EUROPEAN UNIVERSITY INSTITUTE

DEPARTMENT OF ECONOMICS

EUI Working Paper **ECO** No. 2001/9

Returns to Education in Germany: A Variable Treatment Intensity Approach

Sascha Becker

and

Frank Siebern-Thomas

BADIA FIESOLANA, SAN DOMENICO (FI)

Printed in Italy in June 2001
European University Institute
Badia Fiesolana
I-50016 San Domenico (FI)
Italy

Returns to Education in Germany

A variable treatment intensity approach

Sascha Becker and Frank Siebern-Thomas*†

May 2001

Abstract

High school completion rates vary considerably across West-German counties (Landkreise) and are highly correlated with measures of schooling infrastructure. We argue that 'place of childhood' as a proxy of schooling infrastructure is a convincing exogenous source of variation in schooling levels that allows us to identify the causal effect of schooling for well-defined subgroups of the population. Using the variable treatment intensity approach exposed by Angrist and Imbens (1995) we find that individuals from 'poor family background' respond most strongly to the instrument 'place of childhood'. Their response is further most pronounced at low schooling levels whereas the response of individuals with 'rich family background' is most pronounced at higher schooling levels. Finally, this approach allows us to detect changes in the response function over time.

Keywords: RETURNS TO SCHOOLING, CAUSAL EFFECT, INSTRUMENTAL VARIABLES, REGIONAL VARIATION

JEL Classification: I21, J24, J34

---

*European University Institute, Via dei Roccettini 9; I-50016 Fiesole; email: sbecker@iue.it siebern@iue.it

†We would like to thank Andrea Ichino and seminar participants in Florence, Louvain-la-Neuve, Munich, Salerno, and at the CEPR European Summer Symposium in Labour Economics 2001 for their comments. The GSOEP data were used by permission of the DIW. Financial support by the European University Institute is gratefully acknowledged.

1 Introduction

The presence of heterogeneity in returns to schooling seems by now well established. Building on Gary Becker's (1967) model of optimal schooling according to which individuals choose their optimal schooling level by equating marginal benefits from continuing in education with the related marginal costs, recent theoretical contributions by *inter alia* Card (1995a,1995b) and Lang (1993) argue that individuals with different unobservable characteristics like ability, liquidity constraints or discount rates are likely to incur different marginal costs and benefits of further education and hence self-select into specific schooling levels. Such differences in the marginal costs and benefits of schooling imply different returns to schooling at different optimal schooling levels.

This in turn suggests the estimation of the returns to schooling on the basis of adequate instrumental variables and an interpretation of these estimates as *local average treatment effects* (LATE) along the lines of Angrist, Imbens and Rubin (1996): The estimated returns apply only to those individuals who are affected by the underlying instrument, i.e. those who only continue in school one more year because of their being induced to it by the instrument; in the language of the evaluation literature, the instrument is interpreted as an assignment to treatment, with one more year of schooling being interpreted as treatment. Moreover, different instruments will naturally affect different subgroups and hence lead to varying estimates of *the* returns to schooling.[1]

Several empirical studies on the returns to schooling in the US seem to corroborate the LATE interpretation of instrumental variables estimates.[2] Card (1995b) and Kling (2000) e.g. use an indicator of the

[1]For authorative overviews of the recent literature on the identification and estimation of causal effects in economics cf. Angrist and Krueger (1999) and Card (1999).

[2]Further empirical studies on the returns to schooling in a LATE framework are e.g. Angrist (1990), Angrist and Krueger (1991), Angrist and Imbens (1995), and

presence of a college in the county of residence at schooling age as instrument. They argue that this "college proximity" might allow individuals from low income (probably even liquidity constrained) families to attend college who otherwise (i.e. if they would have had to move to another county in order to go to college) would not have done so.[3]

For Germany, Ichino and Winter-Ebmer (1999, 2000)[4] are the only authors we know of that provide LATE estimates of the returns to schooling.[5] IWE (1999) especially contrast estimates obtained on the basis of two different instruments: first, parental educational background, and second, an indicator of the father's serving in the military during World War II. Since parental education as assignment mechanism is likely to affect less able children from well-off families (IWE call them the "stupid rich") the corresponding IV estimate is interpreted as a lower bound of the returns to schooling in Germany. On the other hand, "childhood during war" and particularly "father in war" are considered an extreme form of liquidity constraints that might hinder highly talented children from poor families (the "smart poor") to continue schooling. For this reason, the authors interpret the IV estimate based on the "war instrument" as an upper bound of the returns to schooling in Germany.

In this paper, we extend the IWE (1999)-study in several ways: First, we replicate the Card- and Kling-studies for Germany, making use of an instrument similar to Card's "college proximity". Second, we allow for a variable treatment intensity and try to characterize both the affected subgroups as well as the response functions, and third, we compare results for 1985 with those for 1995, thus testing indirectly for changes in the

---

Kane and Rouse (1993).

[3]Cf. overview of IVE results by Card (1999).

[4]IWE (1999) draws on IWE (2000) where in addition to the instrument 'father in war' an indicator of the individual's having been in the age group 9-15 during the Second World War is used as an instrument. The latter paper is more specifically concerned with the long-run educational cost of World War II, while the first paper is more methodological and aims at providing evidence for heterogeneity in the returns to schooling.

[5]Lauer and Steiner (2000) do actually seem to follow a similar approach but they refrain from interpreting their estimates as local average treatment effects.

returns to schooling, in the instrument effectiveness, and in the response functions over time.

The results obtained on the basis of GSOEP data suggest that similar to the US results by Card and Kling, IV estimates of the returns to schooling are substantially higher than corresponding OLS estimates. We show that individuals from disadvantaged family backgrounds profit most from a better schooling infrastructure prevalent in urban areas.

The remainder of the paper is organized as follows: in the next section we present Becker's model of optimal schooling. In section 3, we present some basic evidence on the relationship between educational attainment and college proximity using regional and GSOEP data. In section 4, we present the GSOEP data and describe our sample. Section 5 discusses the use of IV estimation in our context and presents the instruments used in the empirical analysis. There, we also summarize the results of this analysis and discuss their interpretation. Section 6 concludes.

2 Theoretical considerations

In this section we shortly recall Becker's (1964) model of endogenous schooling in the version laid out by Card (1995b). It provides both the rationale for heterogeneous returns to schooling and the basis for the LATE interpretation of our results.

An individual maximizes

$$U(y, S) = \log y - \phi(S) \tag{1}$$

where y is average earnings per year, S is years of schooling and $\phi(\bullet)$ is the cost of schooling. An individual's opportunities are represented by $y = g(S)$.[6] The first order condition of the optimization problem is

[6]There is considerable discussion in the literature as to which variable best describes the theoretical concept of human capital. Griliches (1977) points out that years of schooling is rather one of the inputs of the human capital production process

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

Now, assume for simplicity that

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

and

$$\phi'(S) = \delta_i(S) = r_i + k_2 S \qquad (k_2 \geq 0) \tag{4}$$

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

$$\log y = b_i S - 0.5 k_1 S^2 \tag{5}$$

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 as shown in figure 1.

INTRODUCE FIGURE 1 ABOUT HERE

This is exactly what is exploited by the LATE-IV approach. A given instrument will affect different margins, i.e. different subpopulations at different schooling levels. As explained in detail in Angrist and Imbens (1995) we can hope to estimate only the average marginal return to schooling for a well-defined subgroup which is affected by the instrument. In the presence of heterogeneity, the notion of a unique return to schooling is hence nonsensical. In section 5.2 we are going to explain this in further detail.

We actually estimate the following system of equations:

than its outcome. To the extent that output measures are unavailable, years of schooling as a proxy for human capital is the best variable we can get to describe what is valued in the labor market.

$$
\begin{aligned}
y &= X\beta + S\gamma + \varepsilon && (6)\\
S &= X\delta + Z\alpha + \eta && (7)
\end{aligned}
$$

where Z is an instrument or set of instruments. For the LATE interpretation of IV to apply to the estimate of γ in (6), the conditions in Imbens and Angrist (1994) have to apply.[7] This approach thus makes a good out of the two main problems faced in a simple OLS regression of (6): the problem of self-selection into schooling and heterogeneity in returns to schooling. The main problem in empirical applications is, of course, to find an adequate instrument as an exogenous source of variation in education choices.

3 Educational outcomes and returns to schooling in Germany: Some background information

In this section, we present descriptive evidence based on regional data for some recent years (1996-1998).[8] We collected data about school completion rates and school infrastructure as well as some information about the state of the labor market at the level of counties (*Kreise*). These data show, in particular, a huge variation in completion rates across counties as well as a positive correlation between completion rates and schooling infrastructure.

[7]Further 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.

[8]The data had to be obtained from the single regional statistical offices (*Statistische Landesämter*) because to our knowledge no consistent educational data base exists at the national level.

3.1 Some background information using regional data

High school completion rates (*Abitur*) in Germany range from roughly 8% (in the *Südwestpfalz)* to 52% (in *Darmstadt*) of all school leavers across counties and hence show astonishingly strong regional variation. 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 plotted 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 (see figure 2). The availability of high schools is in fact seen to be highly correlated with high school completion rates.[9]

INTRODUCE FIGURE 2 ABOUT HERE

A higher average distance to the nearest high school is likely to increase the costs of education. Apart from the (time) opportunity costs of having to travel more, direct costs involve additional transport costs. All other costs do a priori not differ by distance to school. They might differ, however, across the various German regions (*Länder*) which are solely responsible for educational matters. Although there are generally no school fees neither for primary and secondary schools nor for universities, regulations regarding the public provision of books and other material used by students or subsidies for book purchases to low income families as well as regarding transportation subsidies for students do actually differ significantly across the various *Länder.* In many regions subsidies to either transport or book purchase are limited to students up to compulsory school age (i.e. 18 years old) or some other specific age (15 or 16 years old) and have to be borne fully by older students. Last but not least, the schooling years necessary for high school completion amount to 13 years in the West German *Länder* and Brandenburg as opposed to only 12 years in the remaining new German *Länder.* At university, the only fee to pay is for social security and health contributions.[10]

[9] Of course, this is not necessarily a causal relationship driven by the supply of high schools. It could also be that lower demand for higher education causes less supply by the state.

[10] In the later regressions, we try to capture differences in regulations across states

To sum up, using regional data we find lower high school completion rates in rural, less densely populated regions with a poorer schooling infrastructure. In addition, using microdata (GSOEP) we find lower high school completion rates for individuals who grew up in rural as opposed to urban areas (see table 1).[11]

INSERT TABLE 1 ABOUT HERE

Average years of schooling by agglomeration show a similar pattern as can be seen from table 2.

INSERT TABLE 2 ABOUT HERE

Do these differences tell us something about regional variations in the quality of schools and/or high school degrees (as often suggested in the political debate) or are they indicative of regionally varying opportunity costs related to longer schooling?

Our conjecture is that higher costs of education in regions with 'poor schooling infrastructure' reduce private investments in schooling, at least among children from relatively low-income/high discount rate families. This is also suggested by existing empirical studies on the returns to schooling based on instrumental variable estimation (Card (1995b), Kling (2000)). Card finds that the IV estimates of the earnings gain per year of additional schooling (10-14%) are substantially above the earnings gains estimated by a conventional OLS procedure (7.3%). Kling (2000), using Card's data, confirms Card's results and further characterizes the group of students affected by differences in place of childhood.

3.2 Previous studies

Previous results for Germany are based on simple OLS regressions of earnings on schooling. Using years 1984 and 1985 of the German Socioeconomic Panel (GSOEP), Wagner and Lorenz (1989) estimate returns to

by including a set of state dummies.

[11] Using regional data and defining agglomerations by quartiles of population density - which obviously do not coincide with the GSOEP classification - we observe a similar pattern. Going from the most densely to the least densely populated quartile, high schol completion rates in 1997 are 30.92, 23.07, 18.86, and 19.40 respectively.

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 are Ichino and Winter-Ebmer (1999, 2000) and Lauer and Steiner (2000). The former authors 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 subpopulations that are affected by the respective instruments. The latter authors not only estimate the returns to schooling using various estimation methods but also employ IV estimators on the basis of a whole 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 a LATE 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.

4 Data and descriptive evidence

The GSOEP is a representative longitudinal sample of the resident population containing socioeconomic information on private households. It was launched in 1984 with a sample of 12,245 respondents in 5,921 households in West Germany for the two randomly sampled subsamples of German nationals (i.e. people in private households where the head of household is not of either Turkish, Greek, Yugoslavian, Spanish, or Italian nationality) and of foreigners (i.e. people in private households where the head of household is of Turkish, Greek, Yugoslavian, Spanish, or Italian nationality).[12] In 1990, already before official unification, the first wave of the East German Subsample was added. It includes individuals in private households where the head of household is/was a citizen of the German Democratic Republic. In 1995, finally, a special sample of immigrants was for the first time interviewed.

4.1 Sample selection and descriptive statistics

We only keep the full-time employed in 1985 or 1995 who have no missing information on our variables of interest, in particular labor income and schooling. In tables 3, 4, and 5, we show descriptive statistics for the dependent variables, for schooling variables, and for exogenous variables. As for schooling variables, we present both average years of schooling along with degree information.

INSERT TABLE 3 ABOUT HERE

INSERT TABLE 4 ABOUT HERE

INSERT TABLE 5 ABOUT HERE

5 Instrumental Variables Estimation of the Returns to Education

5.1 Choice of instrument

Previous studies have used a broad range of instruments to establish causality in the returns to schooling (see Card, 1999) and the references therein). The choice of an instrument has several important aspects. First, econometrically speaking the instrument should fulfill the exclusion restriction, i.e. have an effect on earnings only via the schooling channel

[12]Note that the foreign sample consists mainly of people who came to Germany in the 1950s and 1960s as well as their descendants who have already assimilated to the native German population. In contrast, the immigrant sample (see below) includes foreigners who only recently came to Germany.

but no direct effect on earnings. Second, heterogeneity in marginal costs and benefits of schooling and therefore the absence of a unique return to schooling for the population as a whole can be exploited by choosing an instrument which describes a quasi-experiment of important policy interest. So, IV estimation is not just the solution to the econometric problem of possibly biased OLS estimates but allows to analyze interesting policy questions. On the basis of these two considerations, we choose our instrument 'place of childhood' which is similar to Card's (1995b) college proximity indicator. It has not yet been used for German data and allows us to address the question as to *who* profits *how* from differences in schooling infrastructure across different places of childhood.

The question on place of childhood in the GSOEP questionnaire is expressed as follows:

"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 ?"

In the sequel, we are going to use three different binary indicators based on this question: 'spent childhood in a city'(pc1), 'spent childhood in a city or big town' (pc2), and 'spent childhood in an urban area' (pc3), i.e. in a city, or in a small or big town. Table 6 shows the percentage of the sample with given instrument status.

INSERT TABLE 6 ABOUT HERE

5.2 Which Effect Can We Identify? The variable treatment approach to the estimation of returns to schooling

The IV estimate of the returns to schooling based on 'place of childhood' as an instrument identifies a causal effect for well-defined subpopulations and schooling levels. The implied natural experiment uses place of childhood as assignment to treatment (Z), the schooling level as treatment (S), and log(monthly earnings) as outcome (Y).

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 $j = 0, 1, 2, ..., J$. The objective is to uncover information about the distribution of $Y_j - Y_{j-1}$, which is the causal effect of the jth 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 (e.g. grown up in the countryside, in a small town, in a town, in a big city) for each individual. We define $S_Z \epsilon \{0, 1, 2, ..., J\}$ 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, indicating that the place of childhood was either in an urban area or in the countryside. S_1 then denotes the years of schooling that would be obtained by an individual growing up in an urban area, and S_0 is the years of schooling of the same individual if he or she grew up in the countryside. In the data, for each individual we observe the triple (Z, S, Y), where Z denotes the place of childhood, $S = S_Z = Z * S_1 + (1-Z) * S_0$ is years of completed schooling, and $Y = Y_S$ is earnings.[13] The main identifying assumption is the following

Assumption 1 (*Independence*)

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

In our case this requires that place of childhood has no effect on

[13] Note 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 observed place of childhood, $S_{obs} = S_{Z_{obs}} = Z_{obs} * S_1 + (1 - Z_{obs}) * S_0$ is observed years of completed schooling, and $Y_{obs} = Y_{S_{obs}}$ is observed earnings as a function of observed schooling.

earnings other than 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.*

Angrist and Imbens (1995) further show that for multivalued 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*

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

where

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

denotes weights and where the response function is defined as

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

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 *ACR weights* ω_j are proportional to the number of people who, because of the instrument, change their treatment from less than j units to j or more units. The *response function* $r(j)$ gives the average difference in the outcome for those who change their treatment from less than j units to j or more units. In the case of further covariates, the analysis is slightly more complicated and requires weighting by the conditional variance of Z.

In our example, IV generates an estimate of the average causal effect among individuals with different marginal benefits from schooling: First, different subgroups are affected by different instruments. Second, individuals in these subgroups are affected by the respective instrument in different ways. And third, the instrument may induce changes of behavior at different levels of schooling.

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

5.3 IV Estimation Results

We started by estimating an OLS regression of earnings on years of schooling controlling for sex, experience and tenure on the job polynomials, yielding estimates in the usual range of 6.7% and 6.6% for 1985 and 1995 respectively.

INSERT TABLE 7 ABOUT HERE

For the reasons given above, these estimates are probably not amenable to an interpretation as the causal effect of schooling on earnings. We therefore performed an IV estimation of the returns to education on the basis of the instruments suggested above. 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 whole 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 7 contains the IV estimation results for the various chosen instrumental variables. Further, first-stage t-statistics and partial R^2 measures are reported as a diagnostic tool following the suggestions of Bound et al. (1995) and Staiger and Stock (1997). In all cases, the instrument quality seems reasonable as suggested by these measures.

The returns estimated using either of these instruments are considerably higher than the OLS estimates. In 1985, the point estimates are 12.6%, 12.5% and 13.3% for the binary instruments 'spent childhood in a city', 'spent childhood in a city or big town', and 'spent childhood in an urban area'. A similar picture arises in the 1995 data. Throughout, the IV estimates are nearly double the size of the OLS estimates. In the light of the LATE framework, these results can be interpreted as the returns to education for those who acquired more education *because* they are living in an area with a good schooling infrastructure.

5.4 Internal validity of the instruments

To check the internal validity of the instrument for identification of the LATE parameter, we have to check the assumptions given in Angrist, Imbens and Rubin (1996).[14] Not all of these assumptions are in general

[14] AIR (1996) prove that the instrumental variables estimate of δ in the heterogenous *treatment* effect model has a *causal* interpretation as *local average treatment effect* under the following assumptions:

(1) Potential *outcomes* for each individual i are unrelated to the *treatment* status of other individuals. (*stable unit treatment value assumption* (*SUTVA*))

(2) Conditional on observables, the *assignment to treatment* is random. (*strongly ignorable assignment to treatment*)

(3) The *treatment* probability is a nontrivial and monotonous function of the instrument, i.e. $E[D_{i1} - D_{i0}] > 0$. (*strong monotonicity*)

rigorously testable. We can only argue and give corroborating evidence as we do in the sequel.

We can be quite confident that the *SUTVA assumption* is satisfied in our sample. It requires potential earnings to be unrelated to the amount of schooling taken by *other* individuals in the sample. This assumption is more likely to be violated in clustered samples.

Strongly ignorable assignment to treatment requires that after controlling for observable characteristics, unobservables like ability should be randomly distributed across different places of childhood. This assumption could be violated if parents endogenously choose to live in an urban area because of better schooling infrastructure. Most of this potential selection into places of living is probably controlled for by observables. In any case, geographical mobility in Germany is quite low by international standards. While Germany has 16 states and about 80 million inhabitants, the US have 51 states and about 250 million inhabitants, so average population per state is relatively similar, the US states being bigger in size, however. While in the US, 3% of the population move across state borders every year, in Germany only 1% of the population move across state borders.[15] Not only are mobility rates low anyway, but the reasons for moving are very unlikely to be related to schooling infrastructure as well. The GSOEP data contain a question on reasons for move. In 1997, respondents can give a maximum of three out of a list of 15 possible reasons. Overall, 8.6% of the movers give "other family reasons" (i.e. not divorce. marriage and leaving parent's home) as reason for move. If at all, families that move to give their kids access to a better schooling infrastructure might show up in this group. For families with kids under age 18 (i.e. those families for whom schooling infrastructure might play a roll), the percentage moving for "other family reasons" is even lower yet (5.2%), thus making "better schooling infrastructure" an

(4) The (unit-level) potential *outcome* variables depend on the assignment status Z_i only through the *treatment* status D_i, i.e. $(Y_{i0}, Y_{i1}) \perp Z_i \,|D_i$. (*exclusion restrictions*)

[15]Data come from the US Census Bureau website (http://www.census.gov/population/socdemo/migration/tab-a-1.txt) and from the website of the German National Statistical Office (http://www.statistik-bund.de/jahrbuch/jahrtab5.htm).

even more unlikely reason for moving. We conclude that our estimates are very unlikely to suffer from violation of the *strongly ignorable assignment to treatment* assumption.

Strong monotonicity compares again two counterfactual situations: 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 is that the cumulative density functions of schooling by instrument status do not cross. As we will show, this holds in our data and makes us confident that violation of the strong monotonicity assumption is not a serious issue here.

The *exclusion restriction* would be violated if there existed a direct effect of the suggested instrument on earnings, e.g. in the form of an 'urban wage premium'. We are in the fortunate situation to have some information about the current place of living. The GSOEP data contain both current state (*Bundesland*) of residence as well as the so-called Boustedt regions.[16] We find that by including these further controls, in 1985 the estimated returns to schooling do not change and in 1995 they even go slightly up. When controlling for state dummies, the coefficients on the Boustedt dummies are found to be statistically insignificant. We

[16]Boustedt (1970) classifies urban regions into seven categories, assigns the neighbouring communities of an urban center to four different sub-categories from "rural" to "urban center".

might therefore conclude that there is no violation of the exclusion restriction through an urban wage premium.

Another reason why the exclusion restriction might be violated is that school quality might vary by place of childhood. In this case, controlling for characteristics of the current place of living is not sufficient because people might have moved and the decision to take further schooling depended on their place of childhood and not on their current place of living. To see if this is a valid objection, we follow an idea similar to Card (1995b) and Kling (2000). They propose to define family background quartiles across which the returns to schooling will vary. 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. Translated to our setup, the idea is that our instrument is unlikely to affect individuals from higher family background quartiles because they have the necessary support by their family to pursue further education even if the respective schools are not nearby. So, using the instrument as such or using the instrument interacted with an indicator of low family background is the same, and gives us one more degree of freedom, namely allows us to control for the main effect of the 'place of childhood' indicator. We will further discuss the construction of the family background quartiles in the following section. There, we also use them to characterise the subgroup of compliers, so they serve a double purpose.

Let us shortly summarize the results of the estimation using the interacted instruments. We find that indeed the main effect of 'growing up in an urban area' is small in size and statistically insignificant.[17] The lower panel of table 7 shows that the point estimates are lower than the ones where we do not control for the main effect of 'growing up in an urban area', but that they are still considerably higher than the OLS estimates. On the basis of this evidence in favor of both the absence of urban wage premia and the validity of the exclusion restriction, we

[17] The coefficients on the main effect pc1 is 0.012 with a standard error of 0.019 in 1985, and 0.011 with a s.e. of 0.020 in 1995.

conclude that the returns to education for the subgroups of compliers, i.e. those individuals who only acquire more schooling when enjoying a good schooling infrastructure, are significantly and substantially higher than the simple OLS estimates. In the following section, we turn to the characterization of the subgroups affected by our instrument.

5.5 External validity of the instruments

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. We suggested above that the effect of schooling infrastructure is more important for children from less advantaged family backgrounds. We follow Card and Kling in defining family background quartiles in the following way: First, we perform a regression of years of schooling on the subgroup of people who spent their childhood in a rural area. Then, based on the parameter estimates obtained, we predict - for all individuals - their 'counterfactual schooling level if they had grown up in a rural area' and split the sample into four quartiles, from the lowest (fbq1) to highest (fbq4).

Table 8 presents some summary statistics on average years of schooling by instrument status and family background quartile for the years 1985 and 1995. Apart from the fact that average years of schooling are higher for those who grew up in urban areas, the table clearly shows that for those who have a higher predicted (counterfactual) schooling level, also actual schooling attainment is higher.

INSERT TABLE 8 ABOUT HERE

Table 9 further shows the distribution of family background and individual variables across these 'counterfactual schooling quartiles'.[18] There is no single individual in the lowest three family background quartiles whose father has a university degree. Conversely, there is virtually

[18]It is interesting to note that in the lowest background quartile, none of individuals report that either their father or mother graduated from high school.

no individual in the two highest background quartiles who has a father without a schooling degree. We also see that a higher percentage of those in the upper family background quartiles did actually grow up in a city.

INSERT TABLE 9 ABOUT HERE

The IV estimate of the returns to schooling can be interpreted as a weighted average of the potentially differing treatment effects across the four background quartiles, γ_q, with the weight given to each quartile q by the product of the proportion of the population in that subgroup (w_q) and the impact on schooling for that subgroup (ΔS_q). This allows us to write

$$\gamma = \sum_{q=1}^{4} \frac{w_q \Delta S_q \gamma_q}{\Delta S}$$

We give the weights w_q in table 10.

INSERT TABLE 10 ABOUT HERE

Table 11 shows the differences in schooling levels by instrument status for the population as a whole (ΔS).

INSERT TABLE 11 ABOUT HERE

Figures 3 and 4 further split up the information of table 10 by family background quartiles for 1985 and 1995 respectively. In 1985, the actual average education difference by instrumental status is much larger for the two lower background quartiles, supporting the suggestion of section 5.3 that instead of our indicator for 'growing up in an urban area' we can equally well use this indicator interacted with poor family background. This allows us to use the main effect of 'growing up in an urban area' in the estimation and thereby control for there being an urban wage premium. We already reported the results of this exercise in the previous subsection.

INSERT FIGURE 3 HERE

INSERT FIGURE 4 HERE

5.6 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 5 shows the difference in the CDFs for the 1985 sample using pc1 as an instrument.[19] It indicates that schooling infrastructure has its largest effect at 11 years of schooling. 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.

INSERT FIGURE 5 HERE

It is even more interesting to break down the response function by background quartiles. Figure 6 shows that the response function of the two lower background quartiles peaks at 10 years of schooling while the response of the two upper quartiles is concentrated among those with 13 or more 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.

INSERT FIGURE 6 HERE

From a policy point of view, this result suggests that the provision of schools beyond 10th grade, i.e. basically the provision of (senior) high schools (*Gymnasien*), can considerably increase the fraction of youths from disadvantaged backgrounds who obtain more schooling.

For 1995, the picture is slightly different. First, figure 7 suggests that for this later cohort, schooling infrastructure increased educational attainment at a later stage in educational careers.

INSERT FIGURE 7 HERE

Overall, in 1995 the response function is flatter and takes on lower values than in 1985. Second, breaking down by background quartiles, we

[19] Figures based on the instruments pc2 and pc3 show a similar pattern and are therefore not shown here.

find that the point of maximum response has moved to the right for all subgroups. Also has the fraction of the population in all subgroups who respond to our instrument decreased (see figure 8)

INSERT FIGURE 8 HERE

The fact that figures 5 and 7 display only non-negative values is equivalent to saying that the CDFs for $Z = 1$ and $Z = 0$ don't cross, a finding that supports the strong monotonicity assumption laid out in section 5.4.

To sum up, there seems to be a decreasing effect of our instrument on lower schooling levels and/or an increasing effect of the instrument on higher schooling levels. This also explains why returns to education seem to have decreased between 1985 and 1995.

6 Summary and conclusions

This study corroborates the general finding of other studies based on IV estimation that OLS estimates are downward biased. It confirms the empirical evidence that different instruments lead to different estimates of the schooling coefficient, underlining the fact that returns to schooling are heterogenous. Our estimates remain within the bounds given by IWE (1999). We find that individuals from 'poor family background' respond most strongly to the instrument 'place of childhood'. Their response is further most pronounced at low schooling levels whereas the response of individuals with 'rich family background' is most pronounced at higher schooling levels. Finally, this approach allows us to detect changes in the response function over time.

The temporal variation of returns to schooling operates through two different channels. First, temporal variation in the covariate weights leads to a reweighting of the returns for different subgroups. We conjecture that there is a decreasing fraction of compliers from a poor family background and/or an increasing fraction of compliers from a rich family background. Second, temporal variation of returns to schooling is also due to temporal variation in the response functions. There seems to be a decreasing effect of our instrument on lower schooling levels and/or an increasing effect of instrument on higher schooling levels.

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 regional government that has decided to devote a certain amount of money to the improvement of upper secondary schooling infrastructure.[20] It then 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 favourable 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 urbanisation, 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.

It 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.

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

References

[1] Angrist, Joshua D., 1990, "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records", *American Economic Review*, 80(3): 313-36.

[2] Angrist, Joshua D. 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*, Vol. 90, No. 430 (June 1995), 431-442.

[3] Angrist, J.D. and A.B. Krueger (1991), *Does Compulsory School Attendance Affect Schooling and Earnings,* Quarterly Journal of Economics 106: 979-1014.

[4] Angrist, J.D. and A.B. Krueger. "Empirical strategies in labor economics", Chapter 23 in: O. Ashenfelter and D. Card, eds., Handbook of Labor Economics, Vol. 3, North-Holland (1999).

[5] Angrist, Joshua D., Guido W. Imbens and Donald B. Rubin. "Identification of Causal Effects Using Instrumental Variables", *Journal of the American Statistical Association*, Vol. 91, No. 434 (June 1996), 444-455.

[6] Becker, Gary S. *Human Capital*, 1st ed. New York: Columbia University Press (for National Bureau of Economic Research), 1964.

[7] Becker, Gary S. *Human Capital and the Personal Distribution of Income*, Ann Arbor: University of Michigan Press, 1967.

[8] Bound, J., D.A. Jaeger and R.M. Baker (1995), Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak, *Journal of the American Statistical Association*, 90(430), 443-450

[9] Boustedt, Olaf (1970): Pendlerverkehr. In: *Handwörterbuch der Raumforschung und Raumordnung*, 2. Aufl., Bd. II, Sp. 2284. Hannover.

[10] Card, David. "Earnings, Schooling and Ability Revisited", in: Solomon Polachek, ed., *Research in Labor Economics*, Vol. 14 (1995a): 23-48.

[11] Card, David, 1995b, 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 (University of Toronto Press, Toronto), 201-222.

[12] Card, David. "The Causal Effect of Education on Earnings", in: O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Vol. 3A (1999): 1801-63.

[13] 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* vol. 20, no.1 (1999): 25-40

[14] 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* 43, no. 4-6 (1999) 889-901.

[15] Ichino, Andrea and Rudolf Winter-Ebmer. "The Long-Run Educational Cost of World War Two", mimeo, 2000.

[16] Imbens, Guido W. and Joshua D. Angrist. "Identification and Estimation of Local Average Treatment Effects", *Econometrica*, Vol. 62, No. 2 (March, 1994), 467-475.

[17] 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?", NBER Working Paper No. W4268, January 1993.

[18] Kling, Jeffrey, "Interpreting Instrumental Variables Estimates of the Returns to Schooling*", Journal of Business and Economic Statistics* (2001), forthcoming

[19] Lang, K. (1993). "Ability Bias, Discount Rate Bias and the Returns to Education", Boston University, mimeo.

[20] Lauer, Charlotte and Viktor Steiner. "Returns to Education in West Germany, An Empirical Assessment", Center for European Economic Research (ZEW), Mannheim, mimeo, 2000.

[21] Lorenz, Wilhelm, and Joachim Wagner (1993), "A Note on Returns to Human Capital in the Eighties: Evidence from Twelve Countries", *Jahrbücher für Nationalökonomik und Statistik*, 211(1-2): 60-71.

[22] Staiger, D. and J.H. Stock (1997), "Instrumental variables regression with weak instruments", *Econometrica*, 65, 557-586.

[23] Wagner, Joachim, and Wilhelm Lorenz (1989), Einkommensfunktionsschätzungen mit Längsschnittdaten für vollzeiterwerbstätige deutsche Männer. (Estimates of Earnings Functions for Full-Time Working German Men Using Longitudinal Data), *Konjunkturpolitik*, 35(1-2): 99-109.

Table 1: High school completion rates by agglomeration

agglomeration	GSOEP 1985	GSOEP 1995
city	18.49	25.00
big town	15.76	20.12
small town	10.94	18.70
in the countryside	8.58	12.97

Table 2: Years of schooling by agglomeration

agglomeration	GSOEP 1985	GSOEP 1995
city	12.01	12.42
big town	12.00	12.37
small town	11.28	11.97
in the countryside	11.10	11.63

Table 3: Summary statistics on outcome variables

Variable	N	Mean	Std.Dev	Min	Max
1985					
gross monthly income	4125	2983.33	1382.73	0	19000
net monthly income	4277	2029.74	959.23	0	13000
1995					
gross monthly income	3242	4524.49	3038.21	0	99999
net monthly income	3287	2984.71	1845.25	0	50000

Table 4: Summary statistics on education

Variable	N	Mean	Std.Dev	Min	Max
		1985			
years of schooling	4617	11.47	2.78	7	19.5
Hauptschule	4617	0.43	0.50	0	1
Realschule	4617	0.34	0.48	0	1
Fachhochschulreife	4617	0.03	0.18	0	1
Abitur	4617	0.09	0.29	0	1
Apprenticeship	4617	0.64	0.48	0	1
University degree	4617	0.10	0.30	0	1
		1995			
years of schooling	3457	12.00	2.87	7	19.5
Hauptschule	3455	0.40	0.49	0	1
Realschule	3455	0.34	0.47	0	1
Fachhochschulreife	3455	0.05	0.23	0	1
Abitur	3455	0.13	0.34	0	1
Apprenticeship	3457	0.69	0.46	0	1
University degree	3457	0.14	0.34	0	1

Source: GSOEP1985 and 1995 (100% version)

Table 5: Summary statistics on exogenous variables

Variable	N	Mean	Std.Dev	Min	Max
		1985			
sex	4617	0.29	0.45	0	1
age	4617	37.18	10.06	20	55
experience	4617	20.70	10.54	0	43
tenure	4606	9.74	8.06	0	56.6
changed place since childhood	3181	0.60	0.49	0	1
		1995			
sex	3457	0.31	0.46	0	1
age	3457	37.00	9.70	20	55
experience	3457	20.00	9.99	1	43
tenure	3457	9.74	8.77	0	41.3
changed place since childhood	2274	0.64	0.48	0	1

Source: GSOEP1985 and 1995 (100% version), own calculations

Table 6: Percentage of sample with given instrument status

	individual grew up in ...	1985	1995
pc1	... a city	21.90	19.09
pc2	... a city or a big town	36.28	33.27
pc3	... some urban area	58.74	54.12

Source: GSOEP 1985 (N=4617) and 1995 (N=3457), own calculations

Table 7: OLS and IV results

	1985	1995
OLS		
	6.71	6.54
	(6.31;7.11)	(6.12;6.96)
IVE: place of childhood		
city (pc1)	12.63	12.58
	(7.89;17.39)	(8.45;16.70)
1st stage t	6.737	7.087
partial R^2	0.0098	0.0141
city or big town (pc2)	12.49	9.67
	(9.00;15.98)	(6.91;12.45)
1st stage t	9.247	9.691
partial R^2	0.0183	0.0265
urban (pc3)	13.28	9.22
	(7.94;18.63)	(6.95;11.48)
1st stage t	6.131	11.387
partial R^2	0.0081	0.0362
IVE: place of childhood * poor family background		
pc1*(poor fbq)	10.65	11.17
	(6.76;14.55)	(7.57;14.77)
1st stage t	-7.848	-7.361
partial R^2	0.0075	0.0105
pc2*(poor fbq)	9.86	11.29
	(7.44;12.28)	(8.34;14.25)
1st stage t	-11.721	-9.512
partial R^2	0.0142	0.0176
pc3*(poor fbq)	9.33	9.68
	(7.58;11.08)	(7.77;11.60)
1st stage t	-15.795	-13.845
partial R^2	0.0219	0.0273

Table 8: Actual average years of schooling by instrument status and family background quartile

	fbq1	fbq2	fbq3	fbq4
1985				
City	10.79	11.45	12.23	13.36
City or big town	10.66	11.36	12.19	13.54
Urban area	10.44	11.05	12.11	13.28
1995				
City	11.46	11.64	12.48	14.13
City or big town	11.59	11.57	12.49	13.97
Urban area	11.18	11.44	12.49	13.88

Source: GSOEP 1985 and 1995, own calculations

Table 9: Distribution of family background and individual variables across those 'counterfactual schooling quartiles'

Background quartile	**1**	**2**	**3**	**4**	**Avg.**
Father's education					
High school degree	0.00	0.26	2.43	20.85	5.76
Professional school	0.25	2.35	4.59	23.34	7.59
University degree	0.00	0.00	0.00	16.68	4.07
No schooling degree	35.30	23.67	0.09	0.09	15.01
Mother's education					
High school degree	0.00	0.00	0.35	6.83	1.75
Professional school	0.25	1.65	3.47	12.87	4.48
University degree	0.00	0.09	0.09	3.11	0.80
No schooling degree	42.29	30.72	0.87	0.80	18.93
Parental presence	0.00	30.64	96.01	95.21	54.86
Place of childhood					
City	19.63	20.19	22.36	25.55	21.90
City or big town	33.87	33.86	35.44	42.15	36.28
Urban	56.11	59.18	56.93	62.91	58.74
Change of place	55.46	61.74	62.02	58.86	60.14
Female	48.19	25.76	30.59	11.36	29.22
Mean age	40.59	35.90	38.04	34.00	37.18

frequency of respective characteristic by family background quartile; definition of quartiles based on regression of schooling level on family background variables (and age) for individuals from rural background and subsequent predictions for all observations as 'counterfactual schooling level if individual had grown up in a rural area'

Table 10: Covariate weights

	fbq1	fbq2	fbq3	fbq4
	1985			
City	23.05	22.95	25.52	28.49
City or big town	24.00	23.22	24.42	28.36
Urban area	24.56	25.07	24.23	26.14
	1995			
City	16.21	26.06	27.88	29.85
City or big town	18.00	25.91	26.61	29.48
Urban area	19.99	25.28	26.19	28.54

Note: w_q is the fraction in each quartile

Source: GSOEP 1985 and 1995, own calculations

Table 11: Differences in schooling by instrument status

	1985	1985	1995	1995
$Z = ...$	0	1	0	1
City	11.32	12.04	11.85	12.59
City or big town	11.17	12.01	11.73	12.53
Urban area	11.10	11.74	11.56	12.36

Source: GSOEP 1985 and 1995, own calculations

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

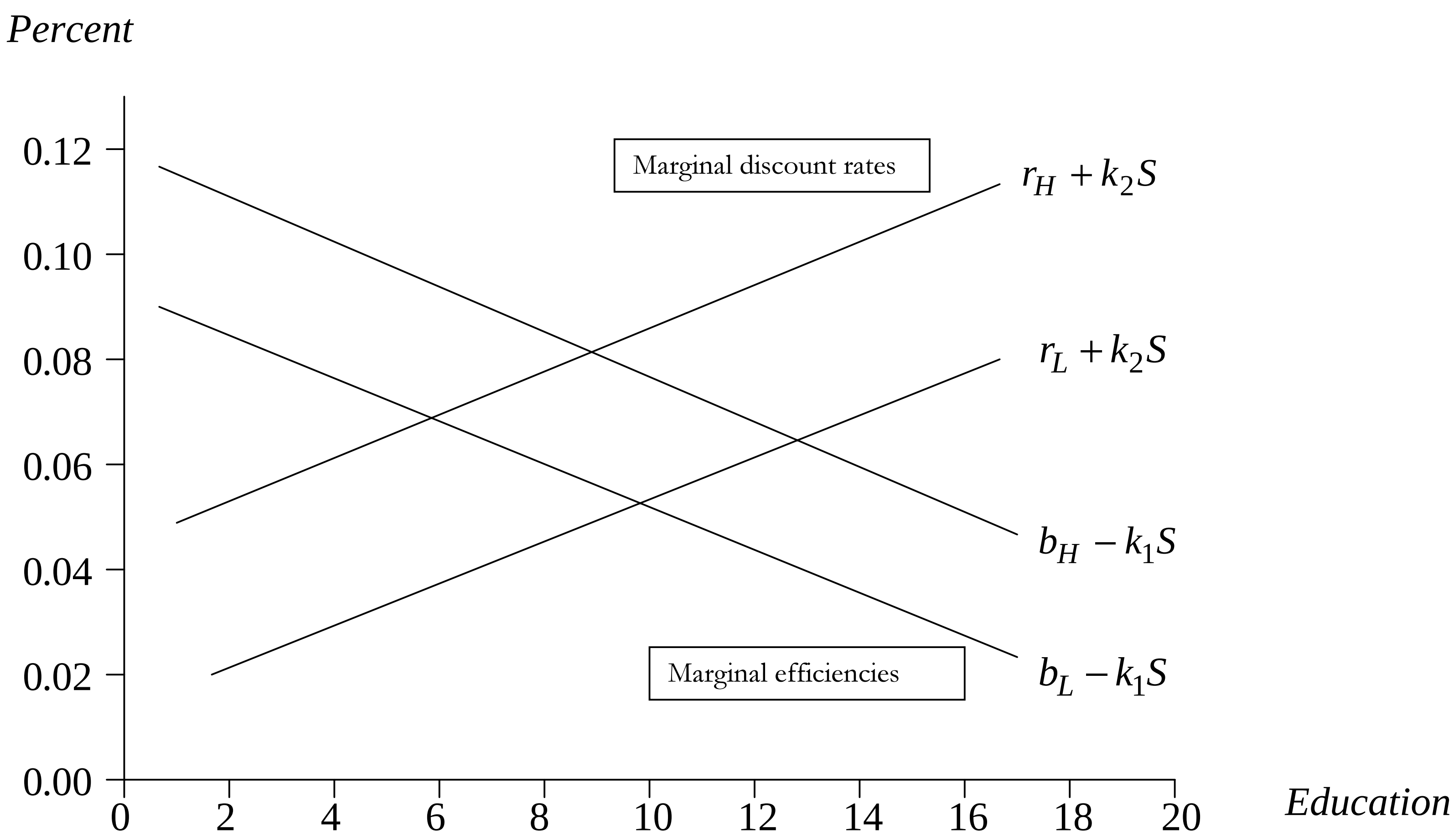
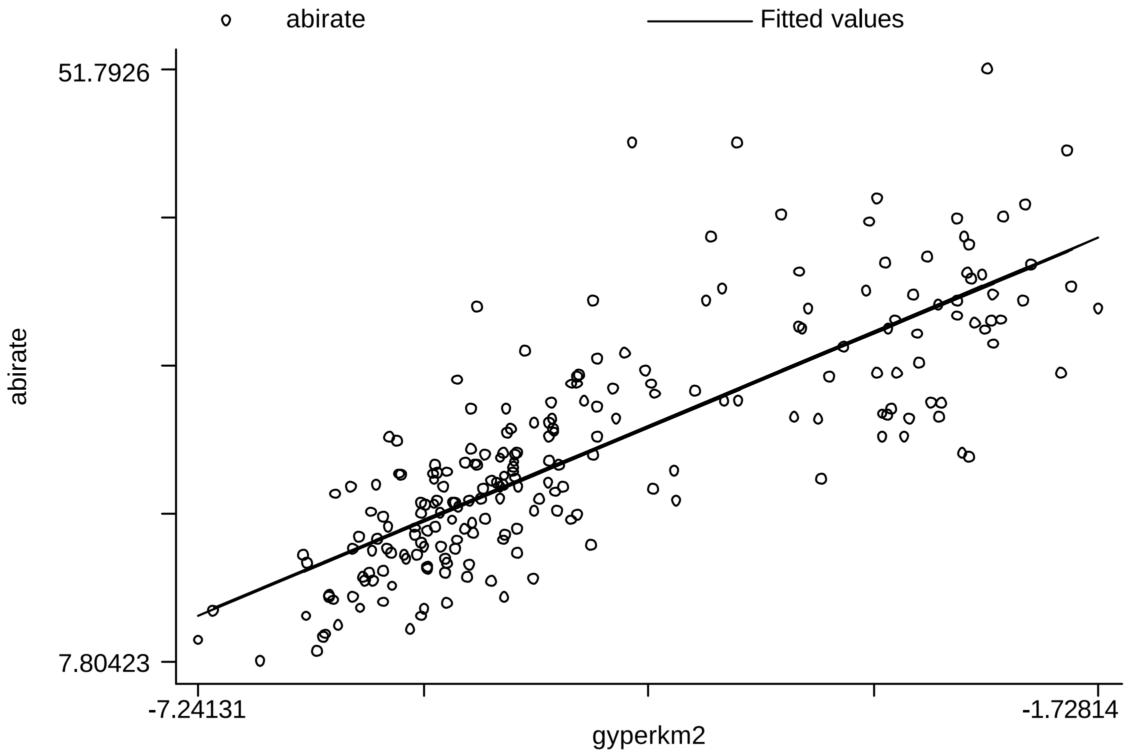


Figure 2: Educational attainment as a function of schooling infrastructure



Note: the figure plots the rate of school leavers having Abitur against the log of the number of "Gymnasium" per square kilometer
Source: data provided by the regional statistical offices (*Statistische Landesämter*) for the years 1996, 1997 or 1998 according to availability.

Figure 3: Actual average education difference by instrumental status using pc1 (1985 data)

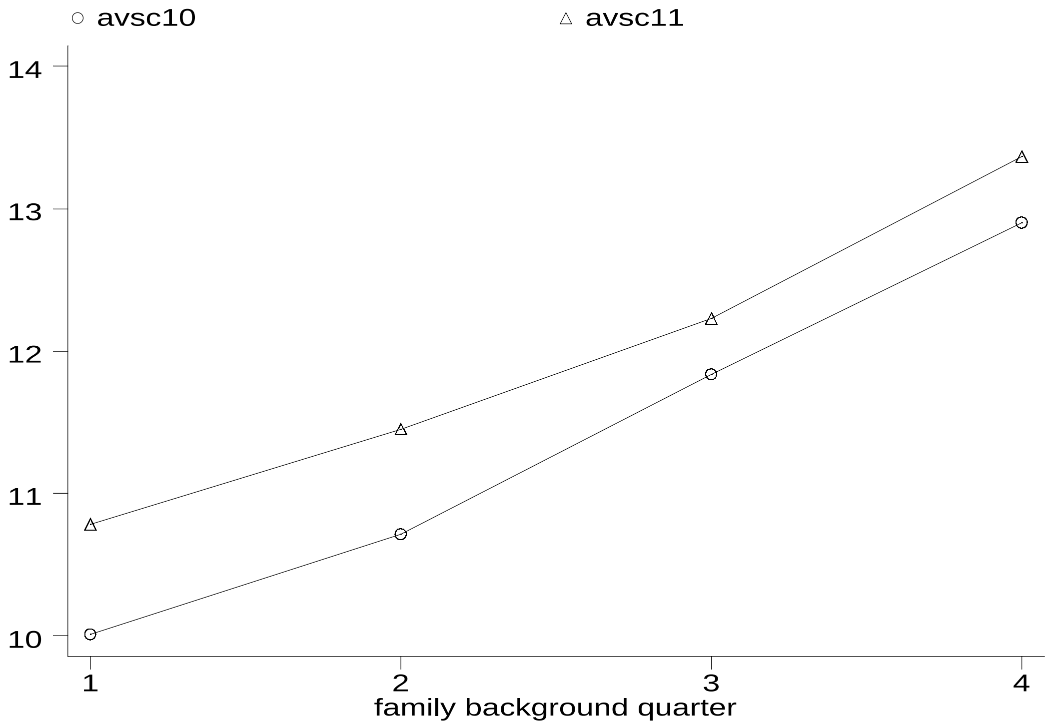


Figure 4: Actual average education difference by instrumental status using pc1 (1995 data)

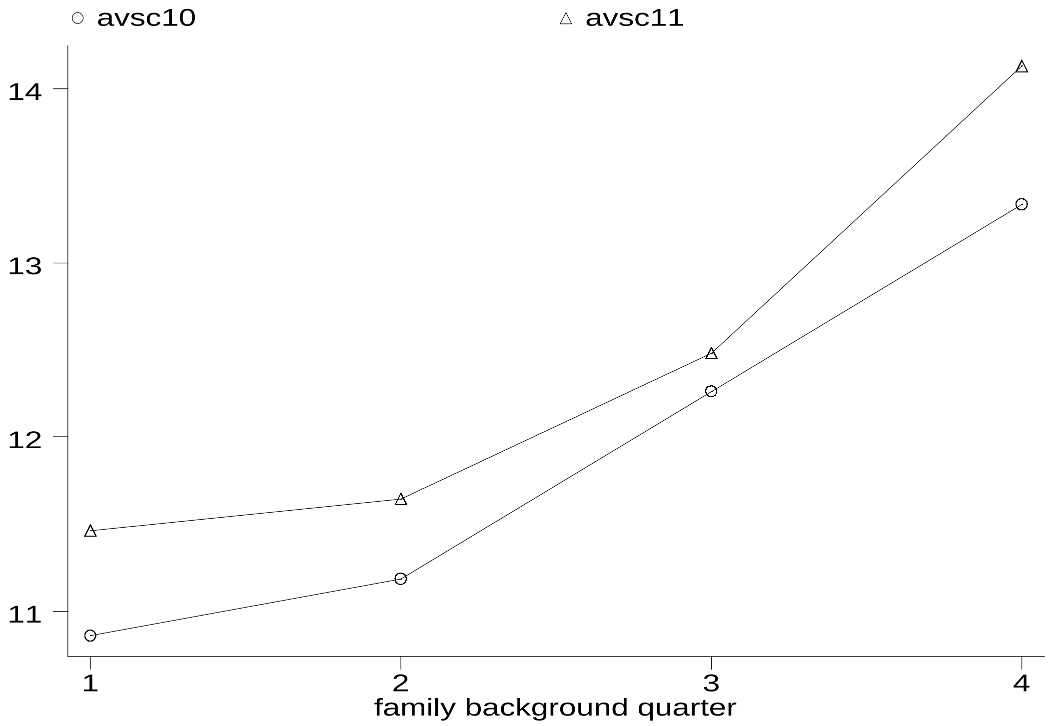
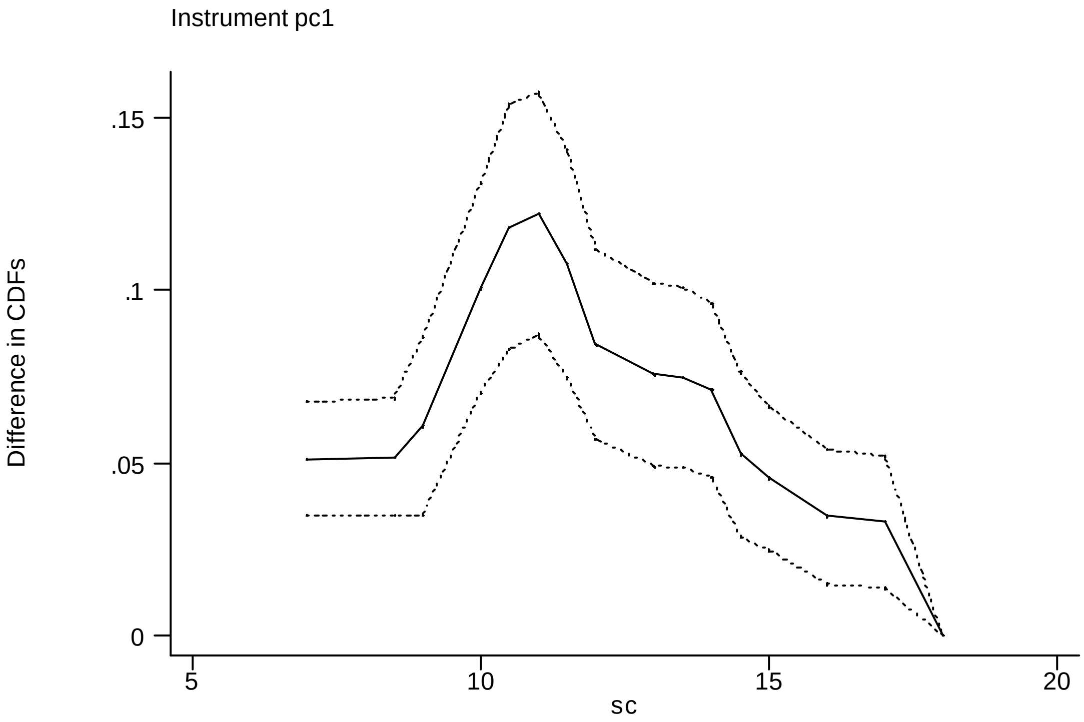
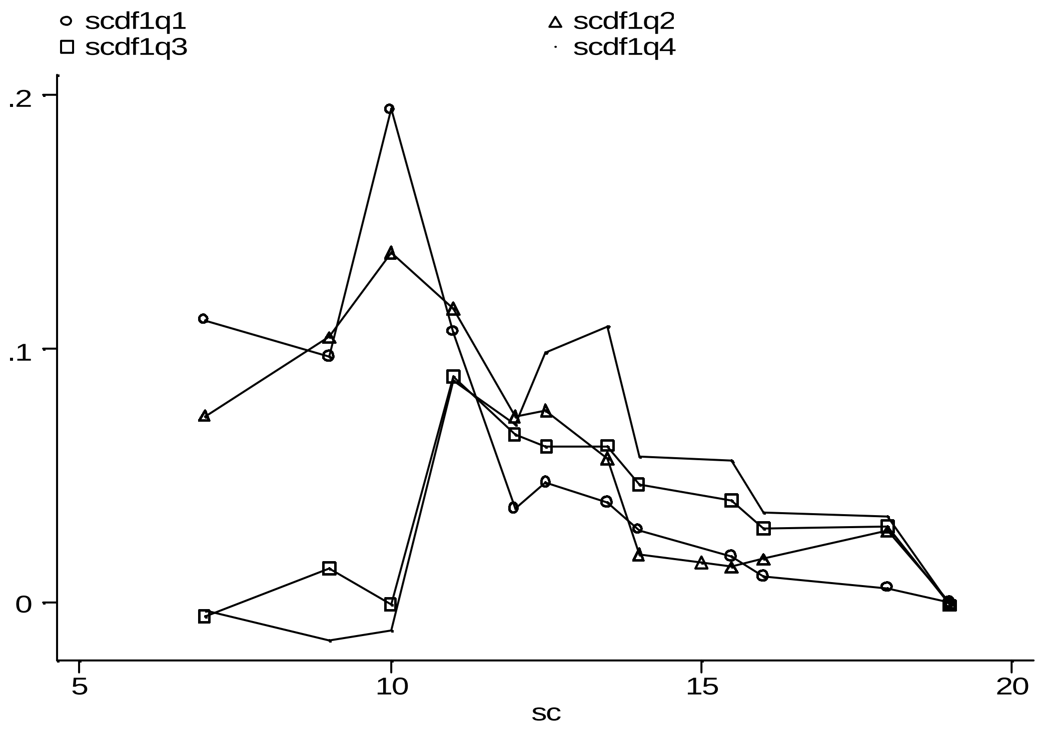


Figure 5: CDF difference using pc1 as an instrument (1985 data)



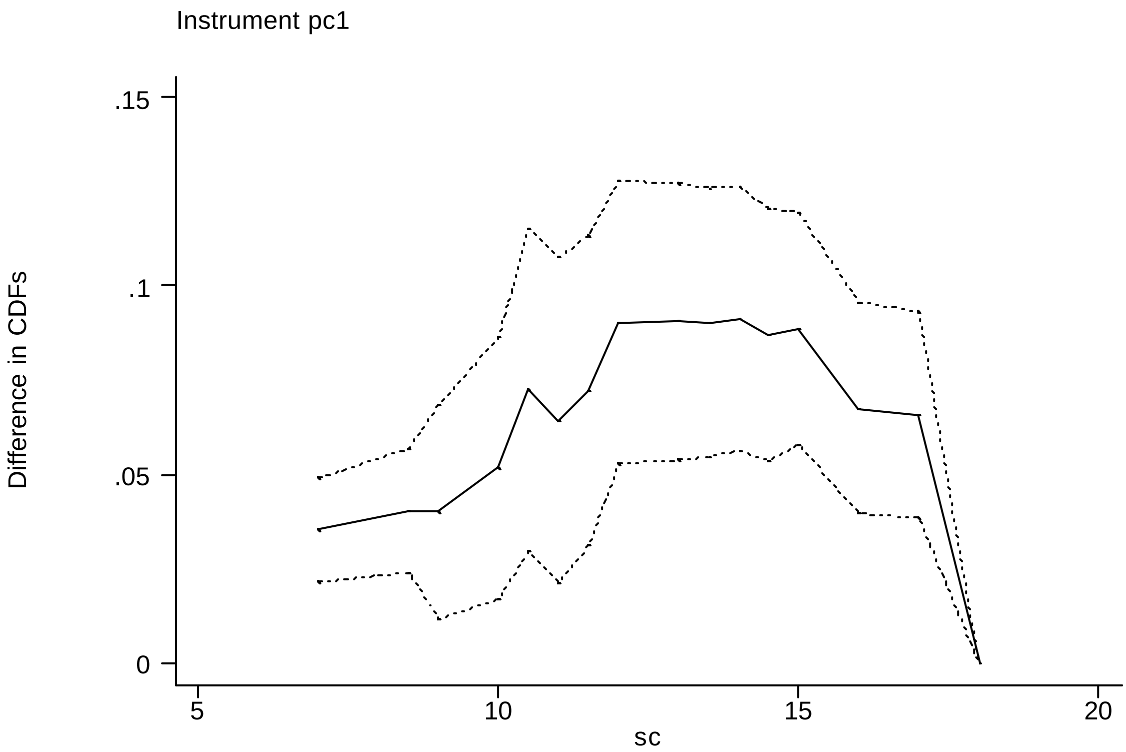
Note: the figure displays the fraction of the population who received at least one more year of schooling due to the instrument.
Calculated as the difference in the CDF: Pr(S<j|Z=0,X)-Pr(S<j|Z=1,X)
The 95% confidence bands are calculated using the conventional formula for a difference in proportions.

Figure 6: CDF difference by family background quartile using pc1 as an instrument (1985 data)



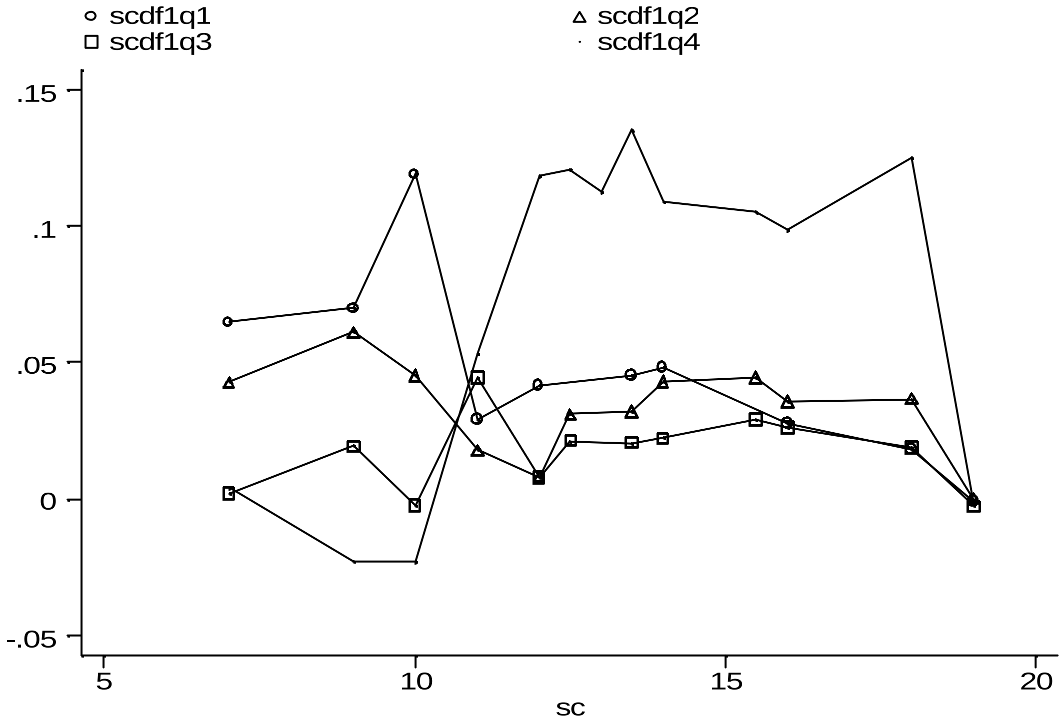
Note: the figure displays the fraction of the population who received at least one more year of schooling due to the instrument.
Calculated as the difference in the CDF: Pr(S<j|Z=0,X,Q)-Pr(S<j|Z=1,X,Q)

Figure 7: CDF difference using pc1 as an instrument (1995 data)



Note: the figure displays the fraction of the population who received at least one more year of schooling due to the instrument.
Calculated as the difference in the CDF: Pr(S<j|Z=0,X)-Pr(S<j|Z=1,X)
The 95% confidence bands are calculated using the conventional formula for a difference in proportions.

Figure 8: CDF difference by family background quartile using pc1 as an instrument (1995 data)



Note: the figure displays the fraction of the population who received at least one more year of schooling due to the instrument.
Calculated as the difference in the CDF: Pr(S<j|Z=0,X,Q)-Pr(S<j|Z=1,X,Q)