

Schooling Infrastructure, Educational Attainment and Earnings*

Sascha O. Becker Frank Siebern-Thomas
CES, CESIFO AND IZA EUROPEAN COMMISSION

This version: October 2007

Abstract

In many countries, students are tracked into a variety of secondary school types. In Germany, tracking takes place at the age of 10. Only one track, high school (*Gymnasium*), leads to a secondary-school diploma (*Abitur*) qualifying for university admission. We show that schooling infrastructure, in particular the local availability of high schools, varies considerably across German counties, and is one crucial determinant of post-compulsory educational attainment. In urban, more densely populated counties, schooling infrastructure is generally better than in rural, less densely populated counties. We find that individuals who grew up in urban areas have a significantly higher educational attainment than individuals who grew up in the countryside. These effects are more pronounced for children from disadvantaged family backgrounds. For them, high school proximity is found to be an important determinant of upper-secondary schooling, which is the prerequisite to access tertiary-level education. The relationship between schooling infrastructure and educational attainment translates into earnings differentials later in life, in particular for this latter group.

Keywords: SCHOOLING INFRASTRUCTURE, SECONDARY-SCHOOL TRACKING, REGIONAL VARIATION

JEL Classification: I21, J24, J31

*We wish to thank David Card and Steve Pischke, as well as seminar and conference participants in Dortmund, Florence, Lausanne, Louvain-la-Neuve, Munich and Salerno for their comments. An earlier version of this paper circulated under the title "Returns to education in Germany - A variable treatment intensity approach". The GSOEP data were provided by the DIW Berlin, Germany. The views expressed represent exclusively the position of the authors and do not necessarily correspond to those of the European Commission. *Corresponding author:* Sascha O. Becker, CES, University of Munich, Schackstr. 4, 80539, Munich, Germany, e-mail: sbecker@lmu.de.

1 Introduction

Schooling infrastructure varies considerably across German counties (*Landkreise*). In urban, more densely populated counties, schooling infrastructure is generally better than in rural, less densely populated counties. By schooling infrastructure, we mean in particular the availability and closeness of the full range of secondary school types. Contrary to other school systems such as in e.g. Belgium, Finland and the UK, the German school system tracks students into differing-ability schools as early as at age 10. The three traditional secondary school types are secondary general school (*Hauptschule*), intermediate school (*Realschule*), and high school (*Gymnasium*).¹ Similar tracking systems can be found in countries such as Austria, Hungary, and the Slovak Republic. Only the *Gymnasium* leads to a secondary-school diploma (*Abitur*) qualifying for university admission. Although it is technically possible to switch tracks in Germany, few students actually do so (see Henz, 1997). The selection into tracks at age 10 is therefore a crucial determinant of educational attainment beyond compulsory schooling (between 8 and 10 years of schooling depending on birth cohort).² The availability of high schools differs significantly between urban and rural areas. While in urban counties, average distance to high schools is relatively small, in the countryside average distance to high school is substantially higher. A larger distance to school increases the costs of education, both the (time) opportunity costs of having to commute longer and the direct transport costs. While these costs may not hamper educational attainment of all students, they are likely to be relevant to those students that are at the margin of continuing school.

We provide evidence for these effects. We find that individuals who grew up in urban areas have a significantly higher educational attainment than individuals who grew up in the countryside. We show that these effects are more pronounced for those with 'low family background' and rationalize this within the model of optimal schooling choice by Gary Becker (1967).

Having established that these educational effects of place of childhood exist, we also evaluate their relevance. We do so by measuring the average earnings loss suffered by those children who, *because of* a childhood spent in the countryside, received less education. The average causal response (ACR) interpretation of instrumental variables (IV) estimates suggested by Angrist

¹Since both the U.S. and the British school systems differ from the German one, there is no ideal translation for the German term *Gymnasium*. Some authors (e.g. Dustmann, 2004) use the British term *grammar school*. Our use of the U.S. term *high school* is not intended to imply that a U.S. high school is the same thing as a German *Gymnasium*.

²For a more extensive discussion of the German school system, see Pischke and von Wachter (2005).

and Imbens (1995) allows us to identify and estimate precisely the effect we would like to measure, namely, the average marginal return to education for individuals who received less education *because of* a childhood spent in a rural area (the *compliers* in the language of Angrist, Imbens and Rubin (1996)). Note that under the conditions required by this interpretation of IV, this is the only average return to schooling that we can identify with our instruments and our sample. However, far from being a limitation, this is precisely the average return in which we are interested, given that our goal is to measure the educational cost/benefit of place of childhood.

Card (1995b) and Kling (2001) present methodologically similar work for the United States. They use college-proximity as an instrumental variable for schooling. In the context of a schooling system that tracks students into different secondary tracks, it is natural to consider schooling infrastructure in a broader sense. The German case is therefore a very relevant case to study in order to understand how the proximity of high schools during childhood influences educational attainment and earnings later in life. While beyond the ambition of this paper, it would also be important to better understand how differences in educational attainment due to early tracking in schooling systems have an impact on future employment prospects and determinants of earnings, including workers' adaptability (to changing labour market requirements) and mobility (between sectors or occupations) as well as their access to lifelong learning.

The article is organized as follows. In the following section, we review Card's (1995b) analytically tractable version of Gary Becker's (1967) optimal schooling model and discuss how heterogeneity in returns to schooling can be exploited to estimate average causal effects for relevant sub-populations. Section 3 presents evidence on the relation between schooling infrastructure and educational attainment in Germany using county-level and individual-level data. Section 4 evaluates the relevance of this effect by measuring the average earnings loss suffered by those children who, because of growing up in a rural area, received less education. Section 5 concludes.

2 Theoretical considerations

In this section, we shortly review Becker's (1967) model of endogenous schooling in the version laid out by Card (1995b). It provides the rationale for heterogeneity in returns to schooling. This heterogeneity can be exploited econometrically in an instrumental variables framework. Angrist and Imbens (1995) discuss two-stage least squares estimation of average causal effects in models with variable treatment intensity. We explain how their approach

can be applied to our case where years of schooling are the treatment.

2.1 Gary Becker’s model of endogenous schooling

An individual maximizes

$$U(y, S) = \log y - \phi(S) \quad (1)$$

where y is average earnings per year, S is years of schooling and $\phi(\cdot)$ is the cost of schooling. An individual’s opportunities are represented by $y = g(S)$. The first order condition of the optimization problem is

$$\frac{g'(S)}{g(S)} = \phi'(S) \quad (2)$$

Now, assume for simplicity that

$$\frac{g'(S)}{g(S)} = \beta_i(S) = b_i - k_1 S \quad (k_1 \geq 0), \quad (3)$$

i.e. there are decreasing marginal benefits to schooling, and

$$\phi'(S) = \delta_i(S) = r_i + k_2 S \quad (k_2 \geq 0), \quad (4)$$

i.e. there are increasing marginal costs to schooling.

The optimal schooling level is then given by $S_i^* = (b_i - r_i)/k$, where $k = k_1 + k_2$. Integrating out (3) yields

$$\log y = b_i S - 0.5 k_1 S^2 \quad (5)$$

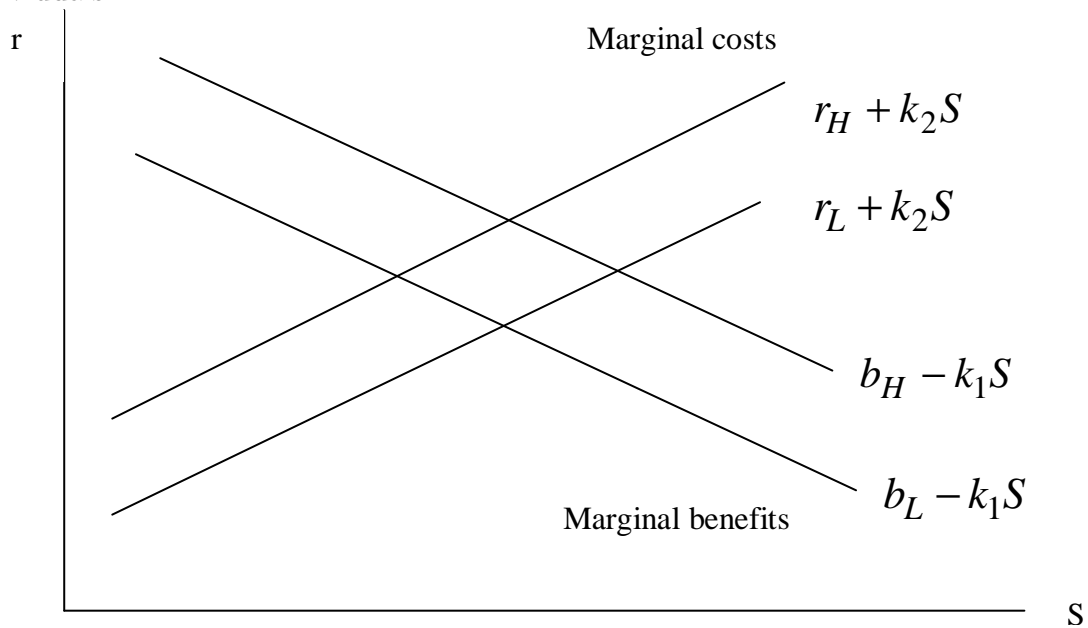
Equations (3) and (4) clearly state the reason for heterogeneous returns to schooling: Individuals are likely to differ in either marginal costs r_i or marginal benefits b_i and are therefore likely to choose different optimal schooling levels.

To illustrate the point, assume there are four population groups, characterized by different intercept parameters $b_H > b_L$ and $r_H > r_L$ for the marginal cost and marginal benefit curves (3) and (4). The four possible combinations of these two values for each parameter characterize four groups of individuals in the population, denoted by (r_L, b_L) (“the stupid rich”), (r_L, b_H) (“the smart rich”), (r_H, b_L) (“the stupid poor”), (r_H, b_H) (“the smart poor”).³ Figure 1 shows that the lowest optimal schooling level

³The group labels are borrowed from Ichino and Winter-Ebmer (1999) who associate higher marginal costs r_H with “the poor” and higher marginal benefits b_H with “the smart”.

arises for those with a low marginal benefit and high marginal cost: (r_H, b_L) ; the highest optimal schooling level arises for those with high marginal benefits and low marginal costs: (r_L, b_H) . In terms of returns to schooling, those with high marginal benefits and high marginal costs, (r_H, b_H) , have the highest returns to schooling. This is the group whose schooling choice beyond compulsory schooling is most likely to be affected by the presence or absence of a good schooling infrastructure, in particular upper-secondary schools (*Gymnasien*). We can think of presence of a good school infrastructure as a downward shift of the marginal cost curve: individuals that, in the absence of a good schooling infrastructure would choose an optimal schooling level arising from the combination (r_H, b_H) would choose a higher schooling level, possibly corresponding to (r_L, b_H) , in the presence of a good schooling infrastructure.

Figure 1: Marginal benefit and marginal cost schedules for different individuals



Note: Illustration of equations (3) and (4) for four different population groups, characterized by parameter combinations (r_L, b_L) (“the stupid rich”), (r_L, b_H) (“the smart rich”), (r_H, b_L) (“the stupid poor”), (r_H, b_H) (“the smart poor”).

2.2 Exploiting heterogeneity in returns to schooling

Heterogeneity in marginal costs and benefits of going to school is exactly what can be exploited empirically by instrumental variables estimation. The Becker model gives rise to the following system of equations:

$$\log y = X\beta + S\gamma + \varepsilon \quad (6)$$

$$S = X\delta + Z\alpha + \eta \quad (7)$$

where Z is an instrument or set of instruments. A given instrument will affect different margins, i.e. different sub-populations at different schooling levels. We can only estimate the average marginal return to schooling for a well-defined subgroup which is affected by the particular instrument.⁴

The model we estimate is an extension of Rubin's Causal Model (RCM) to variable treatment intensity. Assume that each individual would earn Y_j if he or she had j years of schooling for $j = 0, 1, 2, \dots, J$. The objective is to uncover information about the distribution of $Y_j - Y_{j-1}$, which is the causal effect of the j th year of schooling. This will help us understand under which conditions and for which subpopulation of interest γ can be given a causal interpretation. In general, estimates of γ in equation (6) have a causal interpretation only if they have probability limit equal to a weighted average of $E[Y_j - Y_{j-1}]$ for all j in the subpopulation of interest.

We can define potential schooling levels and potential outcomes for all potential values of the instrument for each individual. We define $S_Z \in \{0, 1, 2, \dots, J\}$ to be the number of years of schooling completed by a student conditional on the values of the instrument. Let's initially assume that Z is coded to take on only two values, 1 and 0. S_1 then denotes the years of schooling that would be obtained by an individual growing with $Z = 1$, and S_0 is the years of schooling of the same individual if he or she had been assigned $Z = 0$. In the data, for each individual we observe the triple (Z, S, Y) , where Z is the instrument, $S = S_Z = Z * S_1 + (1 - Z) * S_0$ is years of completed schooling, and $Y = Y_S$ is earnings.⁵ The main identifying assumption is the following

⁴Further assumptions implicit in equations (6) and (7) are log-linearity of earnings in schooling and the absence of degree effects (*sheepskin effects*). See Card (1999) for empirical evidence on the absence of sheepskin effects in the US. In our data, we only find a first-order term of the years of schooling variable to be significant in all our specifications which is again consistent with the absence of sheepskin effects.

⁵Note that, for simplification, we do not use distinct notation for random variables and observations. More correctly, we should denote observations as $(Z_{obs}, S_{obs}, Y_{obs})$, where Z_{obs} denotes the realization of the instrument, $S_{obs} = S_{Z_{obs}} = Z_{obs} * S_1 + (1 - Z_{obs}) * S_0$ is observed years of completed schooling, and $Y_{obs} = Y_{S_{obs}}$ is observed earnings as a function of observed schooling.

Assumption 1 (*Independence*)

The random variables $S_0, S_1, Y_0, Y_1, \dots, Y_J$ are jointly independent of Z .

Assumption 1 is essentially the exclusion restriction and requires that the instrument affects earnings only through its effect on schooling. This implies the existence of unit-level causal effects. To identify a meaningful average treatment effect, the literature typically assumes a constant unit treatment effect, $Y_{ij} - Y_{i,j-1} = \alpha$, for all schooling levels j and all individuals i . Angrist and Imbens (1995), however, impose a nonparametric restriction on the process determining S as a function of Z instead of restricting treatment effect heterogeneity. They impose the following

Assumption 2 (*Monotonicity*)

With probability 1, either $S_1 - S_0 \geq 0$ or $S_1 - S_0 \leq 0$ for each person.

Assumption 2 itself cannot be tested. However, Angrist and Imbens (1995) show that for multi-valued treatments ($J > 1$), assumption 2 has the testable implication that the cumulative distribution function (CDF) of S given $Z = 1$ and the CDF of S given $Z = 0$ should not cross.

From the above assumptions follows the main result in the framework of multivalued treatments:

Theorem 1 *Suppose that Assumptions 1 and 2 hold and that $\Pr(S_1 \geq j > S_0) > 0$ for at least one j . Then*

$$\frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[S|Z = 1] - E[S|Z = 0]} = \sum_{j=1}^J \omega_j \cdot r(j) \equiv \gamma \quad (8)$$

where

$$\omega_j \equiv \frac{\Pr(S_1 \geq j > S_0)}{\sum_{i=1}^J \Pr(S_1 \geq i > S_0)} \quad (9)$$

denotes the covariate weight and where the response function is defined as

$$r(j) \equiv E[Y_j - Y_{j-1} | S_1 \geq j > S_0] \quad (10)$$

This implies that $0 \leq \omega_j \leq 1$ and $\sum_{j=1}^J \omega_j = 1$, so that γ is a weighted average of per-unit average causal effects along the length of an appropriately

defined causal response function. Angrist and Imbens (1995) refer to the parameter γ as the *average causal response* (ACR).

The *covariate weights* ω_j give the weight of the subpopulation, characterized by the respective covariates, in calculating the average treatment effect. The *response function* $r(j)$ shows the weights of the respective schooling levels in computing the average treatment effect. In the presence of further covariates, β has to be interpreted as a variance-weighted average of $\beta(X)$, the ACR in a population with the set of individual characteristics X fixed.

In the empirical part, we present both the weighting function and the response function for our instrument, a measure of schooling infrastructure and thereby try to characterize the affected subgroups and schooling levels.

Before proceeding to the econometric estimates, we motivate our instrument and present evidence on the relationship between schooling infrastructure and educational attainment.

3 Schooling infrastructure and educational attainment

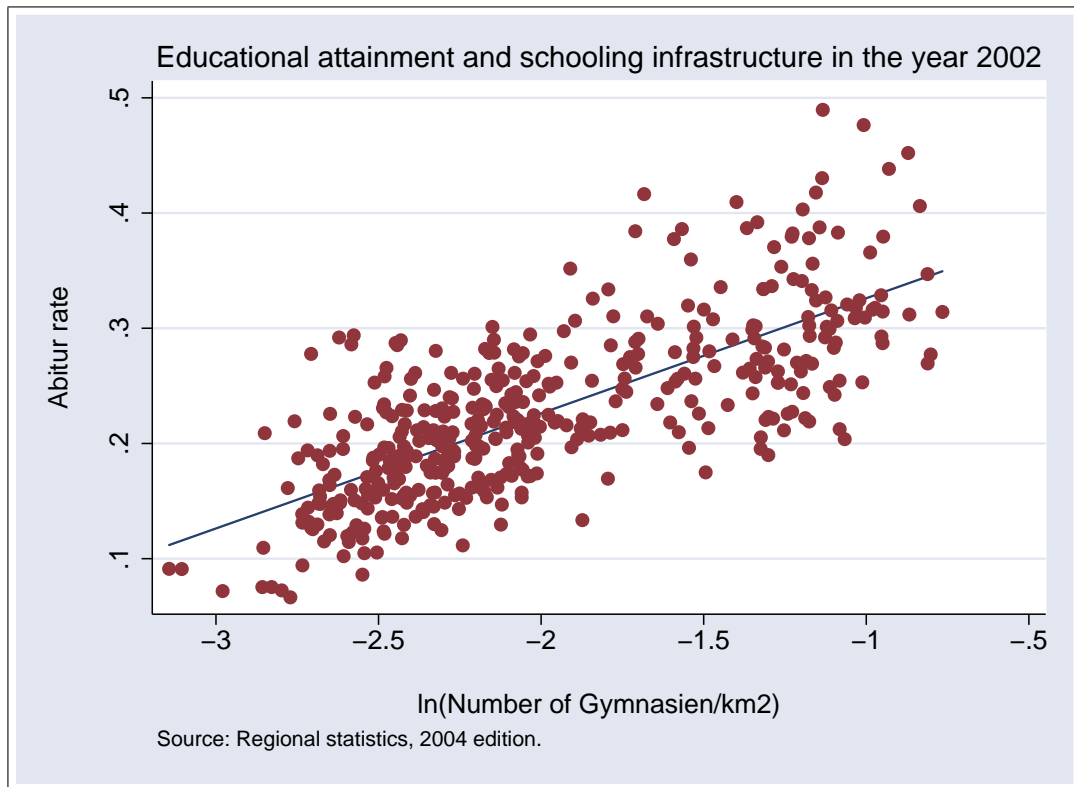
In this section, we present descriptive evidence on the relationship between schooling infrastructure and educational attainment based on two data sources. First, we present descriptive evidence based on county-level administrative data. Second, we turn to micro survey data from the German socioeconomic panel (GSOEP) that we also use in the micro-econometric part of the paper. Both data sets show a huge variation in educational attainment across counties that has a strong positive correlation with schooling infrastructure.⁶

3.1 County-level data

The first data source is county-level data on schooling infrastructure and educational attainment and provided by the German Federal and State Statistical Offices (*Regional Statistics, 2004*). Our discussion focuses on high schools (*Gymnasien*) because successfully completing high school allows access to university. High school completion rates (*Abitur*), i.e. the percentage of school leavers obtaining the secondary-school diploma qualifying for university admission, vary considerably across German counties: from 7% in the

⁶Note that our econometric analysis is restricted to West Germany because the question on place of childhood which is crucial in our analysis was only asked in the GSOEP in 1985, before German unification.

Figure 2: Educational attainment and schooling infrastructure by county



(rural) county of Coburg to 49% in the city of Heidelberg. To see whether there is any systematic relationship between these high school completion rates on the one side and the schooling infrastructure on the other side, we relate the percentage of school leavers having *Abitur* against the (log of) the number of high schools per square kilometer as a measure of schooling infrastructure. By simple geometric arguments, the number of high schools per square kilometer is (inversely) related to the average distance of residents to the nearest high school.⁷ Figure 2 shows that the availability of high schools is in fact seen to be highly correlated with high school completion rates. A larger distance to school increases the costs of education, both the (time) opportunity costs of having to travel longer and the direct transport costs. While these costs may not hamper educational attainment of all students, they may be relevant to those students that are at the margin of pursuing higher education.

⁷We assume, for the sake of simplicity, random location of both individuals and schools *within* county.

Table 1: Percentage of sample with given instrument status

place of childhood	percent	cumulative	binary instrument
city	22.75	22.75	city (<i>pc1</i>)
big town	14.11	36.87	city or big town (<i>pc2</i>)
small town	22.14	59.01	some urban area (<i>pc3</i>)
countryside	40.99	100.00	

Source: 1985 wave of the German Socioeconomic Panel (100% version).

Sample size: N=4096

Sample: full-time employed workers with no missing information on our variables of interest, in particular labor income and schooling.

3.2 Individual-level data

Our second data source is the German Socioeconomic Panel (GSOEP), a household survey comparable to the U.S. Panel Study of Income Dynamics (PSID) or the British Household Panel Survey (BHPS).⁸ The GSOEP does not provide direct measures of schooling infrastructure at the place of childhood. However, the 1985 wave of the GSOEP contains a question on place of childhood that serves as a natural one-dimensional proxy of schooling infrastructure:

'Did you spend the major portion of your childhood up to age 15 in a) a city, b) a big town, c) a small town, or d) in the countryside?'

Table 1 gives the sample distribution of the answers to this question.

In the subsequent analysis we use three binary instruments, denoted by *pc1*, *pc2* and *pc3*, built on the four types of place of childhood, as displayed in table 1.

Answers to this question can be combined with information on educational attainment to provide further descriptive evidence on the relationship between schooling infrastructure and educational attainment. Table 2, panel 1, shows high school completion rates by place of childhood. We find lower high school completion rates for individuals who grew up in rural as opposed to urban areas.⁹ Using the county-level data and defining type of agglomer-

⁸See <http://panel.gsoep.de> for more information on the GSOEP.

⁹Note that the sample is restricted to the full-time employed with non-missing information on the variables of interest, and notably labor income and schooling. The differences in high-school completion rates would be even more pronounced in larger samples, covering also the part-time employed, unemployed and inactive - all of which are groups with

ation by quartiles of population density - which obviously do not perfectly match with the GSOEP classification of place of childhood - we observe a similar pattern. Going from the the least densely to most densely populated quartile, high school completion rates are 14.11, 19.81, 24.33, and 32.47 respectively.

The attentive reader may wonder why high school completion rates differ between county-level data and the GSOEP data. The simple reason is that the GSOEP data include all age groups who were interviewed in 1985, including older cohorts, while the county-level data refer only to school leavers completing their education in the year 2002, i.e. a very recent cohort.¹⁰

Average years of schooling by type of agglomeration show a pattern similar to high school completion rates, as can be seen from the second panel of table 2.

Both county-level data as well as individual-level data thus show that in regions with better schooling infrastructure, educational attainment is higher.

In the next section, we will proceed to evaluate the relevance of this educational effect by measuring the average earning loss suffered by those children who, because of growing up in a rural area - with less favorable schooling infrastructure and therefore higher (direct and indirect) costs of schooling - received less education. This allows us to assess the long-run outcomes of lower educational attainment caused by less favorable schooling infrastructure.¹¹

4 The Effect of Schooling (Infrastructure) on Earnings

The lower panel of table 2 shows that those individuals who grew up in the countryside also earn less than those who grew up in an urban area. The income measure used in the table is the average log of monthly (gross) labor earnings. While there is a number of reasons, most prominently in the New

below average shares of highly educated people.

¹⁰In section 4.6, we will relate this rise in educational attainment to the improvement of schooling infrastructure over the same period. Furthermore, we will use the *differential* increase in improvement of schooling infrastructure as one of our empirical identification strategies.

¹¹Note that the empirical approach is likely to yield a lower bound, since it does not take account of the potential effect of lower educational attainment on future employment prospects and determinants of earnings, including workers' adaptability (to changing labor market requirements) and mobility (between sectors or occupations) as well as their access to lifelong learning.

Table 2: Educational attainment and 1985 labor earnings by place of childhood

percentage with high school degree

city	18.67
big town	15.92
small town	11.36
in the countryside	8.70

average years of schooling

city	11.53
big town	11.30
small town	10.76
in the countryside	10.58

average monthly labor earnings in German Marks

city	3190.33
big town	3106.65
small town	2885.59
in the countryside	2892.76

Source: 1985 wave of the German Socioeconomic Panel (100% version).

Sample size: N=4096

Sample: full-time employed workers with no missing information on our variables of interest, in particular labor income and schooling.

Economic Geography literature (see e.g. Hanson, 2005), why labor earnings depend on type of agglomeration at the *current* place of work, it is less obvious why labor earnings vary by place of childhood. We argued above that place of childhood affects schooling attainment to the extent that schooling infrastructure is poorer in rural areas. While a poor schooling infrastructure may not restrain all students from continuing education, those at the margin between continuing or stopping education may be affected by the proximity to institutions of higher education. In this section, we use an instrumental variables strategy to analyze *how* schooling infrastructure, via its effect on schooling attainment, affects labor earnings, and for *which subgroups* of the population. In other words, we are going to estimate the returns to education for those who are at the margin between continuing education or dropping out of school and whose decision is affected by the available schooling infrastructure. Before describing our own analysis, we give an overview of previous estimates of the returns to education in Germany.

4.1 Previous studies for Germany

Early results for Germany are based on Mincer-style OLS regressions of earnings on schooling. Using years 1984 and 1985 of the German Socioeconomic Panel (GSOEP), Wagner and Lorenz (1989) estimate returns to schooling of 6.5%. In a further study Lorenz and Wagner (1993) give a range of 6.2-7.0% based on the Luxemburg Income Study (LIS 1981) and of 4.0-4.9% using data of the International Social Survey Program (ISSP 1987).

To our knowledge, the only studies using IV estimation for returns to education in Germany are Ichino and Winter-Ebmer (1999, 2004), Lauer and Steiner (2000), and Pischke and von Wachter (2005). Ichino and Winter-Ebmer exploit three different instruments: an indicator of father's education, an indicator of whether an individual was 10 years old during World War II and an indicator of whether their father was in war in this period. Using data from the GSOEP (1986), they give a lower bound of 4.8% and an upper bound of 14% for the return to schooling for those sub-populations that are affected by the respective instruments. Lauer and Steiner (2000) not only estimate the returns to schooling using various estimation methods but also employ IV estimators on the basis of a long list of different instruments. They are above all interested in an analysis of the robustness of the estimated returns to schooling with respect to the various estimation methods and do not provide an interpretation of the obtained IV estimation results. Moreover, the authors conclude that there is no statistical evidence for heterogenous returns to schooling with respect to unobservable characteristics. Pischke and von Wachter (2005) analyze the returns to education

to compulsory schooling in Germany using changes in *compulsory* schooling laws for secondary schools in West German states. They find no return to *compulsory* schooling in Germany in terms of higher wages and conjecture that the result might be due to the fact that the basic skills most relevant for the labor market are learned earlier in Germany than in other countries. Our study is closest to Ichino and Winter-Ebmer (1999,2004) and Lauer and Steiner (2000) because our instruments will mostly pick out differences in secondary schooling.

4.2 IV Estimation Results

As a benchmark to previous results in the literature on Germany, we estimated an OLS regression of earnings on years of schooling controlling for sex, experience and 'tenure on the job' polynomials.¹² We find an estimate of 6.6% which is similar to previous OLS results for Germany (see table , column (1)).

For the reasons given above, the OLS estimates are probably not amenable to an interpretation as the causal effect of schooling on earnings. We therefore focus on an IV estimation of the returns to education on the basis of the instrument 'place of childhood'. The instrumental variables estimates of the returns to schooling on the basis of the chosen instrument have been computed using the two-stage least squares procedure: in the first stage, the years of schooling are regressed on the full list of exogenous variables augmented by the respective instrumental variable using a simple linear probability model; in the second stage, the predicted value of the dependent variable from the first stage regression is then used as additional regressor in the outcome equation instead of the schooling years itself. Table 3 contains the IV estimation results for different specifications. Furthermore, first-stage t -statistics and partial R^2 measures are reported as a diagnostic tool for instrument quality, following the suggestions of Bound et al. (1995) and Staiger and Stock (1997). In all specifications, the instrument quality is good: both first-stage t -statistics and partial R^2 are above the thresholds suggested by these authors.

Table 3, column (2), shows the results when using the binary instrument 'place of childhood in a city' ($pc1$). The estimated return to education is 13.13% (s.e. 4.01). As a robustness check (not reported in the table), we used the instruments 'place of childhood in a city or big town' ($pc2$), and 'place of childhood in an urban area' ($pc3$) to probe to what extent the exact split-up between urban and rural areas matters.

¹²See table A.1 in the appendix for descriptive statistics on the estimation sample.

Table 3: OLS and IV estimates of the Returns to Education

Specification	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	OLS	IV	OLS	IV
1st stage						
<i>Place of childhood = city (pc1)</i>	-	0.37 (0.078)	-	0.32 (0.083)	-	0.23 (0.087)
<i>1st stage F</i>	-	23.21	-	15.33	-	6.71
<i>partial R²</i>	-	0.006	-	0.004	-	0.002
2nd stage						
<i>schooling coefficient</i>	0.066 (0.003)	0.131 (0.040)	0.065 (0.003)	0.137 (0.049)	0.065 (0.003)	0.118 (0.069)
Covariates						
<i>state dummies</i>	no	no	yes	yes	yes	yes
<i>community size dummies</i>	no	no	no	no	yes	yes
<i>second-stage R²</i>	0.39	0.29	0.40	0.27	0.40	0.34
<i>Number of observations</i>	4096	4096	4096	4096	4096	4096

Notes: The Table reports OLS and IV estimates of the coefficient on the *Years of schooling* variable (robust standard errors in parentheses). All regressions control for a quadratic in experience, a quadratic in job tenure, gender, and family background. Experience = age - schooling - 6. In columns 2, 4, and 6, schooling experience, and experience-squared are treated as endogenous, with *Place of childhood* = *city (pc1)*, age, and age-squared as an instrument. Family background variables include indicators for educational and professional attainment of both parents and an indicator for parental presence during childhood.

IV estimation results were very similar: 14.05% (s.e. 2.92) for the instrument *pc2* and 13.56% (s.e. 3.41) for the instrument *pc3*. The exact definition of the binary instrument therefore does not seem to matter a lot and we will concentrate on the instrument *pc1* in the sequel.

In the light of the Angrist and Imbens (1995) framework, these results can be interpreted as the average marginal returns to education for those who acquired more education *because* they grew up in an area with a better schooling infrastructure (i.e. a more urban area).

In the following sections, we are going to concentrate on three questions which are crucial for interpreting the IV estimates. First, following Angrist and Imbens (1995) and Kling (2001), we are going to characterize the subgroups of the population affected by our instrument, thereby giving evidence on the external validity of our instrument. Second, we will characterize the response function, i.e. we will show at which schooling levels the effect of our instrument is most pronounced. Third, we will show several robustness checks to make sure that the instrument 'place of childhood' is valid, i.e. we assess the internal validity of our instrument.

4.3 Characterizing the compliers

If we want to generalize our estimates to some larger populations ("external to the sample"), we have to characterize as closely as possible the subgroups affected by our instrument and the size of the effect on them. Angrist, Imbens and Rubin (1996) call this group the compliers, i.e. those who take further schooling (and earn more later on) only because they grew up in an urban area as opposed to a rural area. We said above that the effect of schooling infrastructure is likely to be more important for children from less advantaged family backgrounds. Growing up in a rural area is so-to-say the worst case scenario in terms of educational opportunities, where only a favorable parental background may help in obtaining further degrees. To test this, we follow Card and Kling in defining an index of family background in the following way:

First, we perform a regression of years of schooling on gender, a quadratic in age and, most importantly, family background variables (parental education, parental presence during childhood) for the subgroup of people who spent their childhood in a rural area.

Second, based on the parameter estimates obtained, we predict - for all individuals - their 'counterfactual schooling level, had they grown up in a rural area'. This gives a single index measure of family background variables, which we use to split the sample into four quartiles, from the lowest (fbq1) to the highest (fbq4). The *actual* years of schooling follow exactly the

predictions: those in the lowest family background quartile have on average 9.31 years of schooling, while those in the second, third and fourth quartiles have, on average, 10.64, 11.40 and 12.42 years of schooling, respectively.

Table 4 describes some key differences in attributes across the four family background quartiles. It is interesting to note that in the lowest background quartile, barely any individual reports that either their father or mother graduated from high school. There is no single individual in the lowest three family background quartiles whose father has a university degree. Conversely, there is virtually no individual in the two highest background quartiles who has a father without a schooling degree. Furthermore, it is interesting to note that younger individuals are more likely to be in higher family background quartiles. This is an indication of the rising trend in educational attainment, in particular for those growing up in a rural area.¹³

The IV estimate of the returns to schooling can be interpreted as a weighted average of the causal effect of a year of schooling within a population subgroup, in our case a family background quartile q . In a population subgroup q , denote by $\Delta Y_q = E[Y|Z = 1, q] - E[Y|Z = 0, q]$ the impact on earnings, by $\Delta S_q = E[S|Z = 1, q] - E[S|Z = 0, q]$ the impact on schooling, by $\gamma_q = E[\gamma_i|q]$ the average return to schooling and let $\omega_q = P(q)$ be weights. This allows us to write

$$\gamma = \frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[S|Z = 1] - E[S|Z = 0]} = \frac{\sum_{q=1}^4 \omega_q \Delta Y_q}{\sum_{q=1}^4 \omega_q \Delta S_q} = \frac{\sum_{q=1}^4 \omega_q \Delta S_q \gamma_q}{\sum_{q=1}^4 \omega_q \Delta S_q} \quad (11)$$

In our application, the instrument is only valid conditional on X . The notation in equation (11) changes accordingly. Let $P(Z|X, q)$ be the conditional probability of growing up in a city ($pc1=1$). The IV estimates controlling for X and q impose weighting from the regression proportional to the conditional variance on Z , $P(Z|X, q)(1 - P(Z|X, q))$, as shown by Angrist (1998). We define $\lambda_{q|X} = E[P(Z|X, q)(1 - P(Z|X, q))|q]$. If the instrument were independent of X and q , then $\lambda_{q|X}$ would be a constant. Furthermore denote by $\Delta S_{q|X} = E[E[S|Z = 1, q] - E[S|Z = 0, q]|q]$ the expected difference (conditional on X and q) in educational attainment by instrument status. The overall weight received by each quartile using two-stage least squares is $\omega_{q|X} = (\lambda_{q|X} \Delta S_{q|X}) / (\sum_q \lambda_{q|X} \Delta S_{q|X})$.

Table 5 shows estimates of $\omega_{q|X}$ and its components. Column (1) contains, for each quartile, $\lambda_{q|X}$, estimated using linear regression including X and

¹³Remember that family background quartiles are predictions based on a regression for the subgroup of people who spent their childhood in a rural area.

Table 4: Probability of characteristics by family background quartile

Background quartile	1	2	3	4	Avg.
Father's education					
High school degree	0.12	3.98	0.30	21.11	6.79
Professional school	0.00	3.18	6.32	9.90	6.93
University degree	0.00	0.00	0.00	16.60	6.36
No schooling degree	64.93	11.54	0.10	0.00	17.65
Mother's education					
High school degree	0.00	0.00	0.40	6.83	1.94
Professional school	0.00	0.00	0.10	3.32	1.26
University degree	0.00	0.29	0.00	3.01	1.18
No schooling degree	74.42	16.87	0.00	0.42	21.80
Own characteristics					
Female	35.64	36.26	34.85	10.37	29.35
Mean age	40.44	37.44	35.77	34.67	37.09

Note: Family background quartiles are computed as follows. Following Card (1995b) and Kling (2001), a predicted value is estimated from a regression of schooling level on family background variables (educational and professional attainment of both parents, parental presence during childhood) and a polynomial in (own) age for the sample of 1679 individuals grown up in a rural area. The 25th, 50th, and 75th percentiles of the predicted values from this sample were used to group all 4096 observations with valid information on our variables of interest, in particular labor income and schooling into four quartiles. In the first/lowest quartile, 15.29 percent of the individuals grew up in a city, in the second quartile 23.78 percent, in the third quartile 23.88 percent, and in the fourth/upper quartile 28.13 percent.

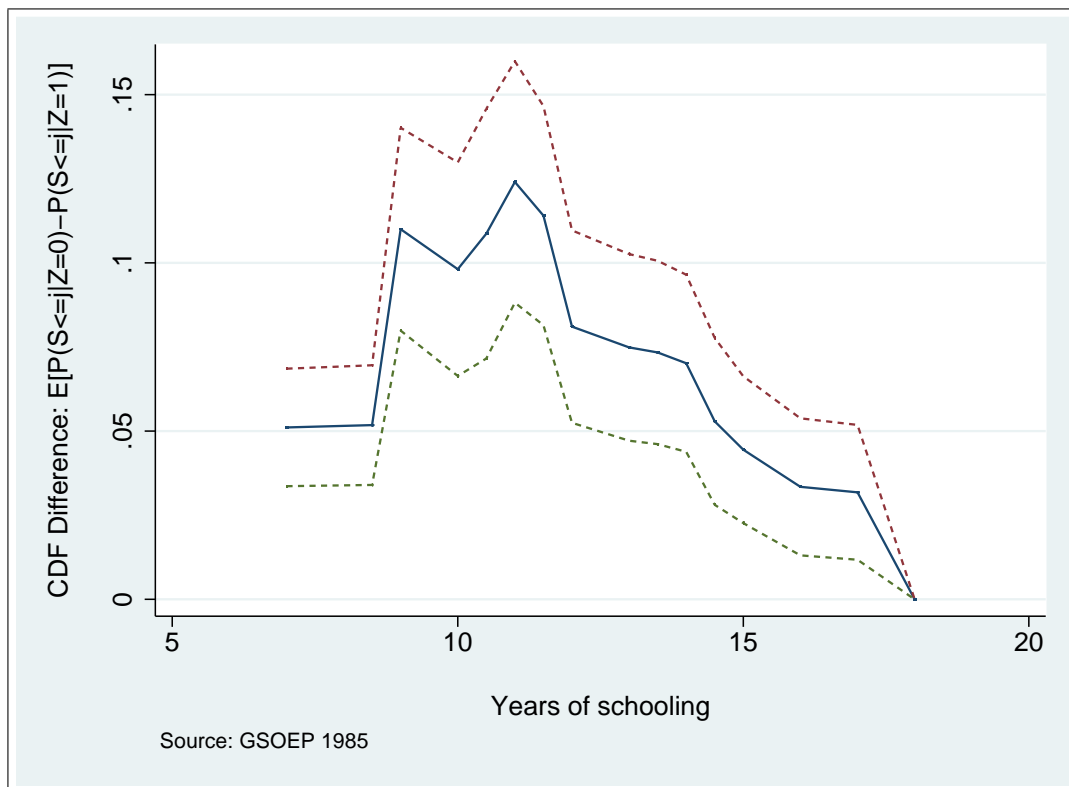
Table 5: Decomposition of IV weighting by family background quartile

q	$\lambda_{q X}$ (1)	$\Delta S_{q X}$ (2)	$\omega_{q X}$ (3)
1st (lowest) quartile	0.13	0.62 (0.18)	0.31
2nd quartile	0.18	0.58 (0.15)	0.41
3rd quartile	0.18	0.28 (0.15)	0.20
4th (highest) quartile	0.19	0.11 (0.15)	0.08

Note: $\lambda_{q|X} = E[P(Z|X, q)(1 - P(Z|X, q))|q]$. $\Delta S_{q|X} = E[E[S|Z = 1, q] - E[S|Z = 0, q]|q]$. $\omega_{q|X} = (\lambda_{q|X} \Delta S_{q|X}) / (\sum_q \lambda_{q|X} \Delta S_{q|X})$. $\lambda_{q|X}$ and $\Delta S_{q|X}$ are computed using linear regression as described in the text.

q to estimate $P(Z|X, q)$, and then taking expectation over the empirical distribution function of X for each value of q (see Angrist, 1998 and Kling, 2001). Column (2) contains $\Delta S_{q|X}$, also computed using linear regression and corresponding closely to the two-stage least squares results in table 3. $\Delta S_{q|X}$ is captured by the coefficients on interactions between Z and q . Column (3) contains the weight $\omega_{q|X}$ that would be used to form a weighted average of the marginal return to schooling obtained from a separate IV earnings regression in each group q . The weights clearly show that the two lower family background quartiles receive more weight in the overall IV estimate. The first and second quartiles receive a weight of together 72% while the two upper quartiles only 28%. Unfortunately, the sample size is too small to obtain reliable estimates of the return to schooling in each quartile separately. But the weights are informative by themselves. To the extent that individuals from lower family background quartiles are likely to have higher marginal returns to schooling, the relatively high overall IV estimates (close to the upper bound provided by Ichino and Winter-Ebmer, 1998) are consistent with those groups obtaining a higher weight.

Figure 3: CDF difference using pc1 as an instrument



4.4 Characterizing the response function

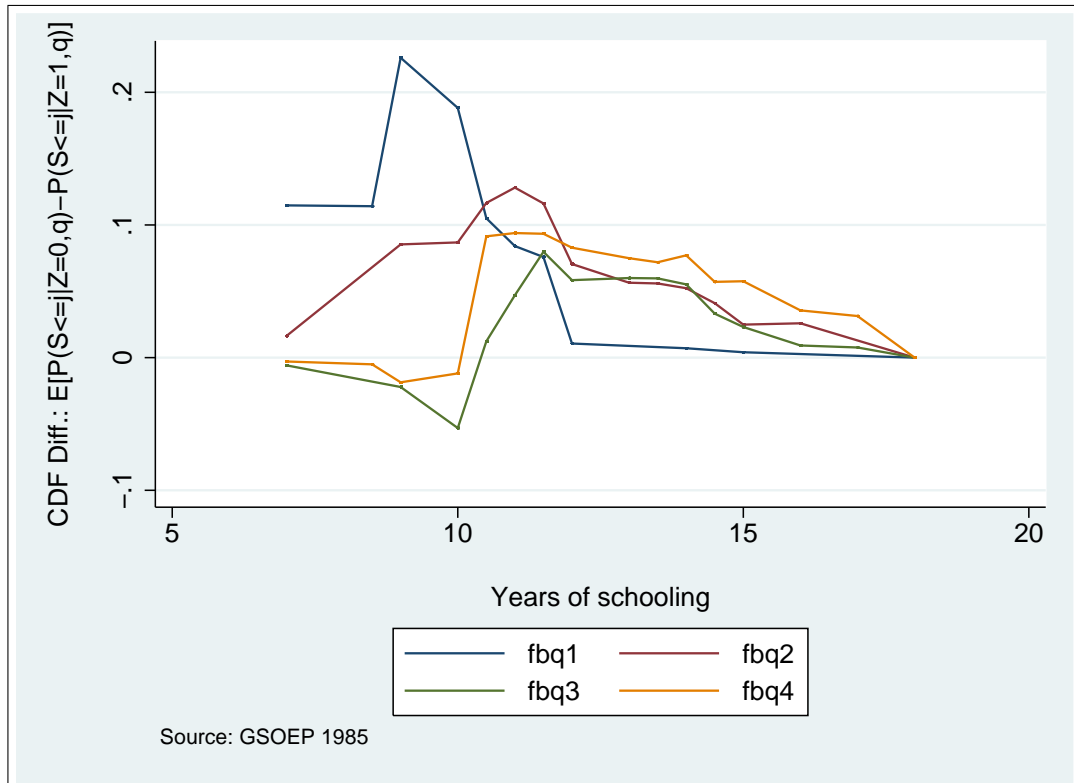
The response function can be estimated from the cumulative distribution functions (CDF) of schooling at different values of the instrument. The difference in the CDFs is equivalent to the fraction of the population who received at least one more year of schooling due to the instrument. Figure 3 displays $Pr(S < j | Z = 0, X) - Pr(S < j | Z = 1, X)$, the difference between the CDF of schooling for $Z = 1$ and $Z = 0$.¹⁴ It indicates that schooling infrastructure has its largest effect at 9 and 12 years of schooling.¹⁵ More specifically we interpret the estimates to indicate that around 10 percent of individuals with similar demographics are induced to obtain more years of schooling due to better schooling infrastructure.

¹⁴The 95% confidence bands are calculated using the conventional formula for a difference in proportions.

¹⁵Grades 9 and 12 refer to two important moments in a student's career, when students decide whether to continue after compulsory schooling and whether to continue to study at university or not, respectively.

Further insight can be gained by breaking down the response function by background quartiles. Figure 4 shows that the response functions for the lower family background quartiles peak earlier than the response function for the upper family background quartiles. For those from lower family background quartiles, the availability of high schools (Gymnasien) nearby is hence an important factor for going beyond compulsory schooling (between 8 and 10 years of schooling, depending on cohort). The fraction of compliers in the lowest family background quartile that obtain 9 or more years of schooling *because of* a good schooling infrastructure is around 20 percent. For those from the upper family background quartiles, on the contrary, the availability of universities nearby seems to matter most. The response function for those groups peaks at around 12-13 years of schooling. Furthermore, the fraction of compliers in the two upper quartiles is overall much lower, again showing that the instrument affects mainly the two lower family background quartiles.

Figure 4: CDF difference by family background quartile using pc1 as an instrument



4.5 Validity of the instrument(s)

The IV results presented so far are only valid provided that the conditions laid out in section 2.2 are fulfilled.

We start by checking whether monotonicity (Assumption 2) holds. Monotonicity has to hold for every individual, an assumption which cannot be tested. Monotonicity signifies that an individual growing up in a city (i.e. in region with good schooling infrastructure) takes at least as much schooling as if he had grown up in the countryside (i.e. in a region with a worse infrastructure). This assumption rules out *defiers*, i.e. individuals who, if growing up in a city, take *less* schooling than if growing up in the countryside. In theory, there might be individuals who take less schooling growing up in an urban area due to e.g. drugs and delinquency, but growing up in a rural area would have obtained more schooling. In a similar way, labor demand in cities might be higher and therefore students might have more outside options in a city as compared to an urban area and for some individuals these outside options might lead to a lower schooling level. While we cannot really rule out that there are some cases like this, for the reliability and interpretability of our estimates it is important that the *fraction* of defiers is nevertheless very small.

One testable implication of strong monotonicity, however, is that the cumulative distribution functions (CDFs) of schooling by instrument status do not cross (see section 2.2). Figure 3 actually shows that this difference is statistically significantly different from zero at all values of schooling, i.e. passes the test suggested by Angrist and Imbens (1995) and makes us confident that violation of the strong monotonicity assumption is not a serious issue here.

It is furthermore very important to test the *exclusion restriction* inherent in Assumption 1. The exclusion restriction could be violated for several reasons.

One potential violation would be the existence of (direct) agglomeration effects on wages (see Hanson, 2005). We are in the fortunate situation to have some information about the current place of living, so we can control for type of agglomeration.

When adding dummies for current state (Bundesland) of residence as additional control variables, the estimation results remain essentially unchanged (see table 3, column (4)).

Within-state heterogeneity in agglomeration types may be substantial, in particular in the bigger German states (e.g. Bavaria, Lower Saxony and North RhineWestphalia). To control more directly for agglomeration effects on wages we use a set of dummies for community size classes. Table 3, column (6), shows that the IV estimates are still very close to the basic

IV specification in column (2). Moreover, the community size dummies are jointly statistically insignificant. We conclude from this evidence that there is no violation of the exclusion restriction through agglomeration effects (urban wage premia).

Since the urban wage premium is probably only related to characteristics of the *current* place of residence, it is a weak 'test' of the exclusion restriction. In fact, the exclusion restriction has to hold for *any* potential effect of place of childhood on earnings. Controlling for characteristics of the *current* place of residence may therefore be insufficient because of mobility between community types since childhood.

To see if this is a valid objection, we follow an idea similar to Card (1995b) and Kling (2001). In order to test whether college proximity is a legitimate instrument, they use the interaction of college proximity with an indicator for low parental background as an instrument and control for the main effect of college proximity. Applied to our context, the idea of this setup is that individuals from higher family background quartiles are likely to be less affected by place of childhood because they have the necessary support by their family to pursue further education in any case, even if the respective schools are not nearby. If this assumption is correct - and figure 4 gave support to it - only individuals in a low family background quartile will be affected by our instrument. In this case, using the instrument as such or using the instrument interacted with an indicator of low family background identifies essentially the same group of compliers, but gives us one more degree of freedom, which allows us to control for the main effect of 'place of childhood'.

Table 6, columns (1)-(3) show that indeed the main effect of 'growing up in an urban area' is small in size and statistically insignificant in all specifications, which we take as evidence that the exclusion restriction is likely to hold.¹⁶

The point estimates of the returns to schooling are generally somewhat lower than the ones where we do not control for the main effect of 'growing up in an urban area'. Column (1) shows the return to schooling when the interaction is between *pc1* and the lowest family background quartile. Columns (2) and (3) use the interaction between *pc1* and the lowest two respectively the lowest three family background quartiles. As we go from the most disadvantaged groups (first family background quartile) to less and less disadvantaged groups, the point estimates go down, consistent with our expectations that the most disadvantaged groups have the highest marginal returns to growing up in an urban area.

¹⁶Also note that the family background variables are never jointly significant.

Table 6: IV estimates of the Returns to Education: Place of childhood interacted with family background indicators

<i>IV: pc1 * fbq</i>	lowest	lowest two	lowest three
<i>returns to schooling</i>	0.113 (0.019)	0.101 (0.019)	0.096 (0.017)
<i>main effect of pc1</i>	-0.001 (0.019)	0.007 (0.018)	0.011 (0.019)
<i>second-stage R²</i>	0.34	0.37	0.37
<i>Number of observations</i>	4096	4096	4096

Notes: The Table reports IV estimates of the coefficient on the *Years of schooling* variable. The IV estimates use *Place of childhood = city (pc1)* interacted with indicators of family background as instruments. In column (1), *pc1* is interacted with a dummy for the lowest family background quartile. In column (2), *pc1* is interacted with a dummy for the two lowest family background quartiles. In column (3), *pc1* is interacted with a dummy for the three lowest family background quartiles. All regressions control for a quadratic in experience, a quadratic in job tenure, and gender. Robust standard errors are reported in parentheses.

Table 7: IV estimates of the Returns to Education: Birth cohort effects

<i>birth cohorts</i>	1930-1945	1930-1950	1930-1965
<i>schooling coefficient</i>	0.187 (0.051)	0.170 (0.044)	0.131 (0.040)
<i>second-stage R²</i>	0.02	0.05	0.29
<i>Number of observations</i>	1680	2253	4096

Notes: The Table reports OLS and IV estimates of the coefficient on the *Years of schooling* variable. The IV estimates use *Place of childhood = city (pc1)* as an instrument. All regressions control for a quadratic in experience, a quadratic in job tenure, and gender. Robust standard errors are reported in parentheses.

4.6 Birth cohort effects

To further assess the robustness of our identification strategy, this section analyzes differences in returns to education for different birth cohorts. In section 3.2, we gave evidence on the rise in educational attainment by comparing a cohort of recent school leavers (using regional data) to older cohorts (using the GSOEP). It turns out that the upward trend in educational attainment over the last decades - not only in Germany, but worldwide (see e.g. Barro and Lee, 2001) - went along with an expansion in the number of schools. The most interesting part of the increase in schooling provision is the *differential* increase in number of schools in rural and urban areas. In Appendix A.1, we provide descriptive evidence that, after WW II, the number of high schools increased faster in rural areas. This suggests that growing up in a rural area constituted a more severe educational cost for earlier cohorts as opposed to more recent cohorts.

Therefore, as illustrated by figure 1, earlier cohorts (having relatively higher marginal costs r_H) can be expected to have both lower optimal schooling levels as well as higher marginal returns to schooling than more recent cohorts (having relatively lower marginal costs r_L). To the extent that the instrumental variable 'place of childhood' identifies the compliers, we should see higher returns to education for earlier cohorts when using our instrument.

Table 7 shows that the returns to education are indeed higher for earlier cohorts (of compliers). For the birth cohorts 1930-1945, returns to education

identified by the instrument 'place of childhood' are considerably higher (18.7 percent) than for all birth cohorts together (13.1 percent). A similar pattern arises when looking at the 1930-1950 cohort with returns to education of 17.0 percent. When estimating returns for the younger birth cohorts, 1945-1965, separately, estimates drop to 8.1 percent, with larger standard errors (6.5), indicating that the instrument is not as strong for more recent cohorts as it is for earlier cohorts. In general, this evidence is consistent with a reduction over time in the importance of school proximity as an inducement for educational attainment.¹⁷

5 Summary and conclusions

We show that educational attainment differs considerably by whether place of childhood was in an urban or in a rural area. In urban, more densely populated counties, schooling infrastructure is generally better than in rural, less densely populated counties. In particular, the supply of high schools (*Gymnasien*), the only secondary school track allowing university entry, is higher in urban counties. As a consequence, in urban counties, average distance to high school is relatively small, and therefore the direct (commuting) and indirect (time) costs of going to high school are relatively small. In the countryside, average distance to high school is substantially higher and therefore the costs of taking further schooling beyond compulsory schooling are higher. Indeed, we give evidence that educational attainment is higher in counties with better schooling infrastructure.

Having established that these educational effects of place of childhood exist, we evaluate their long-run consequences in terms of labor earnings. We do so by measuring the average earnings loss suffered by those children who, *because of* a childhood spent in the countryside, received less education. The average causal response (ACR) interpretation of instrumental variables (IV) techniques suggested by Angrist and Imbens (1995) allows us to identify the subgroups of the population most affected by (non-)availability of schooling infrastructure.

We show that individuals from different family backgrounds respond differently to our instrument 'place of childhood'. We find a stronger effect on low schooling levels for individuals with 'poor family background' and

¹⁷It would be interesting to redo the analysis by background quartiles and schooling levels (characterising compliers and response functions) for the chosen cohorts, since despite lower estimates for younger cohorts, school proximity for some subpopulations, notably low family background, may have remained important. However, the number of observations is too small to yield reliable estimates for the birth cohort subsamples.

a stronger effect on high schooling levels for individuals with 'rich family background'.¹⁸

Furthermore, we show that older birth cohorts obtained both less schooling and had higher marginal returns to schooling than more recent cohorts, consistent with the empirical observation that schooling infrastructure improved relatively more in rural areas.

The finding that educational attainment crucially depends on the provision of post-compulsory schooling in proximity to the place of living, has important policy implications. Consider the case of a local government that has decided to devote a certain amount of money to the improvement of upper secondary schooling infrastructure.¹⁹ The local government faces the decision *where* to build the school, in an urban area or in a rural area, or similarly whether to build one big school in a city or some smaller schools in the countryside. If the per student cost of providing further places at school is constant independent of where schools are built, our results clearly indicate that students living in areas with a less favorable schooling infrastructure would probably benefit most from such an investment because of their above average marginal returns to education. To the extent that schooling infrastructure is correlated with the degree of urbanization, providing a better schooling infrastructure especially in rural areas could thus considerably increase the incentives for individuals from disadvantaged family background to acquire more education and thus improve their long-run prospects in the labor market.²⁰

Our results shed light on an important difference between countries that track students into different secondary school types and countries that do not track students. While in the United States, college proximity is an important constraint in obtaining post-compulsory education (see Card (1995b) and

¹⁸The estimated returns to schooling are at best a lower bound, since the empirical approach does not consider the potential impact of lower educational attainment on future employment prospects and determinants of earnings, including workers' adaptability (to changing labor market requirements) and mobility (between sectors or occupations) as well as their access to lifelong learning.

¹⁹We do not address the cost-benefit issue here, i.e. we do not ask whether investing in schooling infrastructure is beneficial as such. In contrast, we take an individual-level perspective and take the provision of funds by the government as given in this thought experiment.

²⁰It is important to note, though, that the policy implication might be quite different for the case in which the federal government increases schooling infrastructure in the country as a whole. In this case there might be general equilibrium effects that decrease the return to education in the long run due to an overall higher supply of better-educated individuals (see Heckman et al., 1999). The policy implications of this paper do therefore refer to the optimal allocation of schools but not necessarily to the optimal overall spending on schooling infrastructure.

Kling (2001)), in Germany, tracking at age 10 makes proximity of high schools (*Gymnasien*) a relevant factor at a much earlier stage (see also Dustmann, 2004). In future research, it would be interesting to see if our results also hold in other countries with tracking systems (e.g. Austria, Hungary and the Slovak Republic) and also confront them with similar estimates for countries without such early tracking systems such as e.g. Belgium, Finland or the UK.

References

- Angrist, Joshua D.**, “Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records,” *American Economic Review*, June 1990, 80 (3), 313–336.
- and **Alan B. Krueger**, “Does Compulsory School Attendance Affect Schooling and Earnings?,” *Quarterly Journal of Economics*, November 1991, 106 (4), 979–1014.
- and — , “Empirical Strategies in Labor Economics,” in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3A, Amsterdam, New York and Oxford: Elsevier Science, North-Holland, 1999, pp. 1277–1366.
- and **Guido W. Imbens**, “Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity,” *Journal of the American Statistical Association*, June 1995, 90 (430), 431–442.
- , — , and **Donald B. Rubin**, “Identification of Causal Effects Using Instrumental Variables,” *Journal of the American Statistical Association*, 1996, 91 (434), 444–472.
- Becker, Gary S.**, *Human Capital and the Personal Distribution of Income*, Ann Arbor: University of Michigan Press, 1967.
- Bound, John, David A. Jaeger, and Regina M. Baker**, “Problems with Instrumental Variables Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak,” *Journal of the American Statistical Association*, 1995, 90, 443–450.
- Card, David**, “Earnings, Schooling, and Ability Revisited,” *Research in Labor Economics*, 1995, 14, 23–48.

- , “Using Geographic Variation in College Proximity to Estimate the Return to Schooling,” in Louis N. Christofides, E. Kenneth Grant, and Robert Swidinsky, eds., *Aspects of labour market behaviour: Essays in honour of John Vanderkamp*, Toronto, Buffalo and London: University of Toronto Press, 1995, pp. 201–22.
- , “The Causal Effect of Education on Earnings,” in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3A, Amsterdam, New York and Oxford: Elsevier Science, North-Holland, 1999, pp. 1801–63.
- Dustmann, Christian**, “Parental Background, Primary to Secondary School Transitions, and Wages,” *Oxford Economic Papers*, 2004, 56 (2), 209–230.
- Hanson, Gordon**, “Market Potential, Increasing Returns, and Geographic Concentration,” *Journal of International Economics*, September 2005, 67, 1–24.
- Heckman, James J., Lance Lochner, and Christopher Taber**, “Human Capital Formation and General Equilibrium Treatment Effects: A Study of Tax and Tuition Policy,” *Fiscal Studies*, March 1999, 20 (1), 25–40.
- Henz, Ursula**, “Der Beitrag von Schulformwechseln zur Offenheit des allgemeinbildenden Schulsystems,” *Zeitschrift für Soziologie*, 1997, 26 (1), 52–69.
- Ichino, Andrea and Rudolf Winter-Ebmer**, “Lower and Upper Bounds of Returns to Schooling: An Exercise in IV Estimation with Different Instruments,” *European Economic Review*, 1999, 43, 889–901.
- and — , “The Long-Run Educational Cost of World War II,” *Journal of Labor Economics*, January 2004, 22 (1), 57–86.
- Imbens, Guido W. and Joshua D. Angrist**, “Identification and Estimation of Local Average Treatment Effects,” *Econometrica*, 1994, 62, 467–475.
- Kane, Thomas J. and Cecilia E. Rouse**, “Labor Market Returns to Two- and Four-Year Colleges: Is a Credit a Credit and Do Degrees Matter?,” *American Economic Review*, June 1995, 85 (3), 600–614.

- Kling, Jeffrey R.**, “Interpreting Instrumental Variables Estimates of the Returns to Schooling,” *Journal of Business and Economic Statistics*, July 2001, 19 (3), 358–64.
- Lang, Kevin**, “Ability Bias, Discount Rate Bias and the Returns to Education,” *Boston University*, 1993, *mimeo*.
- Lauer, Charlotte and Viktor Steiner**, “Returns to Education in West Germany, An Empirical Assessment,” *ZEW Discussion Paper*, 2000, 00-04.
- Lorenz, Wilhelm and Joachim Wagner**, “A Note on Returns to Human Capital in the Eighties: Evidence from Twelve Countries,” *Jahrbücher für Nationalökonomik und Statistik*, 1993, 211 (1-2), 60–71.
- Pischke, Jörn-Steffen and Till von Wachter**, “Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation,” *NBER Working Paper 11414*, 2005.
- Staiger, Douglas and J. Stock**, “Instrumental Variables Regression with Weak Instruments,” *Econometrica*, 1997, 65 (3), 557–586.
- Wagner, Joachim and Wilhelm Lorenz**, “Einkommensfunktionsschätzungen mit Längsschnittdaten für vollzeiterwerbstätige deutsche Männer. (Estimates of Earnings Functions for Full-Time Working German Men Using Longitudinal Data),” *Konjunkturpolitik*, 1989, 35 (1-2), 99–109.

Table A.1: Descriptive statistics for the variables used in the analysis

Variable	Mean	Std.Dev	Min	Max
female	0.29	0.46	0	1
age	37.09	10.05	20	55
experience	20.15	10.52	0	42
tenure	9.75	8.00	0	41
monthly earnings	2989.29	1378.89	60	19000
years of schooling	10.94	2.51	7	18
<i>community size</i>				
≤ 2.000	0.05	0.21	0	1
2.000 - 5.000	0.08	0.27	0	1
5.000 - 20.000	0.24	0.42	0	1
20.000 - 50.000	0.18	0.38	0	1
50.000 - 100.000	0.09	0.29	0	1
100.000 - 500.000	0.17	0.38	0	1
≥ 500.000	0.20	0.40	0	1

Source: 1985 wave of the German Socioeconomic Panel (100% version).

Sample size: N=4096

Sample: full-time employed workers with no missing information on our variables of interest, in particular labor income and schooling.

A.1 Differential development of schooling infrastructure in rural vs urban areas

To illustrate the conjecture, presented in section 4.6, that differences in schooling infrastructure between rural and urban areas were more pronounced in older cohorts, we collected data on the construction of high schools since World War II. This information is not available in electronic format, so we had to collect printed volumes from several decades. Statistics of this kind are collected at the level of individual states (Länder). Since the effort involved in visiting the libraries of the statistical offices of the 16 individual German states is considerable and the purpose of our exercise is just to illustrate the conjecture, we deliberately chose to provide evidence only for Bavaria, Germany's largest state.

Figure A.1 shows that the growth in the number of high schools (Gym-

nasien) was stronger in rural counties (*Landkreise*) as opposed to urban counties (*kreisfreie Städte*). This is evidence for the catching-up of rural areas in schooling infrastructure.

This descriptive evidence suggests that growing up in a rural area constituted a more severe educational cost for earlier cohorts as opposed to more recent cohorts and is consistent with the empirical finding that returns to education are lower for more recent, less infrastructure-constrained, birth cohorts.

Figure A.1: Total number of high schools (*Gymnasien*) in Bavaria since 1950 - rural vs. urban

