Daniel A. Kamhöfer, Hendrik Schmitz, and
Matthias Westphal

HETEROGENEITY IN MARGINAL NON-MONETARY RETURNS TO HIGHER EDUCATION

LIfBi

**LEIBNIZ INSTITUTE FOR
EDUCATIONAL TRAJECTORIES**

NEPS
National Educational Panel Study

Working Papers of the German National Educational Panel Study (NEPS)
at the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg

The NEPS Working Papers publish articles, expertises, and findings related to the German National Educational Panel Study (NEPS).
The NEPS Working Papers are edited by a board of researchers representing the wide range of disciplines covered by NEPS. The series started in 2011.

Papers appear in this series as work in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character.

Any opinions expressed in this series are those of the author(s) and not those of the NEPS Consortium.

The NEPS Working Papers are available at
https://www.neps-data.de/projektübersicht/publikationen/nepsworkingpapers

Contact: German National Educational Panel Study (NEPS) – Leibniz Institute for Educational Trajectories – Wilhelmsplatz 3 – 96047 Bamberg – Germany – contact@lifbi.de

Heterogeneity in Marginal Non-monetary Returns to Higher Education

Daniel A. Kamhöfer

University of Duisburg-Essen and
CINCH-Health Economics Research Center, Essen

Hendrik Schmitz

University of Paderborn and
CINCH-Health Economics Research Center, Essen

Matthias Westphal

Ruhr Graduate School in Economics,
University of Duisburg-Essen and
CINCH-Health Economics Research Center, Essen

Email addresses of the authors:

daniel.kamhoefer@uni-due.de (Kamhöfer)

hendrik.schmitz@uni-paderborn.de (Schmitz)

matthias.westphal@uni-due.de (Westphal)

Supplementary materials to this paper may be found under the following link:

https://www.dropbox.com/sh/wvqule32bvh8t89/AADou8fGmGcZtQ6fXL4RKT_ca?dl=0

Heterogeneity in Marginal Non-monetary Returns to Higher Education

December 2015

Abstract

In this paper we estimate the effects of college education on cognitive abilities and health exploiting exogenous variation in college availability and student loan regulations. By means of semiparametric local instrumental variables techniques we estimate marginal treatment effects in an environment of essential heterogeneity. The results suggest heterogeneous but always positive effects on cognitive skills and homogeneously positive effects for all health outcomes but mental health, where the effects are around zero throughout. We find that likely mechanisms of positive physical health returns are effects of college education on physically demanding activities on the job and health behavior such as smoking and drinking while mentally more demanding jobs might explain the skill returns.

Keywords: Returns to higher education, cognitive abilities, health, marginal treatment effect

JEL Classification: *C31, H52, I12, I21*

1 Introduction

"The whole world is going to university – Is it worth it?" *The Economist*'s headline read in March 2015.[1] After decades of economic research there is still a need for investigation of the returns to higher education. According to recent surveys by Barrow and Malamud (2015) and Oreopoulos and Petronijevic (2013) convincing causal evidence for positive earnings returns to college is rare. Yet, they summarize positive average monetary effects of higher education in most studies. Much less work has been done on non-monetary effects of higher education and in institutional settings outside the US. In this paper we estimate the effects of college education on two non-pecuniary outcomes, namely cognitive abilities and health using a rich and representative German data set.

Cognitive abilities and health belong to the most important non-monetary determinants of individual well-being. Moreover, the stock of both factors also influences the economy as a whole (see, among many others, Heckman et al., 1999, and Cawley et al., 2001, for cognitive abilities and Acemoglu and Johnson, 2007, Cervellati and Sunde, 2005, and Costa, 2015, for health). Yet, non-monetary returns to college education are not fully understood so far (Oreopoulos and Salvanes, 2011). Despite this lack of knowledge we furthermore jointly focus on health and cognitive skills since they are closely interrelated and – most arguably – complementary for the production of health, skills, and human capital. The existing literature studies either one separately. Only a few studies analyze the health returns to college education. Grimard and Parent (2007) and de Walque (2007) use avoidance behavior of potential US draftees to instrument the college decision. For the complying subpopulation, college education reduces the probability that a non-smoker starts smoking and increases the probability that a smoker ceases smoking. Currie and Moretti (2003) use college proximity as instrument and find that women with college education are less likely to smoke and that they are more likely to bear healthier children. To the best of our knowledge there is no study so far that estimates the effect of college education on cognitive abilities.[2]

We use a slightly modified version of the marginal treatment effect approach introduced and forwarded by Heckman and Vytlacil (2005). The main feature of this approach is to explicitly model the choice for education, thus turning back from a mere statistical view of exploiting exogenous variation in education to identify casual effects towards a description of the behavior of economic agents. Translated into our research question, the MTE is the effect of education on cognitive abilities and health for individuals at the margin of taking higher education. The MTE can be used to generate all conventional treatment

---

[1]*The Economist*, edition March 28th to April 3rd 2015.
[2]Glymour et al. (2008), Banks and Mazzonna (2012), Schneeweis et al. (2014), and Kamhöfer and Schmitz (2015) analyze effects of secondary schooling on cognitive abilities.

parameters, such as the average treatment effect (ATE). On top of this, the distribution of marginal effects is also informative in its own right: the MTE does not just reveal effect heterogeneity but also some of its underlying structure (for instance, selection into gains). This may be an important property that the local average treatment effect – as identified by conventional two stage least squares methods – would miss.

Our outcome variables (cognitive ability[3] and health) are measured between 2010 and 2012 while individuals take up college education between 1958 and 1990 in our sample. Thus, we analyze long-run effects. We use two instruments in order to receive a sufficient amount of exogenous variation in the college decision. The first one is the availability of college places in the area of residence at the time of the secondary school graduation. We exploit arguably exogenous expansions of college capacities in the German "educational expansion" between the 1960s and 1980s that generates variation of colleges over time and regions. As a second instrument we use changes in student loan regulations in Germany.

By deriving the entire distribution of treatment effects over the support of the probability of college attendance, this paper contributes to the literature mainly in two important ways. First, this is the first study that analyzes the effect of college education on cognitive abilities. Therefore, we add an important mechanism that helps to explain potential earnings returns to college education. Second, by going beyond the point estimate of the LATE, we provide a more comprehensive picture in an environment of essential heterogeneity. The results suggest heterogeneous but always positive effects for all measures of cognitive abilities. Health returns are positive and homogenous except for a zero effect in the case of mental health. We also study likely mechanisms of positive physical health returns and find effects of college education on physically demanding activities on the job and health behavior such as smoking and drinking. Mentally more demanding jobs might explain the skill returns.

The paper is organized as follows. Section 2 briefly introduces the German educational system and describes the exogeneous variation we exploit. Section 3 outlines the empirical approach. Section 4 presents the data. The main results are reported in Section 5 while Section 6 addresses some of its potential underlying pathways. Section 7 concludes.

---

[3]See Section 4 for a detailed definition of cognitive abilities. We use the terms "cognitive abilities", "cognitive skills", and "skills" interchangeably.

2 Institutional background and changes over time

2.1 The German higher educational system

After graduating from secondary school, adolescents in Germany either enroll into higher education or start an apprenticeship. The latter is part-time training-on-the-job and part-time schooling. This vocational training usually takes three years and individuals often enter the firm (or another firm in the sector) as a fulltime employee afterwards. To be eligible for higher education in Germany, individuals need a university entrance degree. In the years under review, only academic secondary schools (*Gymnasien*) with 13 years of schooling in total award this degree (*Abitur*). Although the tracking from elementary schools to secondary schools takes place rather early at the age of 10, students can switch secondary school tracks in every grade. It is also possible to enroll into academic schools after graduating from basic or intermediate schools in order to receive a university entrance degree.

In Germany, mainly two institutions offer higher education: universities/colleges[4] and universities of applied science (*Fachhochschulen*). The regular time to receive the in those days common *Diplom* degree (master's equivalent) was 4.5 years at both institutions. Colleges are usually large institutions that offer degrees in various subjects. Moreover, colleges also offer the opportunity to earn a doctoral degree. The other type of higher educational institutions, universities of applied science, are usually smaller than colleges and often specialized in one field of study (e.g., business schools). Moreover, universities of applied science have a less theoretical curriculum and teaching structure that is similar to schools. Nearly all institutions of higher education in Germany charge no tuition fees. However, students have to cover their own costs of living. On the other hand, their peers in apprenticeship training earn a small salary. Possible budget constrains and the availability of financial aid are likely determinants of the decision to enroll into higher education.

2.2 Exogenous variation in college education over time

While the higher educational system as described in Section 2.1 did not change in the years under review, the educational accessibility (in terms of mere quantity but also their distribution within Germany) as well as financial affordability of tertiary education changed significantly, providing us with two sources of exogenous variation. This

---

[4]We use the words university and college as synonyms to refer to German *Universitäten* and closely-related institutions like technical universities (*Technische Universitäten*/*Technische Hochschulen*), an institutional type that combine features of colleges and universities applied science (*Gesamthochschulen*) and universities of the armed forces (*Bundeswehruniversitäten*/*Bundeswehrhochschulen*).

so called "educational expansion" falls well into the period of study (1958-1990). Within this period relaxed credit constraints and the shrinking transaction costs of studying may have changed incentives and the mere presence of a new or growing college could also have nudged individuals towards higher education that otherwise would not have studied. In this paper, we consider three processes in order to quantify the educational expansion. The first is the openings of new colleges, the second process is the extension in capacity of all colleges (we refer to both as college availability) and the third the introduction of a student loan program (BAfoeG) in Germany.

College availability

College availability as an instrument for higher education was introduced to the literature by Card (1995) and has frequently been employed since then (e.g., Currie and Moretti, 2003), also to estimate the MTE (e.g., Carneiro et al., 2011, and Nybom, 2014). We exploit the rapid increase in the number of new colleges and in the number of available spots to study as exogenous variation in the college decision.

Factors that have driven the increase in the number of colleges and their size can be summarized into four groups: (i) The large majority of the population had a low level of education. This did not only result from the WWII but also from the "anti-intellectualism" (Picht, 1964, p.66) in the Third Reich. (ii) An increase in the number of academic secondary schools at the same time (as analyzed in Kamhöfer and Schmitz, 2015, and Jürges et al., 2011, for instance) qualified a larger share of school graduates to enroll into higher education (Bartz, 2007). (iii) A change in production technologies led to an increase in firm's demand for high-skilled workers – especially, given the low level of educational participation (Weisser, 2005). (iv) Political decision makers were afraid that "without an increase in the number of skilled graduates the West German economy would not be able to compete with communist rivals" (Jürges et al., 2011, p.846, in reference to Picht, 1964).

Although these reasons (maybe except for the firm's demand for more educated workers) affected all of the 11 West German federal states – that are in charge of educational policy – in the same way, the measures taken and the timing of actions differ widely between states. Because of local politics (e.g., the balancing of regional interests and avoiding clusters of colleges) there was also a large amount of variation in college openings within the federal states, see the Supplementary Materials A to the paper for a much more detailed description.

Between 1958 (the earliest secondary school graduation year in our sample) and 1990 the number of colleges in Germany doubled from 33 to 66.[5] Since we use birth cohort and

---

[5]All data are taken from the German Statistical Yearbooks, 1959-1991, see German Federal Statistical Office (1991). We only use colleges and no other higher educational institutes described in Section 2 (e.g.,

district fixed effects as well as state-specific time trends in the empirical approach, the instrument measures regional differences in the variation of increased opportunities to receive college education (see Figure A1 in the Appendix for the spatial variation over the time). In particular, the opening of new colleges introduces discrete discontinuities in choices sets that cannot be exploited using cross-sectional data (as most other research on college openings does). Given the rich set of control variables (including the socioeconomic environment before the college decision and various sets of fixed effects, see Section 4) and the political process of college opening decisions, it should not be an issue that regions which traditionally had a college differ from those without a college, e.g., in terms of local economic conditions or other conditions potentially correlated with health or abilities.

The same development described above and discussed in the Supplementary Materials A led to an increase in the size of existing colleges and, therefore, in the number of available spots to study as well. The average number of students per college was 5,013 in 1958 and 15,438 in 1990. Of the 33 colleges in 1958, 30 still existed in 1990 and had an average size of 23,099 students. The total number of students increased from 155,000 in 1958 to 1 million in 1990. Figure 1 shows the trends in college openings and enrolled students (weighted by the number of inhabitants) per federal state. While the actual numbers used in the regressions vary on the much smaller district level, the state level figures simplify the visualization of the pattern.

Details on how we exploit the variation in college availability in the empirical specification are discussed in Section 4.4 after presenting the data.

Student loan regulations

Another policy intervention that introduced exogenous variation in the college decision in Germany – and which we use as an instrument – is the introduction of a large-scale student loan program known as BAfoeG (named after the Federal Training Assistant Act, *Bundesausbildungsförderungsgesetz*, that made the program federal law). The BAfoeG loan scheme (that still exists today) was introduced in 1971 and gave for the first time in the German educational history every student a legal claim of receiving a minimum financial amount that covers the basic costs of living. See German Federal Statistical Office (2009) for a history of BAfoeG regulations. Supplementary Materials A state the exact formula for calculating BAfoeG payoffs as well as an example. In principle, individuals whose family's income falls below some threshold are eligible for student loans while

---

universities of applied science). Administrative data on openings and the number of students are not available for other institutions than colleges. However, since other higher educational institutions are small in size and highly specialized, they should be less relevant for the higher education decision and, thus, neglecting them should not affect the results.

Figure 1: Number of colleges and students over the time
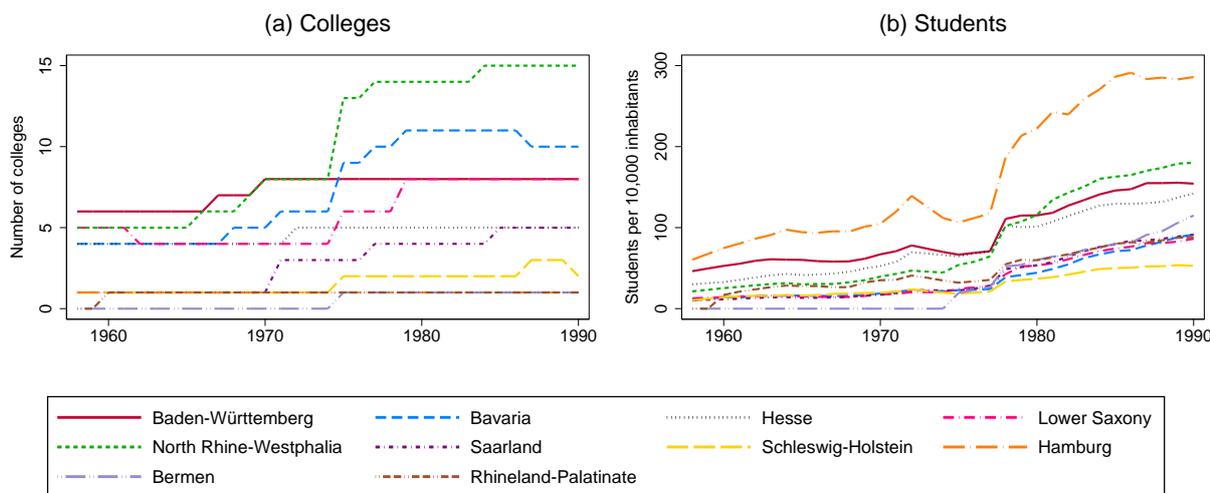
Notes: Own illustration. College opening and size information are taken from the German Statistical Yearbooks 1959–1991 (German Federal Statistical Office, 1991). Yearly information on the district-specific population size is based on personal correspondence with the statistical offices of the federal states. Data are available on request.

those above the threshold are not. While changes in the eligibility threshold over time (see Supplementary Materials A) affect all individuals who face the decision to study in the same year, the instrument also exploits discontinuities around eligibility thresholds (see below). The identifying assumption is analogous: given observable characteristics, whether an individual is on the left-hand side or the right-hand side of the threshold value is independent of her or his cognitive abilities and health. Given the discontinuous character around the thresholds, we consider this plausible.

Figure 2 illustrates how the BAfoeG payment scheme (right axis) and the probability to study (left axis) depends on parental income (measured in the deviation from/distance to the eligibility threshold on the horizontal axis). BAfoeG eligibility generally depends on the difference between the family's net income and its financial needs that is reflected in an income threshold. The threshold was adjusted every year and varies with the financial needs of the family, e.g., the number of siblings living in the household. The horizontal axis in Figure 2 shows the difference between the eligibility threshold and (imputed) household net income.[6] We refer to this difference as surplus income because the government expects families to use this part of their income to support their offspring's higher education. When the family's surplus income is not sufficient to cover the student's basic

---

[6]The NEPS data used in this study (see Section 4 for details on the data set) do not include information on the family's income during the respondents' youth. However, they include a variety of information on the living conditions at the age of 15, see Table A1 (e.g., single parent, number of siblings as well as parental education and occupation). Because this information and the household income are also included in the German Socio-economic Panel Study (SOEP, see Wagner et al., 2007), we use SOEP waves 1984-1990 in order to impute the family income for NEPS respondents aged 15 in 1971-1990. Since we only use the imputed family income for the illustrations in Figures 2 and S1 as well as Table S2, the other results are not affected even if there is a measurement error in the imputed family income.

Figure 2: The BAfoeG benefit scheme and propensity to study by eligibility
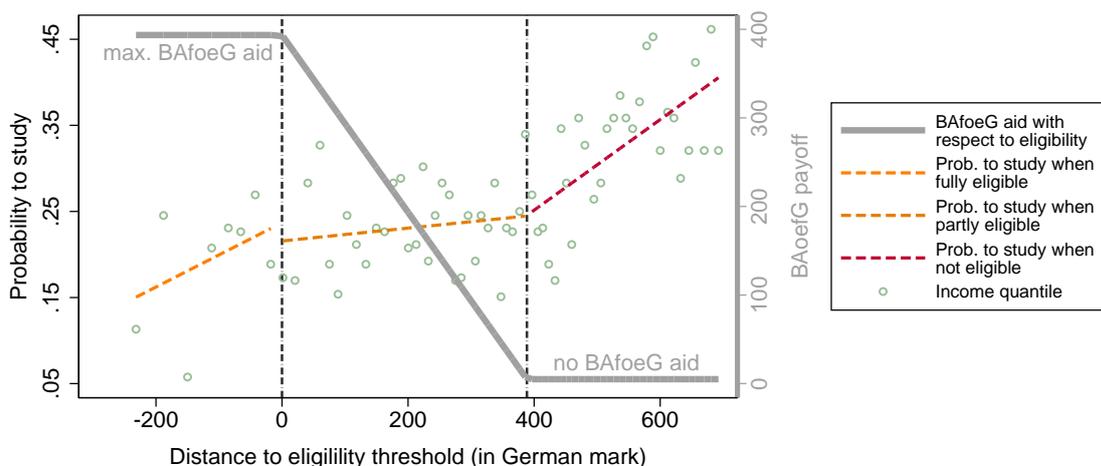
Notes: Own illustration based on NEPS-Starting Cohort 6 data, SOEP data and public records on BAfoeG eligibility taken from German Federal Government (1992). The horizontal axis gives the difference between the individual income threshold for eligibility according to the family characteristics the time an individual is aged 15 in the NEPS and the (SOEP imputed) family income in German mark 1971. The dashed black lines on the horizontal axis at 0 and 389 represent discontinuities in the BAfoeG payoff scheme as depicted by the gray line. A negative value on the horizontal axis indicates that an individual is eligible for financial aid because the family income is below the threshold. The right hand side axis gives the amount of financial aid (in German mark in prices of 1971) the individual is eligible for given the surplus income on the horizontal axis. Although the maximal value of financial aid varies yearly, we use the 1971-1990 average in order to simplify the visualization. The left axis gives the probability of higher education. The green circles depict the probability to study by the quantile of the family's income surplus. Because we aim at exploiting discontinuities in the relationship between income surplus and the probability to study, we analyze the relationship in the neighbourhood of the thresholds (+/-300 German mark). The orange and red lines represent the fitted values of a linear spline regression of the probability to study on the family's income surplus.

costs of living, individuals are (partly) eligible for BAfoeG financial aid. Therefore, the relationship between surplus income (or eligibility) on the horizontal axis and BAfoeG payoff on the left-hand side axis exhibits two thresholds (dashed black lines) that may cause a discontinuity. The first threshold is at a surplus income of zero. I.e., the family income equals the amount of money the family needs to cover there costs of living when the individual does not go to college. If the family income is on the left-hand side from this zero threshold, the government does not expect the family to make a financial contribution to their offspring's higher education, instead the government pays the maximal financial aid. The maximal financial aid should cover the basic costs of living for the student. Although it was yearly adjusted, we take the average value over all years in this graph, that is 389 German mark[7], to simplify the graphical relationship. If the family has a surplus income on the right-hand side of the zero threshold, the family is expected to spend the surplus for their child's higher education. In this case, the government only pays the difference between the amount the family can contribute and the amount the student needs (389 German mark). If the family has a surplus income of more than 389 German mark, i.e., lays on the right-hand side of the zero threshold, it is expected to cover the student's costs of higher education entirely. Between the zero threshold and the 389

---

[7]All values are in prices of 1971. In today's terms 389 German mark equal € 645 or $707.

threshold, one German mark less family income leads to an increase in the BAfoeG payoff of one German mark (as depicted by the gray line).

The second purpose of Figure 2 is to provide evidence how BAfoeG payments change the effect of family income and the probability to study. The green circles plot the probability to study by the quantile of the surplus income. Because we aim at exploiting discontinuities in the relationship between income surplus and the probability to study, we analyze the relationship in the neighbourhood of the thresholds (+/-300 German mark). The orange and red lines state the fitted values of regressing the probability to study on the surplus income within the BAfoeG eligibility bins. Within the first bin (left-hand side of the zero threshold), an increase in the (still negative) surplus income of 100 German mark is associated with an increase in the probability to study of 3.7 percentage points. While the effect is even higher in the no-BAfoeG area (right-hand side of the 389 threshold), 5.3 percentage points, the effect is only 0.07 percentage points for surplus income increases between the thresholds. This effect is close to the zero effect we would expect because an increase in the family's surplus income leads to an one-to-one decrease in the BAfoeG payoff – as indicated by the gray line. The significant flatter relationship between surplus income and the probability to study when BAfoeG payments matter ensures us that BAfoeG is an important determinant to the decision to go to college.

However, it might be noteworthy to mention that, for this figure, we use the imputed family income at the age of 15 that is based on socioeconomic factors, not the actual family income. Thus, one has to be careful in projecting the mechanisms of BAfoeG regulations. This figure nonetheless provides illustrative evidence that BAfoeG regulations could be effective and what source of exogenous variation we are aiming to exploit.

3 Empirical Strategy

Our estimation framework widely builds on Heckman and Vytlacil (2005) and Carneiro et al. (2011). Derivations and in-depth discussion of most issues can be found there. We start with the potential outcome model, where

$$
\text{Potential outcome}_i = \begin{cases} Y_i^1 & \text{with treatment} \\ Y_i^0 & \text{without treatment.} \end{cases}
$$

The observed outcome Y_i either equals Y_i^1 in case an individual received a treatment – which is college education here – or Y_i^0 in the absence of treatment. Obviously, treatment

participation is voluntary, rendering a treatment dummy D_i in a simple linear regression endogenous.

In the marginal treatment effect framework, this is explicitly modelled by using a choice equation, that is, we specify the following latent index model:

$$
\begin{aligned}
Y^1 &= X'\beta_1 + U_1 & \text{(1)} \\
Y^0 &= X'\beta_0 + U_0 & \text{(2)} \\
D^* &= Z'\delta - V \qquad \text{where } D = \mathbf{1}[D^* \geq 0] = \mathbf{1}[Z'\delta \geq V] & \text{(3)}
\end{aligned}
$$

The vector X contains observable, and U_1, U_0 unobservable factors that affect the potential outcomes.[8] D^* is the latent desire to take up college education which depends on observed variables Z and unobservables V. Z includes all variables in X and some more (i.e., the instruments). Whenever D^* exceeds a threshold (set to zero without loss of generality), the individual opts for college education, otherwise she does not. U_1, U_0, V are potentially correlated, inducing the endogeneity problem (as well as heterogenous returns) as the researcher observes $Y (= DY^1 + (1 - D)Y^0), D, X, Z$, but not U_1, U_0, V.

Following this model, individuals are indifferent between high school education and the next best alternative (e.g., an apprenticeship) whenever the index of observables $Z'\delta$ is equal to the unobservables V. Thus, for indifferent individuals also the unobservables are known and observable. This property is exploited in the estimation. Since for every value of the index $Z'\delta$ one needs individuals with and without higher education, it is important to meaningfully aggregate the index by a monotonous transformation that for example returns the quantiles of $Z'\delta$ and U.

One can do achieve this by applying the cumulative distribution of $Z'\delta$ (for instance, the standard normal distribution) to the left and the right of the equation: $Z'\delta \geq V \Leftrightarrow \Phi(Z'\delta) \geq \Phi(V) \Leftrightarrow P(Z) \geq U_D$ where $P(Z) \equiv P(D = 1|Z) = \Phi(Z'\delta)$. If we vary the excluded instrument in $Z'\delta$ from the lowest to the highest value while holding the covariates X constant, more and more individuals will select into higher education. Those who react to this shift also reveal their rank in the unobservable distribution. Thus, the unobservables are fixed given the propensity sore and it is feasible to evaluate any outcome for those who select into treatment at any quantile U_D that is identified by the instrument-induced change of the higher education choice. In general, estimating marginal effects by U_D does not require stronger assumptions than those required by the LATE since Vyt-

---

[8]Note that the general derivation does not require linear indices. However, it is standard to assume linearity when it comes to estimation. Moreover, note that we dropped individual subscripts for simplicity.

lacil (2002) showed its equivalence.[9] Yet, strong instruments are beneficial for robustly identifying effects over the support of $P(Z)$. This, however, is testable.

The marginal treatment effect (MTE), then, is the marginal (gross) benefit of taking the treatment for those who are just indifferent between taking and not-taking it and can be expressed as

$$MTE(x, u_D) = \frac{\partial E(Y|x, p)}{\partial p}$$

This is the effect of an incremental increase in the propensity score on the observed outcome. This is identified by those who switch from not taking the treatment to taking the treatment when the propensity score is slightly increased – that is, by those at the margin of taking college education.

The MTE varies along the line of U_D in case of heterogeneous treatment effects which arise if individuals self-select into the treatment based on their expected idiosyncratic gains. This is a situation Heckman et al. (2006) call "essential heterogeneity". This is an important structural property that the MTE can recover: If individuals react already at low values of the instrument, where the observed part of the latent desire of selecting into, say, higher education ($P(Z)$) is still very low, a prerequisite for yet going to college is that V is marginally lower. These individuals could choose college against all (observed) odds because they are more intrinsically talented or motivated as indicated by a low V. If this is translated into higher future gains ($U_1 - U_0$), the MTE would exhibit a significant negative slope: As $P(Z)$ rises, marginal individuals need less and less compensation in terms of unobserved and expected returns to yet choose college – this is called selection into gains.[10]

In this case the common treatment parameters ATE, ATT, and LATE do not coincide. The MTE can be interpreted as a more fundamental parameter than the usual ones as it captures all local switching effects and not only some weighted average of those.

$$TE_j(x) = \int_0^1 MTE(x, u_D) h_j(x, u_D) du_D$$

---

[9]In this model the exclusion restriction is implicit since Z has an effect on D^* but not on Y^1, Y^0. Monotonicity is implied by the choice equation since D^* monotonously either increases are decreases the higher the values of Z.

[10]As Basu (2011, 2014) notes, essential heterogeneity is not restricted to sorting into gains but is always an issue if selection is based on factors that are not completely independent of the gains. Thus, in health economic applications, where gains are arguably harder to predict for the individual than, say, monetary returns, essential heterogeneity is also an important phenomenon.

---

where $TE_j(x)$ denotes some treatment effect j and $h_j(x, u_D)$ the respective weights (see, e.g. Heckman and Vytlacil, 2007, for the exact expressions of the weights for common parameters).

The main component for estimating the MTE is the conditional expectation $E(Y|X, p)$. Heckman and Vytlacil (2007) show that

$$
\begin{aligned}
E(Y|X, p) &= X'\beta_0 + X'(\beta_1 - \beta_0) \cdot p + E(U_1 - U_0 | D = 1, X) \cdot p \\
&= X'\beta_0 + X'(\beta_1 - \beta_0) \cdot p + K(p)
\end{aligned}
\tag{4}
$$

where $K(p)$ is some not further specified function of the propensity score if one wants to avoid distributional assumptions of the error terms. Thus, the estimation of the MTE involves estimating the propensity score in order to estimate equation (4) and, finally, taking its derivative with respect to p.

In our application, we impose the Conditional Independence Assumption of the instrument

$$
(U_1, U_0, V) \perp\!\!\!\perp Z | X
$$

meaning that the error terms are independent of Z given X. That is, after conditioning on X a shift in the instruments Z (or the single index $P(Z)$) has no effect on the potential outcome distributions. However, estimating the two sources of heterogeneity by the interaction terms $X'(\beta_1 - \beta_0)p$ in (4) is not possible while maintaining conditional exclusion restrictions on the instruments. Therefore, Carneiro et al. (2011) make the stronger assumption of unconditional independence:[11] $(U_1, U_0, V) \perp\!\!\!\perp Z, X$. The intuitive reason for its necessity is that the interaction term absorbs partial correlation of X and p that might also co-vary with the error term. The estimated parameters of the interaction terms, thus measures a complementary X-effect with respect to the reference category but not the ceteris paribus effect of either one of its components. As a consequence the fixing of X does not work unless one assumes zero correlation of Z and the unobservables U_1, U_0 also between different values of X as in the Unconditional Independence Assumption.

One possible solution to get by with the Conditional Independence Assumption would be to condition on X non-parametrically by estimating separate MTEs for every data cell determined by X. This is hardly ever feasible due to a lack of observations and powerful instruments within each such cell. The other solution in bringing the empirical framework to the data, is to take a more pragmatic approach and estimate marginal effects that

---

[11]They do, however, also provide analyses that partly relax this assumption. Another study that does not need to impose unconditional independence is Brinch et al. (2012).

only vary over the unobservables while fixing the X-effects at mean value. This means to deviate from (4) by restricting $\beta_1 = \beta_0 = \beta$ except for the intercepts α_1, α_0 in (1) and (2) such that $E(Y|X, p)$ becomes:

$$E(Y|X, p) = X'\beta + (\alpha_1 - \alpha_0) \cdot p + K(p) \tag{5}$$

Thus, we allow for different levels of potential outcomes, while we keep conditioning on X. Even with the true population effects that are varying over X, note that the derivative of Equation (4) w.r.t. the propensity score is constant in X. Hence, only the level of the MTE changes for certain subpopulations determined by X, the curvature remains unaffected. Thus, estimation of Equation (5) delivers an MTE that has a level which is averaged over all subpopulations without changing the curvature. So all crucial elements of the MTE are preserved, since we are interested in the average effect and its heterogeneity with respect to the unobservables for the whole population. How this heterogeneity is varying for certain subpopulations is of less importance and also the literature has focused on MTEs where the X-part is averaged out. On the other hand we gain with this approach by considerably relaxing our identifying assumption.

In estimating (5), we follow Carneiro et al. (2010, 2011) again and use semi-parametric techniques using the estimator suggested by Robinson (1988). Alternatively one might use a flexible approximation of $K(p)$ based on a polynomial of the propensity score as done by Basu et al. (2007).[12]

Semi-parametrically, the MTE can only be identified over the support of P. The greater the variation in Z (conditional on X) and, thus $P(Z)$, the larger the range over which the MTE can be identified. This may be considered a drawback of the MTE approach, in particular, because treatment parameters that have weight unequal to zero outside the support of the propensity score are not identified using semi-parameteric techniques. This is sometimes called the "identification at infinity" requirement (see Heckman, 1990) of the MTE. However, we argue that the MTE over the support of P is already very informative. We use semi-parametric estimates of the MTE and restrict the results to the empirical ATE or ATT that are identified for those individuals who are in the sample (see Basu et al., 2007).

---

[12]This amounts to estimating $E(Y|X, p) = X'\beta + (\alpha_1 - \alpha_0) \cdot p + \sum_{j=1}^{k} \phi_j p^j$ by OLS and using the estimated coefficients to calculate $\widehat{MTE}(x, p) = (\widehat{\alpha}_1 - \widehat{\alpha}_0) + \sum_{j=1}^{k} \widehat{\phi}_j j p^{j-1}$.

4 Data

4.1 Sample selection and college education

Our main data source are individual level data from the German National Educational Panel Study (NEPS), see Blossfeld et al. (2011). NEPS data map the educational trajectories of more than 60,000 individuals in total. The data set consists a multi-cohort sequence design and covers six age groups, called "starting cohorts": newborns and their parents, pre-school children, children in school grades 5 and 9, college freshmen students, and adults. Within each starting cohort the data are organized in a longitudinal manner, i.e., individuals are interviewed repeatedly. For each starting cohort, the interviews cover extensive information on competence development, learning environments, educational decisions, migrational background, and socioeconomic outcomes.

We aim at analyzing longer term effects of college education and, therefore, restrict the analysis to the "adults starting cohort". For this age group five waves are available with interviews conducted between 2007/2008 (wave 1) and 2012/2013 (wave 5)[13], see LIfBi (2015). Moreover, NEPS data include detailed retrospective information on the educational and occupational history as well as the living conditions at the age of 15 – about three years before individuals decide for higher education. From the originally 17,000 respondents in the adults starting cohort, born between 1944 and 1989, we exclude observations for four reasons: First, we focus on individuals from West Germany due to the different educational system in the former German Democratic Republic (GDR), thereby dropping 3,500 individuals living in the GDR at the age of the college decision. Second, to allow for long term effects we make a cut-off at college attendance before 1990 and drop 3,800 individuals who graduated from secondary school in 1990 or later. Third, we drop 1,600 individuals with missing spatial information. An attractive (and for our analysis necessary) feature of NEPS data is that they include information on the district (German *Kreis*) of residence during secondary schooling which is used to assign instruments in the selection equation. The fourth reason for losing observation is that the dependent variables are not available for each respondent, see below. Our final sample includes between 2,587 and 8,018 observations, depending on the outcome variable.

The explanatory variable "college degree" takes on the value 1 if an individual has any higher educational degree, and 0 otherwise. Dropouts are treated as all other individuals without college education. About one fourth of the sample has a college degree, while three fourth do not.

---

[13]It turns out that we only use information from waves 2, 3, and 4 and only cross-sectional information as each outcome variable only appears in one specific wave. See below for more information.

4.2 Dependent variables

Cognitive abilities

Traditionally, the development of cognitive abilities is subject to psychological research. Cognitive abilities summarize the "ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought" (American Psychological Association, 1995), where the sum of these abilities is referred to as intelligence. Psychologists distinguish several concepts of intelligence with different cognitive abilities; however, they all include measures of verbal comprehension, memory and recall as well as processing speed.

College education is likely to affect the observed level of cognitive abilities through the so-called cognitive reserve – the mind's ability to tolerate brain damage (Meng and D'Arcy, 2012). It is well-documented (see, for instance, the review articles of Lindenberger, 2014, and Salthouse, 2004) that individuals suffer an age-related decline in brain functioning. According to the "cognitive reserve hypothesis" this decline might be eased through the cognitive reserve. Neuropsychological research suggests that the cognitive reserve works through a "more efficient utilization of brain networks or [an] enhanced ability to recruit alternate brain networks as needed" (Stern, 2002, p.448) and is affected by the level of education, see, e.g., Meng and D'Arcy (2012). The literature considers two mechanisms – or a combination of both – through which education may affect the cognitive reserve (see Meng and D'Arcy, 2012): First, college education leads to a higher cognitive reserve when leaving college but the rate of decline in brain functioning is the same for college graduates and non-graduates. Thus, the college surplus in skills remains stable over time. Second, the initial cognitive reserve is the same for both groups but individuals with college education suffer a lower rate of decline and, hence, exhibit a higher cognitive reserve at older ages.

Although comprehensive intelligence tests take hours, a growing number of socioeconomic surveys have much shorter proxies included that measure specific skill components. The short ability tests are usually designed by psychologists and the results are highly correlated with the results of more comprehensive intelligence tests (cf. Lang et al., 2007, for a comparison of cognitive skill tests in the German Socio-economic Panel with larger psychological test batteries). The NEPS includes three kinds of competence tests which cover various domains of cognitive functioning: reading speed, reading competence, and mathematical competence.[14] All competence tests were conducted in 2010/2011 (wave 3) as paper and pencil tests under the supervision of a trained interviewer and the test language was German.

---

[14]For a general overview over test designs and applications in the NEPS, see Weinert et al. (2011).

Table 1: Descriptive statistics dependent variables

	(1)	(2)	(3)	(4)	(5)	(6)
	Cognitive ability components			Health measures		
	Read. speed	Read. comp.	Math liter.	PCS	MCS	Health satis.
Observations	3,559	4,116	2,587	4,305	4,305	8,018
with college degree (in %)	28.9	28.0	28.8	29.0	29.0	26.2
Raw values						
Mean with degree	39.99	30.20	13.36	53.30	51.26	7.50
Mean without degree	35.84	22.51	9.11	50.14	50.59	7.08
Maximum possible value	51	39	22	100	100	10
Standardized values						
Mean with degree	0.35	0.69	0.64	0.24	0.05	0.14
Mean without degree	−0.14	−0.27	−0.26	−0.10	−0.02	−0.05

Notes: Own calculations based on NEPS-Starting Cohort 6 data.

The first test measures reading speed.[15] The participants receive a booklet consisting of 51 short true-or-false questions and the test duration is 2 minutes. Each question has between 5 and 18 words. The participants have to answer as many questions as possible in the given window. The test score is the number of correct answers. Since the test aims at the answering speed, the questions only deal with general knowledge and use easy language. One question/statement, for example, reads "There is a bath tub in every garage.". The mean number of correct answers in our estimation sample is 39.99 (out of 51) for college graduates and 35.84 for others, see Table 1. For more information, see Zimmermann et al. (2014).

The reading competence test measures understanding of texts. It lasts 28 minutes and covers 32 items. The test consists of three different tasks. First, participants have to answer multiple choice questions about the content of a text, where only one out of four possible answers is right. In a decision-making task, the participants are asked whether statements are right or wrong according to the text. In a third task, participants need to assign possible titles out of a list to sections of the text. The test includes serval types of texts, e.g., comments, instructions, and advertising texts (LIfBi, 2011). Again, the test

---

[15]The test measures the "assessment of automatized reading processes", where a "low degree of automation in decoding [...] will hinder the comprehension process", i.e., understanding of texts (Zimmermann et al., 2014, p.1). The test was newly designed for NEPS but based on the well-established Salzburg reading screening test design principles (LIfBi, 2011).

score reflects the number of correct answers. Participants with college degree score on average 30.20 and without 22.51 (out of 39).[16]

The mathematical literacy test evaluates "recognizing and [...] applying [of] mathematics in realistic, mainly extra-mathematical situations" (LIfBi, 2011, p.8). The test has 22 items and takes 28 minutes. It follows the principle of the OECD-PISA tests and consists of the areas quantity, space and shape, change and relations, as well as data and change, and measures the cognitive competencies in the areas of application of skills, modelling, arguing, communicating, representing, as well as problem solving; see LIfBi (2011). Individuals without college degree score on average 9.11 (out of 22) and persons who graduated from college receive 4.25 points more.

Due to the rather long test duration given the total interview time, not every respondent had to do all three tests. Similarly to the OECD-PISA tests for students, individuals were randomly assigned a booklet with either all three or two out of the three tests. 3,559 individuals did the reading speed test, 4,116 the reading competence test, and 2,587 math. Since the tests measure different competencies that refer to distinct cognitive abilities, we may not combine the different test scores into an overall score but give the results separately (see Anderson, 2007). The plots in the left-hand side of Figure 3 present the standardized (mean = 0, SD = 1) test score distributions by college education. The density with college degree is clearly located to the right of the density without degree.
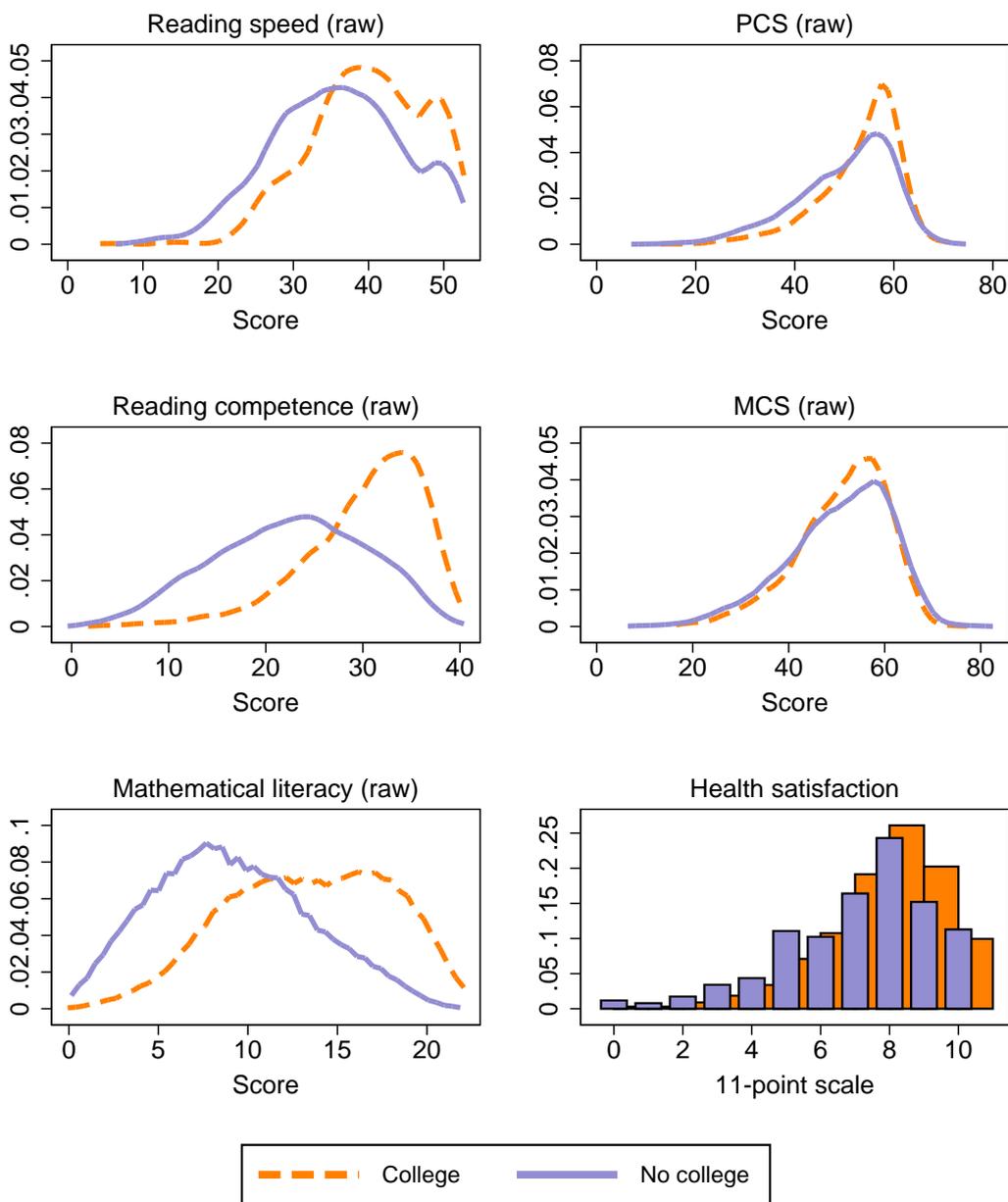
Health measures

Three variables from the health domain are used as outcome measures: the Physical Health Component Summary Score (PCS), the Mental Health Component Summary Score (MCS), both wave 4 (2011/2012), and health satisfaction from wave 2 (2009/2010).

The summary scores for mental and physical health (MCS and PCS) are based on the SF12 questionnaire, which is an internationally standardized set of 12 items regarding eight dimensions of the individual's health status. These eight dimensions comprise physical functioning, physical role functioning, bodily pain, general health perceptions, vitality, social role functioning, emotional role functioning and mental health. A scale ranging from 0 to 100 is calculated for each of these eight dimensions. The eight dimensions or subscales are then aggregated to the two main dimensions mental and physical health, using weights derived from an explorative factor analysis. The aggregated scales (MCS and PCS) are standardized and transformed to have a mean of 50 and a standard deviation of 10 with lower values indicating lower health states (Andersen et al., 2007) – to make the scores comparable to the skill outcomes, we rescale them to mean 0 and standard deviation 1. As MCS and PCS are based on the same set of subscales, they might overlap to a certain degree. However, due to the different weighting schemes, they are

---

[16]The total number of possible points exceeds 32 because some items were worth more than one point.

Figure 3: Distribution of dependent variables by college graduation

Notes: Own illustration based on NEPS-Starting Cohort 6 data.

expected to absorb different components of overall health. Negative and small correlation coefficients between MCS and PCS confirm this.

Health satisfaction is measured on an 11-point scale from 0 (very unsatisfied) to 10 (very satisfied). It is a subjective measure but these kinds of measures have been shown to have high predictive power for morbidity and subsequent mortality (see, e.g., Idler and Benyamini, 1997, for a review of studies which use self-rated health as a health outcome). Furthermore, it gives a more complete picture of overall health than many single objective

measures can do. It combines physical and mental health and might be the preferred measure when we think of health in terms of the utility derived from it.

Columns (4) to (6) of Table 1 report sample means and panels 4-6 in Figure 3 show the distribution of the health measures across individuals by college graduation. Those with college degree have, on average, a better physical health score and a higher health satisfaction. However, the differences are less pronounced than in the case of cognitive abilities. With respect to mental health, both groups differ only marginally.

4.3   Control variables

Based on their birth and graduation year, individuals made their college decision between 1958 and 1990. The NEPS allows us to consider important socioeconomic characteristics that probably both affect the college education decision as well as the outcomes today (variables denoted with X in Section 3). This is *general demographic information* such gender, migrational background, and family structure, *parental characteristics* like parent's educational background. Moreover, we include two blocks of controls that were determined before the educational decision was made. *Pre-college living conditions* include family structure, parental job situation and household income at the age of 15, while *pre-college education* includes educational achievements (number of repeated grades and secondary school graduation mark).

Table A1 in the Appendix provides more detailed descriptions of all variables and reports the sample means by treatment status. Apart from higher abilities and a better health status (as seen in Table 1), individuals with a college degree are more likely to be males from an urban district without a migrational background. Moreover, they are more likely to have healthy parents (in terms of mortality). Other variables seem to differ less between both groups.

In the regressions we also include a full set of individual cohort fixed-effects, cohort-effects of mother and father, district fixed effects as well as federal state-specific time trends (see Mazumder, 2008, and Stephens and Yang, 2014, for the importance of the latter).

4.4   Instruments

All the processes of college expansion discussed in Section 2.2 probably shifted individuals also with a lower desire to study into college education. Such powerful exogenous variation is beneficial for our approach as we try to identify the distribution of treatment

effects by desire to study. We assign each individual the college availability and student loan regulations as instruments (that is, variables in Z but not in X). In doing so, we use the information on the district of the secondary school graduation and the year of the college decision. The latter is the year of secondary school graduation plus, when applicable, one year of compulsory military or civilian service. The district – there are 326 districts in West Germany – is either a city or a certain rural area.

Regarding student loan regulations, we implicitly exploit the exogenous kinks the propensity for college education by using the the level of the individual *Bafoeg eligibility threshold* as an instrument (see Section 2.2). If parental household income crosses this threshold from below, the maximum BAfoeG aid is gradually reduced (a graphical representation for this threshold is the right dashed line in Figure 2). This instrument varies over individuals (as the threshold is calculated using individual characteristics as the number of children per family; for details consult the Supplementary Materials A.). As we control for all these factors causing the individual variation, identifying power is still enhanced by different trends in the adjustment of the components over time.

Regarding college availability, the operationalization is as follows. The question is how to exploit the regional variation in openings and spots most efficiently as it is almost infeasible to control for all distances to all colleges simultaneously. Our approach to this question is to create an index that best reflects the educational environment in Germany and combines the distance with the number of college spots:

$$Z_{it}^C = \sum_{j}^{326} K(dist_{ij}) \times \left(\frac{\#students_{jt}}{\#inhabitants_{jt}} \right). \tag{6}$$

The college availability instrument Z_{it}^C basically includes the number of college spots (measured by the number of students) per inhabitant in district j (out of 326 districts in total) individual i faces year in t weighted by the distance between i's home district and district j. Weighting the number of students by the population of the district takes into account that districts with the same number of inhabitants might have colleges of a different size. This local availability is then weighted by the Gaussian kernel distance $K(dist_j)$ between the centroid of the home district and the centroid of district j (if a district has more than one college, we use the total number of students in the district). The kernel puts a lot of weight to close colleges and a very small weight to more distant colleges. Since individuals can choose between many districts with colleges, we calculate the sum of all district-specific college availabilities within the kernel bandwidth.

Figure A2 in the Appendix shows the average distance to the closest college 5 years before and after a new college opened (in $t = 0$) in the district. By definition, after a college opened in the district, the distance is 0 (as long as these universities do not shut down

Table 2: Descriptive statistics of instruments and background information

	(1)	(2)	(3)	(4)
		Stati	stics	
	Mean	SD	Min	Max
Instrument 1: College availability	0.471	0.267	0.046	1.128
Background information on college availability (implicitly included in the instrument)				
Distance to nearest college	28.097	26.605	0	171.355
College in district at the time of graduation	0.120	0.325	0	1
Colleges within 100km	5.873	3.424	0	16
College spots per inhabitant within 100km	0.035	0.019	0	0.172
Instrument 2: BAfoeG eligibility threshold[a]	893.7	629.0	0	5,127.8

Notes: Own calculations based on NEPS-Starting Cohort 6 data and administrative data. For college availability data from the German Statistical Yearbooks 1959–1991 (German Federal Statistical Office, 1991) is used. Distances are calculated as the Euclidean distance between two respective district centroids. Information on the BAfoeG eligibility threshold is taken from German Federal Government (1992). Note that the BAfoeG eligibility threshold is an individual income threshold reflecting the financial needs of the family. See the Supplementary Materials A for details. BAfoeG eligibility threshold in German Marks in prices of 1971 (year of BAfoeG loan introduction).

[a] The maximal BAfoeG eligibility threshold is an outlier caused by household high a high number of children, the 95 quantile is 1,816 German Marks.

again). Due to the opening of the college, the average distance to the closest college falls by approximately 45 km on average. This indicates that the reduction of the commuting distance to the next college is relevant for the individuals in our sample.

Table 2 presents the descriptive statistics for both instruments. For the compound instrument on college availability, we also provide background information on certain descriptive measures on distance and student density. As shown above, we implicitly condense this information in one measure.

5 Results

5.1 OLS

Although we are primarily interested in analyzing the returns to college education for the marginal individuals, we start with ordinary least squares (OLS) estimations as a benchmark. Columns (1) to (3) in Table 3 report results for the three measures of cognitive abilities, while columns (4) to (6) do the same for health. Each cell reports the coefficient of college education of a separate regression. Standardization allows interpreting the coefficients in units of a standard deviation. Panel A shows the OLS results. After conditioning on observables, individuals with a college degree read, on average, 0.4 SD

Table 3: Regression results for OLS and First Stage estimations

	(1)	(2)	(3)	(4)	(5)	(6)
	Cognitive ability component			Health measure		
	Read. speed	Read. comp.	Math liter.	PCS	MCS	Health satis.
Panel A: OLS results						
College degree	0.394***	0.763***	0.713***	0.263***	0.038	0.162***
	(0.039)	(0.033)	(0.042)	(0.037)	(0.038)	(0.028)
Panel B: 2SLS first-stage results						
College availability	1.890***	1.806***	1.919***	1.958***	1.958***	1.819***
	(0.095)	(0.088)	(0.113)	(0.085)	(0.085)	(0.060)
BAfoeG eligibility	0.014***	0.015***	0.015***	0.011***	0.011***	0.013***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)
Number of observations	3,559	4,116	2,587	4,305	4,305	8,018
First-stage F statistic instr.	222.40	243.63	161.33	288.67	288.67	525.15

Notes: Own calculations based on NEPS-Starting Cohort 6 data. Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Regressions also include a full set of control variables as well as year-of-birth, district fixed effects, and federal state-specific trends.

faster than persons without college education. Moreover, they approximately have a by 0.7 SD better text understanding and mathematical literacy. While PCS and health satisfaction are also higher, there is no significant relation with MCS. All in all the results are pretty much in line with the differences in standardized means as shown in Table 1, slightly attenuated, however, due to the inclusion of control variables.

Panel B reports the first stage results of the 2SLS estimations. All coefficients of the instruments point into the expected direction and are individually significant. As to be expected, the coefficients barely change across the outcome variables (as the first-stage specifications only differ in the number of observations across the columns) and whether we use the instruments separately or simultaneously. Tests on joint significance of the instruments give F values well above 10, thereby fulfilling the Staiger and Stock (1997) rule of thumb.

In order to get a feeling for the effect size of college availability in the first-stage, we consider, as an example, the opening of the college in the city of Essen in 1972. In 1978, about 11,000 students studied in Essen. To illustrate the effect of the opening, we assume a constant population size of 700,000 inhabitants. The kernel weight of new spots in the same district is 0.4 (= $K(0)$). According to Equation (6), the instrument value increases by 0.006 (rounded). Given the coefficient of college availability of 1.9, an individual who

made the college decision in Essen in 1978 had a 1.14 percentage points higher probability to go to college due to the opening of the college in Essen (compared to an individual who made the college decision in 1971). This seems to be a plausible effect. The effect of the college opening in Essen on individuals who live in districts other than Essen is smaller, depending on the distance to Essen.

Panel B of Table 3 also gives the effect of an increase in the family income eligibility threshold for BAfoeG aid of 100 German mark (in prices of 1971). The probability of deciding for college education increases by about 1.5 percentage points (around 1.1 in columns (4) to (6)). An average increase of the eligibility threshold (15 German mark in the years under review), increases the probability to study by 0.23 percentage points (0.17 in columns (4) to (6)).

5.2 Marginal treatment effects

Figure 4a shows the distribution of the propensity scores used in estimating the MTE by treatment and control group.[17] They are obtained by logit regressions of the college degree on all Z and X variables. The regression results are reported in Table A2 in the Appendix. For both groups, the propensity score varies from 0 to about 1. Moreover, there is a common support of the propensity score almost on the unity interval. Variation in the propensity score after keeping the X variables fixed is used to identify local effects.
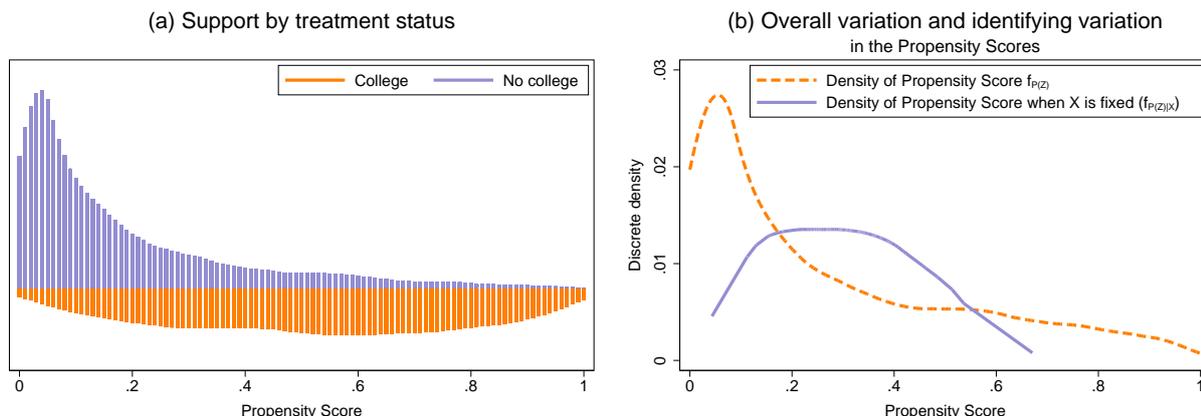
This variation is presented in Figure 4b. It shows the conditional support of P when the linear X-index of observables is held fixed ($f_{P(Z)|X}$). Here, the support ranges nearly from 0 to 0.8 only caused by variation in the instruments – the identifying variation. This is important in the semiparametric estimation since it shows the regions in which we can credibly identify (conditional on our assumptions) marginal effects without having to rely on inter- or extrapolations to regions where we do not have identifying variation.

We calculate the MTE using a local linear regression with a bandwidth that ranges from 0.11 to 0.29 depending on the outcome variable.[18] In Supplementary Materials B, we compare this with the series approximation approach and find both to produce similar results. The MTE distribution is generated by evaluating the derivative of Equation (5) in Section 3.

---

[17]The figure presents the results for the health satisfaction sample. The results do not change for the other outcome variables.

[18]We assess the optimal bandwidth in the local linear regression using Stata's `lpoly` rule of thumb. Our results are also robust to the inclusion of higher order polynomials in the local (polynomial) regression. The optimal, exact bandwidths are: for reading competence 0.13, for reading speed 0.13, math score 0.11, health satisfaction 0.15, MCS 0.12, PCS 0.29.

Figure 4: Distribution of propensity scores

(a) Support by treatment status

(b) Overall variation and identifying variation in the Propensity Scores

Notes: Own illustration based on NEPS-Starting Cohort 6 data. The left panel shows the propensity score (PS) density by treatment status. The right panel illustrates the joint PS density (dashed line). The solid line shows the PS variation solely caused by variation in Z, since the X-effects have been integrated out. Further note that in the right panel the densities were both normalized such that they sum up to one over the 250 points where we evaluate the density.

Figure 5 shows the MTE distributions for all outcome variables. The left column of Figure 5 plot the results for cognitive skills. For reading speed and reading competence we see that individuals with low values of U_D have highest skill returns to college education. Low values of U_D mean that these are the individuals who are very likely to study as already smaller values of $P(z)$ exceed U_D, see the transformed choice equation in Section 3. Thus, there is obviously selection into gains with respect to reading skills. For individuals with a high value of U_D, i.e., a low internal desire to go to college, the marginal returns are lower but still positive and significantly different from zero.

The left part of the MTE curve looks similar for mathematical literacy (bottom left). However, for the half of the population that is less likely to study (those with $U_D > 0.5$), the marginal math skill returns to college education even increase. This might not be explained with a selection but with lower math skills in the counterfactual situation. While reading skills might be needed in most occupations with and without college degree, math skills may be particularly important for jobs with college degree. In other words, both individuals who are very likely to study anyway but also those who are unlikely to go to college (would) benefit highest in terms of mathematical skills. In Section 5.3 we calculate treatment parameters from the MTE estimations and discuss effect sizes there.

The right column of Figure 5 presents the MTE distributions for health outcomes. In all three cases do we find homogeneous effects and, therefor, no evidence of selection into gains. As to be expected from the 2SLS results (which, in this case, coincide with the ATE), the effects are smaller compared to those for cognitive abilities, but still considerable in magnitude (again, see the next subsection for discussions of effect size). Again, however, the effect on mental health is neither economically, nor statistically different from zero.

Figure 5: Marginal Treatment Effects for cognitive abilities and health
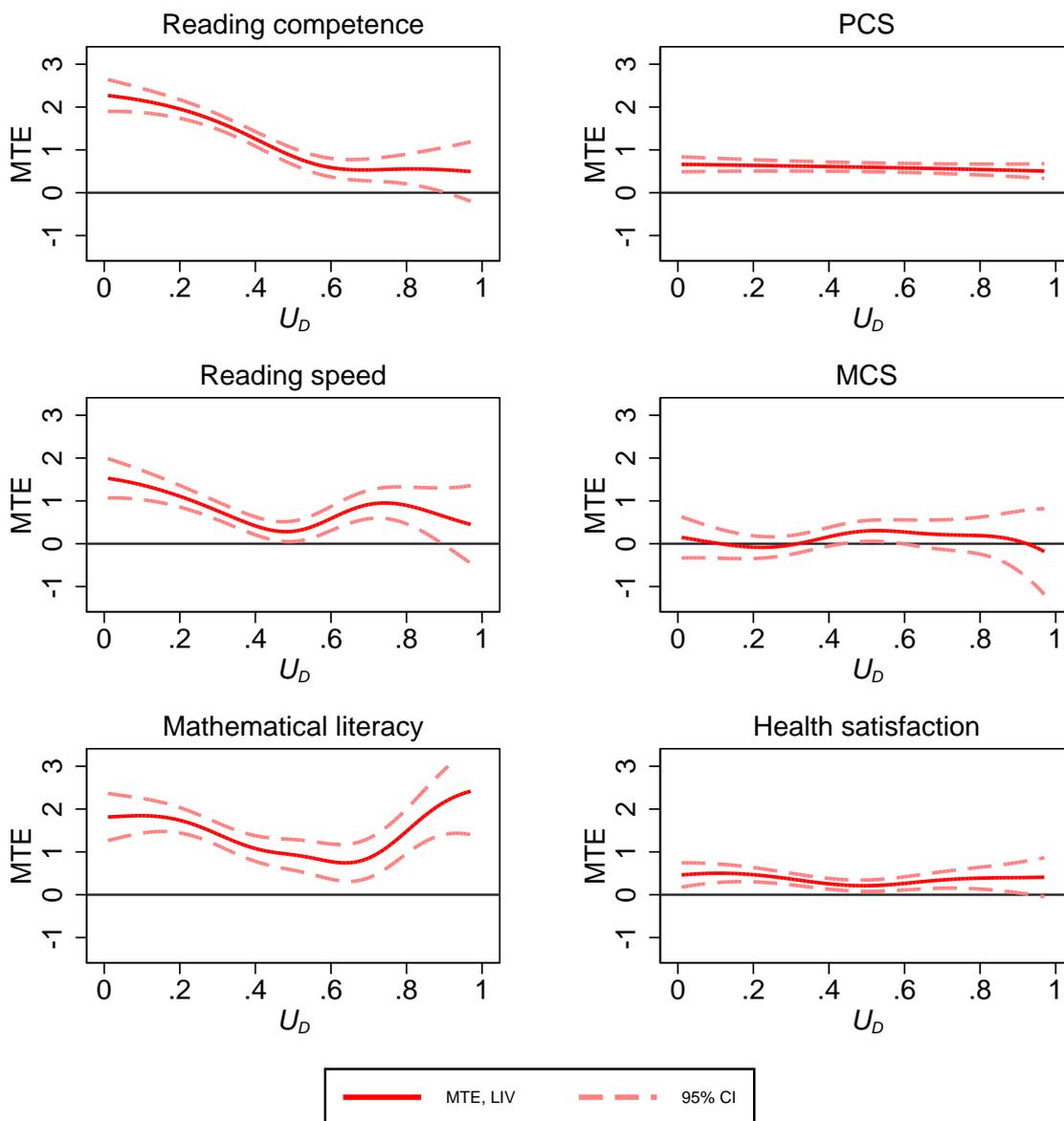
Notes: Own illustration based on NEPS-Starting Cohort 6 data. All outcomes are standardized to mean 0 and standard deviation 1. The dashed lines give the 95% confidence intervals. Calculations based on a local linear regression where the influence of the control variables was isolated using a semiparametric Robinson estimator (Robinson, 1988) for each outcome variable. The optimal, exact bandwidths for the local linear regressions are: for reading competence 0.13, for reading speed 0.13, math score 0.11, health satisfaction 0.15, MCS 0.12, PCS 0.29.

Thus, both the likely and the unlikely students (would) benefit from going to college in terms of physical health but not mental health.

In Supplementary Materials B, we show the results of robustness checks. First, we use series approximation instead of local IV estimation in order to estimate the MTE. The resulting MTE distributions are nearly the same over all values of U_D. Other robustness checks employ different specifications of the college availability instrument: (i) using different kernel bandwidths to weight the college distance, (ii) only use the sum of the kernel

weighted distances (bandwidth 250 km) to calculate the college availability (the college size is not taken into account), and (iii) college availability is boiled down to a binary indicator (that takes the value one if a college is in the district of secondary school graduation), as for instance in Card (1995). Although the condensation of college availability in Equation (6) is somewhat arbitrary, these robustness checks hint that the specification of the instrument does not affect our findings.

5.3   Treatment parameters

Table 4 reports the conventional treatment parameters estimated using the MTE and the respective weights as described above and more formally derived and explained in, for example, Heckman and Vytlacil (2007). In particular, we calculate the average treatment effect (ATE), the average treatment effect on the treated (ATT) and the average treatment effect on the untreated (ATU). The estimated weights applied to the returns for each U_D on the MTE curve are shown in Figure 6. Whereas the local average treatment effect is an average effect weighted by the conditional density of the instrument, the ATT (vice versa for the ATU) for example gives more weight to those individuals that select already into higher education at low U_D values (indicating low intrinsic reluctance for higher education). The reason is that their likelihood of being in any 'treatment group' is higher compared to individuals with higher values of U_D. The ATE places equal weight over the whole support.

The LATE parameters in column (4) are calculated using conventional 2SLS (though roughly the same parameter would be obtained if respective LATE-weights would be applied to the MTE) and both instruments[19]. In all cases but mental health the coefficients on college education approximately double with respect to the OLS estimates. Increasing 2SLS coefficients (compared to OLS) seem to be counterintuitive as one often expects OLS to overestimate the true effects. Yet, this is not an uncommon finding and in a world with heterogeneous effects often explained by the group of compliers that potentially has higher individual treatment effects than the average individual (Card, 2001). This is directly obvious by comparing the LATE to column (1) which is another indication of selection into gains. Regarding the other treatment parameters (but physical health), the LATE lies within the range of the ATT and the ATU.

Note that these are the "empirical", conditional-on-the-sample parameters as calculated in Basu et al. (2007), that is, the treatment parameters conditional on the the common support of the propensity score. The population ATE, however, would require full sup-

---

[19]The results are fairly robust with respect to employing either one of the instruments separately; see Supplementary Materials B for details.

Table 4: Estimated treatment parameters

	(1)	(2)	(3)	(4)
	Treatment Parameters			
	ATE	*ATT*	*ATU*	*LATE*
Mathematical literacy	1.16	1.39	1.05	1.31
	(0.11)	(0.14)	(0.13)	(0.11)
Reading competence	1.14	1.58	0.95	1.39
	(0.09)	(0.12)	(0.10)	(0.10)
Reading speed	0.68	0.93	0.60	0.73
	(0.11)	(0.15)	(0.12)	(0.10)
Health satisfaction	0.33	0.38	0.30	0.39
	(0.08)	(0.11)	(0.08)	(0.08)
MCS	0.07	0.03	0.10	0.13
	(0.25)	(0.28)	(0.26)	(0.10)
PCS	0.40	0.46	0.44	0.63
	(0.09)	(0.12)	(0.10)	(0.10)

Notes: Own calculations based on NEPS-Starting Cohort 6 data. LATE results are calculated using 2SLS. All other parameters: Parameter estimation based on a series approximation with third order polynomial; PS not interacted with the observables. Standard error estimated using a conventional bootstrap with 100 iterations.

port on the unity interval.[20] As depicted in Figure 4, we do not have full support in the data at hand. Although we observe individuals with and without college degree for most probabilities to study, we cannot observe an individual with a probability arbitrarily close to 100% without college degree (and arbitrarily close to 0% with a degree). Instead, the parameters in Table 4 were computed using the marginal treatment effects on the common support only. However, as this reaches from 0.002 to 0.969 (in the health satisfaction sample as one example) it seems fair to say that this probably comes very close to the true parameters.[21]

Table 4 is informative in particular for two reasons. First, it boils down the MTE to single numbers such that the average effect size is immediately clear. And second, differences between the parameters again emphasize the role of effect heterogeneity. Together with the bootstrapped standard errors the table reveals that the ATT and the ATU structurally differ from each other for all skill outcomes, whereas the parameters of the health out-

---

[20]The ATT would require for every university graduate in the population a non-graduate with the same Propensity Score (including 0%). For the ATU one would need the opposite: a graduate for every non-graduate with the same Propensity Score including 100%.

[21]Note that we use the results from the series approximation instead of the Robinson estimator for this table. This is because standard errors are calculated by bootstrapping and the much longer estimation time for the semiparametric method prohibits its use in the bootstrap. Note again that, on the support, the MTEs do not strongly differ between both methods, see Figure S2.

Figure 6: Treatment parameter weights conditional on the propensity score
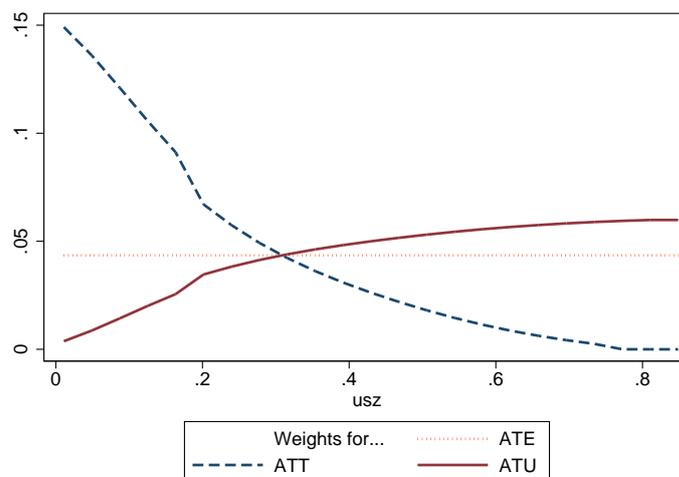
Notes: Own illustration based on NEPS-Starting Cohort 6 data.

comes do not. Hence, the treatment group of university graduates seem to profit from higher education in terms of the skills but not in terms of health with respect to the non-graduates. One reason is that they choose to study because of their idiosyncratic skill returns. Yet, it is more likely to be windfall gains that go along with monetary college premiums that the decision was more likely to be targeted at. Nonetheless, this also is evidence for selection into gains.

The effect sizes for all (ATE), for the university degree subgroup (ATT), and for those without higher education (ATU) in Table 4 capture the overall returns to college education, not the per-year effects. On average, the per-year effect is approximately the overall effect divided by 4.5 years (the regular time it takes to receive a *Diplom* degree), if we assume linear additivity of the yearly effects. The per-year effects for mathematical literacy and reading competence speed are about 25% of a standard deviation for all parameters. For reading speed the effects are around 15% of an SD. These effects are of considerable size, yet slightly smaller than those found in the previous literature on different treatments and, importantly, different compliers. For instance, ability returns to an additional year of compulsory schooling were found to be up to 0.5 SD (see, e.g., Banks and Mazzonna, 2012).

To get an idea of the total effect of college education on, say, math skills, the following example might help. If you start at the median of the standardized unconditional math score distribution ($\Phi(0) = 50\%$), the average effect of 1.16 of a standard deviation, all other things the same, will make you end up at the 87% quantile of that distribution ($\Phi(0 + 1.16) = 87\%$) – in the thought experiment of being the only treated in the peer group.

As suggested by the distributions of the marginal treatment effects in Figure 5, the health returns to higher education are smaller than the skill returns, still they are around 10% of an SD per year (except for the zero effect on mental health). Given the previous literature, the results seem reasonable.

6   Potential mechanisms

Health and skills evolve hand in hand which makes it difficult to unravel. For example, mentally demanding activities on the job may affect the long-lasting ability returns to college education, but might also contribute to health via mental well-being. Physically demanding activities on the other hand may contribute to long-lasting physical health effects of college education, but could lead to a skill decline if mental activities are crowded out. The NEPS data allow us to investigate numerous potential mechanisms that may govern the long-lasting effects of education on cognitive skills and health.

6.1   Skill mechanisms

Two (possibly accompanying) reasons might explain why college education still affects the cognitive reserve decades after leaving the college: first, (some) individuals improve their skills in college and keep this advantage over the life-cycle against the control group of individuals who did not attend college. Second, the age related decline in cognitive skills is slower for those with college education. When individuals with college education engage in more cognitively demanding activities, e.g., more sophisticated jobs, this might mentally exercise their minds (Rohwedder and Willis, 2010). This effect of mental training is sometimes referred to as use-it-or-lose-it hypothesis, see Rohwedder and Willis (2010) and Salthouse (2006). If such an exercise effect leads to alternating brain networks that "may compensate for the pathological disruption of preexisting networks" (Meng and D'Arcy, 2012, p.2), a higher demand of cognitive skills (as a result of college education) increases the individual's cognitive capacity. Because the competence tests in NEPS are fairly different from merely engaging in activities like reading (although both things require the same skills), we think it is unlikely that the cognitive skill tests reflect some retesting effect (Ferrer et al., 2004). In other words, the competence tests really measure cognitive skills instead of a familiarity with activities like reading texts.

In order to investigate if a more cognitively demanding job might be a potential mechanism (as, e.g., suggested by Fisher et al., 2014), we use information on the individual's activities on the job. All outcome variables considered in this chapter are binary, their sample means and the definitions of the indicators are given in Panel A in Table 5, OLS

Table 5: Sample means and definition of potential mechanisms

	Sample mean	Definition
Panel A: Mentally demanding activities on the job (potential skill mechanisms)		
Math: counting	0.845	=1 if job requires counting (e.g., change at a cash register)
Math: simple comp.	0.814	=1 if job requires summations and subtractions
Math: percentages	0.706	=1 if job requires calculating with percentages and fractions
Math: volumes	0.307	=1 if job requires calculating volumes and areas
Reading	0.777	=1 if respondent often spends more than 2 hours reading
Writing	0.693	=1 if respondent often writes more than 1 page
Learning new things	0.664	=1 if respondent reports to learn new thinks often
Panel B: Physically demanding activities on the job (potential health mechanisms)		
Standing position	0.311	=1 if often working in a standing position for 2 or more hours
Uncomfortable pos.	0.195	=1 if respondent needs to bend, crawl, lie down, keen or squat
Walking	0.245	=1 if job often requires walking, running or cycling
Carrying	0.307	=1 if often carrying a load of at least 10 kg
Heat or cold	0.122	=1 if often exposed to great heat or cold
Panel C: Health behaviors (potential health mechanisms)		
Adipositas	0.152	=1 if Body Mass Index (=mass in kg/height in m^2) exceeds 30
Smoking	0.273	=1 if currently smoking
Alcohol frequency	0.404	=1 if alcohol consumption at least twice a week
Alcohol amount	0.185	=1 if three or more drinks when consuming alcohol
Sport	0.716	=1 if any sporting exercise in the last 3 months

Notes: Own calculations based on NEPS-Starting Cohort 6 data. Definitions are taken from the data manual. For the mathematical operations the indicator takes the value 1 if an individual reports use at least this operation or a more sophisticated operation.

and 2SLS results are reported in Supplementary Materials B. Figure 7 shows the MTE distributions for the effect of college education on cognitively demanding activities on the job. The left-hand side column presents four qualitatively distinguished mathematical operations that range from counting to calculating volumes. The right-hand side column shows the effects on reading (more than 2 hours), writing texts (more than 1 page) and often learning new things on the job.

For the first three mathematical operations, counting, simple computations, and percentage calculation, the MTE distributions in Figure 7 are significantly above zero for individuals with a rather high desire to study and close to zero for individuals with high unobservable costs of studying. Up to a value of U_D of 0.4 the MTE distributions of the mental activities exhibit a similar pattern as the marginal effects for mathematical literacy in Figure 5: the higher the unobservable desire for higher education, the higher the returns. For individuals with U_D close to zero the effect of college education on the probability to require counting, simple computations or percentage calculation on the

job increases by about 30 percentage points (pp). For the most demanding mathematical operation, calculating volumes, the confidence interval of the MTE distribution always includes zero.

The (marginal) effect of higher education on the probability to read or write long texts on the job is positive as well. Moreover, like the MTE distributions for reading competence and reading speed in Figure 5, the effect size is declining in U_D. Individuals with a high desire to study are up to 30 pp and 60 pp more likely to read and write, respectively, longer texts as part of thier job. For individuals with a low desire to study the effect declines to 10 pp but is still positive.

To sum up, college education does not only affect the engagement in more mentally de-manding activities on the job (comparison of college graduates and non-graduates) but distribution of the marginal returns within the group of graduates is comparable to the selection into gains we observe for cognitive abilities. The lower the unobservable costs of studying, the higher the returns. Although the potential mechanisms discussed here do not explain the entire distribution of the cognitive skill returns to college education and there are most likely several other characteristics that govern the long-lasting effects of education, it seems plausible that the mechanisms we consider explain the skill returns to some extent.

Figure 7: Marginal Treatment Effects for potential skill mechanisms on the job
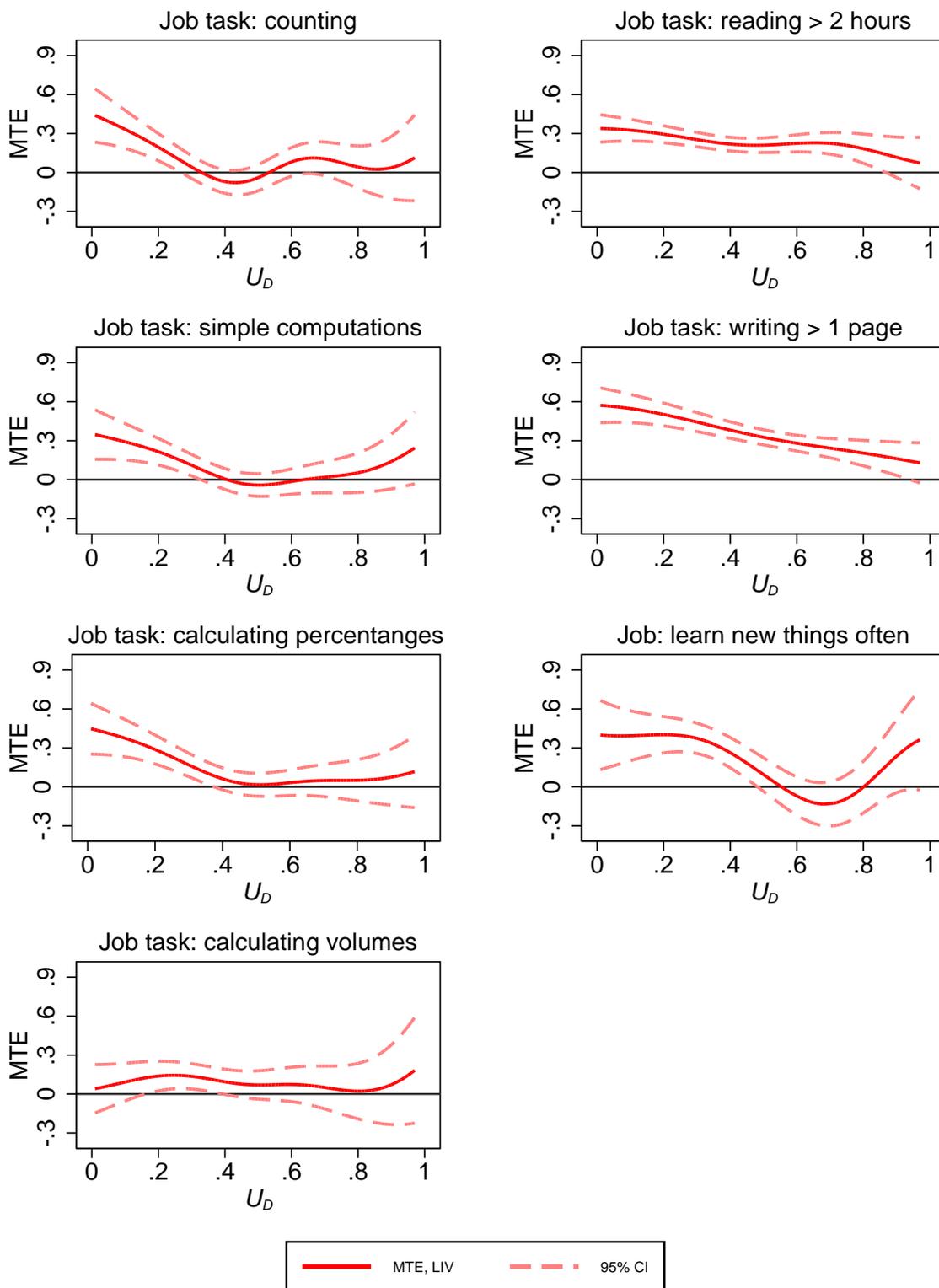
Notes: Own illustration based on NEPS-Starting Cohort 6 data. The outcomes are non-standardized. The dashed lines give the 95% confidence intervals. The number of observations are: counting 3,491, simple computations 3,491, percentage calculation 3,491, volumes calculation 3,491, reading 7,933, writing 3,491, and learning new things 3,491. The first-stage F statistics on the instruments range from 265.85 to 525.15. Calculations based on a local linear regression where the influence of the control variables was isolated using a semiparametric Robinson estimator (Robinson, 1988) for each outcome variable. The optimal, exact bandwidths for the local linear regressions are: counting 0.12, simple computations 0.14, percentage calculation 0.15, volumes calculation 0.14, reading 0.15, writing 0.20, and learning new things 0.12.

6.2 Health mechanisms

Concerning the health mechanisms, we study job-related and behavioral health effects. NEPS data cover engagement in five physical activities on the job: working in a standing position, working in an uncomfortable position (like bending often), walking or cycling long distances, carrying heavy loads, and being exposed to extreme temperatures while working. Panel B in Table 5 gives sample means and definitions. The binary indicators are coded as 1 if the respondent reports to engage in the activity (and 0 otherwise). The MTE distributions are given in Figure 8.

College education reduces the probability of engaging in the physically demanding activities. The MTE distributions are always significantly below zero (but for working in an uncomfortable position with U_D above 0.7). Given that college graduates are more likely to have an office job than non-graduates, this finding is fairly unsurprising. The effect of higher education on physically demanding activities is in line with the finding that college education increases objective health (PCS measure) and satisfaction with health. However, unlike PCS and health satisfaction, the probability to engage in physically demanding activities exhibits some heterogeneity in the returns. For individuals with a high desire to study, college education leads to a stronger reduction in the probability to suffer physically demanding activities than for individuals with a low desire to study.

Besides physical activities on the job, health behaviors may be considered as one important dimension of the general formation of health over the life-cycle, see Cutler and Lleras-Muney (2010). To analyze this we resort to the following variables in our data set: a binary indicator for adipositas (body mass index exceeds 30) as a compound lifestyle measure and more direct behavioral variables like an indicator for smoking, the frequency (1 if alcohol at least twice a week) and the amount (1 if at least three or more drinks when consuming alcohol) of alcohol consumption, as well as physical activity as measured by an indicator of having taken any sport exercise in the last 3 months. Panel C in Table 5 reports the sample means while Figure 9 displays marginal treatment effects.

Figure 8: Marginal Treatment Effects for potential health mechanisms on the job

Figure 9: Marginal Treatment Effects for potential health behaviors
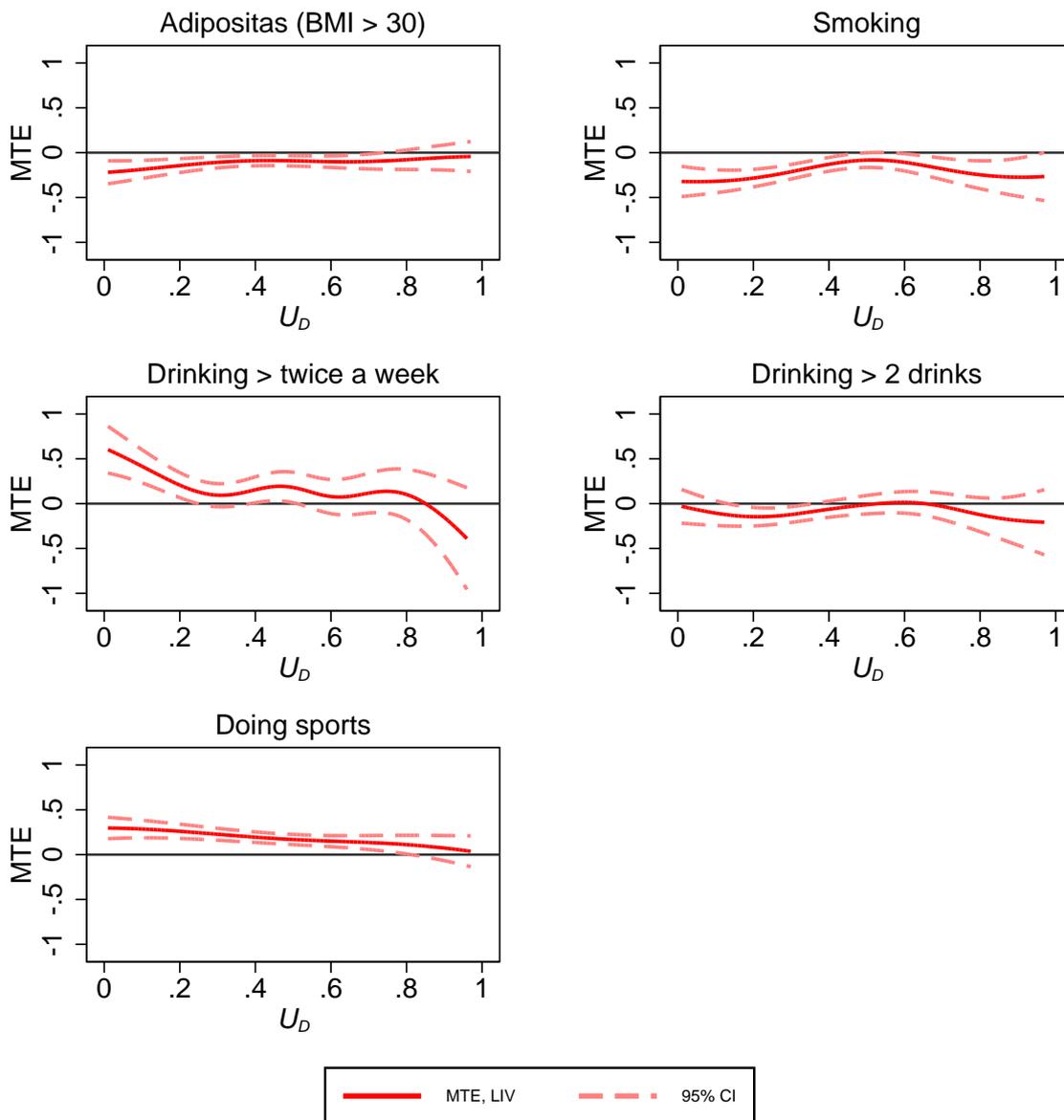
Notes: Own illustration based on NEPS-Starting Cohort 6 data. The outcomes are non-standardized. The dashed lines give the 95% confidence intervals. The number of observations are: adipositas 4,228, smoking 4,279, alcohol frequency 4,276, alcohol amount 3,916, sport 4,279. The first-stage F statistics on the instruments ranges from 270.16 to 287.33. Calculations based on a local linear regression where the influence of the control variables was isolated using a semiparametric Robinson estimator (Robinson, 1988) for each outcome variable. The optimal, exact bandwidths for the local linear regressions are: adipositas 0.16, smoking 0.15, alcohol frequency 0.10, alcohol amount 0.12, sport 0.21.

College education leads to a decrease in the probability to suffer adipositas for nearly all college graduates but those who have high unobservable costs of studying. The effect size reaches up to 20 pp for some individuals. College education significantly decreases the probability of smoking for nearly all graduates in the sample by up to 30 pp. This is in line with LATE estimates of the effect of college education in the US of Grimard and Parent (2007) and de Walque (2007). Regarding alcohol consumption there seems to be a positive effect of education on the frequency of alcohol consumption (especially for individuals with low unobservable costs of studying) but a negative effect on the amount, given that people drink. The latter effect is not statistically different of zero. While one should be cautious in interpreting the effect of the frequency as there is a strong heterogeneity, it might hint at the common finding that regular but moderate alcohol consumption is not harmful to health while binge drinking is certainly harmful. The effect of college education on the probability of doing sports is positive over the whole effect distribution. For individuals with a high desire to study the effect is up to 30 pp and it decreases along unobservable costs but is always positive and for most parts of the distribution significantly different from zero.

All in all, college education affects all potential health mechanisms in the expected direction. The potential mechanisms exhibit stronger heterogeneity in the effect size, however. Since health is a high dimensional measure, the potential mechanisms at hand are of course not able to explain the health returns to college education entirely. Nevertheless, the findings encourage us in our interpretation of the effects of college education on physical health and health satisfaction.

7 Conclusion

This paper uses the Marginal Treatment Effect framework introduced and advanced by Heckman and Vytlacil (2005, 2007) to estimate non-monetary returns to college education under essential heterogeneity. We use representative data from the German National Educational Panel Study (NEPS). Our outcome measures are cognitive abilities and health. The former are assessed using state-of-the-art cognitive competence tests on individual reading speed, text understanding, and mathematical literacy. As expected, cognitive abilities and health are positively correlated with having a college degree in our data set. Using instruments that exploit exogenous variation in the supply of colleges and student loan eligibility, we estimate marginal returns to college education. We find that there is indeed heterogeneity in the effect of college education on cognitive abilities. People select into college education in accordance with their gains in cognitive abilities. Yet the effect is positive almost over the entire distribution.

For physical health and health satisfaction we find homogeneously and significantly positive effects of college education. Thus, there does not seem to be selection into gains in terms of health. Moreover, we find no evidence that individuals benefit in terms of mental health from education. The distribution of the mental health returns is flat and around zero. Potential mechanisms of skill returns are more demanding jobs that slow down the cognitive decline in later ages. Regarding health we find positive effects of higher education on BMI, non-smoking, sports participation and moderate alcohol consumption.

The results generally suggest positive non-monetary returns to higher education. However, they clearly show that this is not the case for each and every individual. While some might be better of without more education (in particular in terms of net benefits when opportunity costs are taken into account) the average individual clearly seems to benefit. Provided that the continuing technological progress has skills become more and more valuable, more education is certainly an answer to the technological change for the average individual.

One limitation of this paper is that we are not able to stratify the analysis by study subject. This is left for future work.

Acknowledgments

carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.

References

Acemoglu, D. and Johnson, S. (2007). Disease and Development: The Effect of Life Expectancy on Economic Growth. *Journal of Political Economy*, 115(6):925–985.

American Psychological Association (1995). Intelligence: Knowns and Unknowns, Report of a task force convened by the American Psychological Association.

Andersen, H. H., Mühlbacher, A., Nübling, M., Schupp, J., and Wagner, G. G. (2007). Computation of Standard Values for Physical and Mental Health Scale Scores Using the SOEP Version of SF-12v2. *Schmollers Jahrbuch: Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 127:171–182.

Anderson, J. (2007). *Cognitive Psychology and its Implications*. Worth Publishers, New York, 7 edition.

Banks, J. and Mazzonna, F. (2012). The Effect of Education on Old Age Cognitive Abilities: Evidence from a Regression Discontinuity Design. *The Economic Journal*, 122(560):418–448.

Barrow, L. and Malamud, O. (2015). Is college a worthwhile investment? *Annual Review of Economics*, 7:519–555.

Bartz, O. (2007). Expansion und Umbau – Hochschulreformen in der Bundesrepublik Deutschland zwischen 1964 und 1977. *Die Hochschule*, 2007(2):154–170.

Basu, A. (2011). Estimating decision-relevant comparative effects using instrumental variables. *Statistics in Biosciences*, 3:6–27.

Basu, A. (2014). Person-Centered Treatment (PeT) effects using instrumental variables: An application to evaluating prostate cancer treatments. *Journal of Applied Econometrics*, 29:671–691.

Basu, A., Heckman, J. J., Navarro-Lozano, S., and Urzua, S. (2007). Use of instrumental variables in the presence of heterogeneity and self-selection: an application to treatments of breast cancer patients. *Health Economics*, 16(11):1133–1157.

Blossfeld, H.-P., Roßbach, H.-G., and von Maurice, J. (2011). Education as a Lifelong Process – The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft*, 14:Special Issue.

Brinch, C. N., Mogstad, M., and Wiswall, M. (2012). Beyond LATE with a discrete instrument. Heterogeneity in the quantity-quality interaction of children. Discussion Papers 703, Research Department of Statistics Norway.

Card, D. (1995). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. In L. Christofides, K. G. and Swidinsky, R., editors, *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, pages 201–222. University of Toronto Press.

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

Carneiro, P., Heckman, J. J., and Vytlacil, E. (2010). Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin. *Econometrica*, 78(1):377–394.

Carneiro, P., Heckman, J. J., and Vytlacil, E. J. (2011). Estimating Marginal Returns to Education. *American Economic Review*, 101(6):2754–81.

Cawley, J., Heckman, J., and Vytlacil, E. (2001). Three observations on wages and measured cognitive ability. *Labour Economics*, 8(4):419–442.

Cervellati, M. and Sunde, U. (2005). Human Capital Formation, Life Expectancy, and the Process of Development. *American Economic Review*, 95(5):1653–1672.

Costa, D. L. (2015). Health and the economy in the united states from 1750 to the present. *Journal of Economic Literature*, 53(3):503–70.

Currie, J. and Moretti, E. (2003). Mother's Education And The Intergenerational Transmission Of Human Capital: Evidence From College Openings. *The Quarterly Journal of Economics*, 118(4):1495–1532.

Cutler, D. M. and Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of Health Economics*, 29(1):1 – 28.

de Walque, D. (2007). Does education affect smoking behaviors?: Evidence using the Vietnam draft as an instrument for college education. *Journal of Health Economics*, 26(5):877–895.

Ferrer, E., Salthouse, T., Stewart, W., and Schwartz, B. (2004). Modeling Age and Retest Processes in Longitudinal Studies of Cognitive Abilities. *Psychology and Aging*, 19(2):243–259.

Fisher, G., Stachowski, A., Infurna, F., Faul, J., Grosch, J., and Tetrick, L. (2014). Mental work demands, retirement, and longitudinal trajectories of cognitive functioning. *Journal of Occupational Health Psychology*, 19(2):231–242.

German Federal Government (various issues, 1973-1992). Unterrichtung durch die Bundesregierung – Bericht nach Paragraph 35 des Bundesausbildungsförderungsgesetzes zur Überprüfung der Bedarfssätze, Freibeträge sowie Vomhundertsätze und Höchstbeträge Paragraph 21 Abs. 2. Technical report, German Federal Government (Deutsche Bundesregierung), Berlin.

German Federal Statistical Office (2009). Wirtschaft und Statistik. Technical Report 2, German Federal Statistical Office (Statistisches Bundesamt), Wiesbaden.

German Federal Statistical Office (various issues, 1959–1991). Statistisches Jahrbuch für die Bundesrepublik Deutschland. Technical report, German Federal Statistical Office (Statistisches Bundesamt), Wiesbaden.

Glymour, M., Kawachi, I., Jencks, C., and Berkman, L. (2008). Does childhood schooling affect old age memory or mental status? Using state schooling laws as natural experiments. *Journal of Epidemiology and Community Health*, 62(6):532–537.

Grimard, F. and Parent, D. (2007). Education and smoking: Were Vietnam war draft avoiders also more likely to avoid smoking? *Journal of Health Economics*, 26(5):896–926.

Heckman, J. (1990). Varieties of selection bias. *The American Economic Review*, 80(2):pp. 313–318.

Heckman, J. J., Lochner, L. J., and Todd, P. E. (1999). Earnings Equations and Rates of Return: The Mincer Equation and Beyond. In Hanushek, E. and Welch, F., editors, *Handbook of the Economics of Education*, volume 1. Elsevier.

Heckman, J. J., Urzua, S., and Vytlacil, E. (2006). Understanding Instrumental Variables in Models with Essential Heterogeneity. *The Review of Economics and Statistics*, 88(3):389–432.

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

Heckman, J. J. and Vytlacil, E. J. (2007). Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 6 of *Handbook of Econometrics*, chapter 71. Elsevier.

Idler, E. and Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behaviour*, 38:21–37.

Jürges, H., Reinhold, S., and Salm, M. (2011). Does schooling affect health behavior? Evidence from the educational expansion in Western Germany. *Economics of Education Review*, 30(5):862–872.

Kamhöfer, D. and Schmitz, H. (2015). Reanalyzing Zero Returns to Education in Germany. *Journal of Applied Econometrics*, forthcoming.

Lang, F., Weiss, D., Stocker, A., and von Rosenbladt, B. (2007). The returns to cognitive abilities and personality traits in Germany. *Schmollers Jahrbuch: Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 127(1):183–192.

LIfBi (2011). Starting Cohort 6 Main Study 2010/11 (B67) Adults Information on the Competence Test. Technical report, Leibniz Institute for Educational Trajectories (LIfBi) – National Educational Panel Study.

LIfBi (2015). Startkohorte 6: Erwachsene (SC6) – Studienübersicht Wellen 1 bis 5. Technical report, Leibniz Institute for Educational Trajectories (LIfBi) – National Educational Panel Study.

Lindenberger, U. (2014). Human cognitive aging: Corriger la fortune? *Science*, 346(6209):572–578.

Mazumder, B. (2008). Does education improve health? A reexamination of the evidence from compulsory schooling laws. *Economic Perspectives*, (Q II):2–16.

Meng, X. and D'Arcy, C. (2012). Education and Dementia in the Context of the Cognitive Reserve Hypothesis: A Systematic Review with Meta-Analyses and Qualitative Analyses. *PLoS ONE*, 7(6):e38268.

Nybom, M. (2014). The Distribution of Lifetime Earnings Returns to College. Working Paper Series 2/2014, Swedish Institute for Social Research.

Oreopoulos, P. and Petronijevic, U. (2013). Making College Worth It: A Review of the Returns to Higher Education. *The Future of Children*, 23(1):41–65.

Oreopoulos, P. and Salvanes, K. (2011). Priceless: The Nonpecuniary Benefits of Schooling. *Journal of Economic Perspectives*, 25(1):159–84.

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

Robinson, P. M. (1988). Root- N-Consistent Semiparametric Regression. *Econometrica*, 56(4):931–954.

Rohwedder, S. and Willis, R. J. (2010). Mental Retirement. *Journal of Economic Perspectives*, 24(1):119–38.

Salthouse, T. A. (2004). What and When of Cognitive Aging. *Current Directions in Psychological Science*, 13(4):140–144.

Salthouse, T. A. (2006). Mental Exercise and Mental Aging: Evaluating the Validity of the "Use It or Lose It" Hypothesis. *Perspectives on Psychological Science*, 1(1):68–87.

Schneeweis, N., Skirbekk, V., and Winter-Ebmer, R. (2014). Does Education Improve Cognitive Performance Four Decades After School Completion? *Demography*, 51(2):619–643.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586.

Stephens, M. J. and Yang, D.-Y. (2014). Compulsory Education and the Benefits of Schooling. *American Economic Review*, 104(6):1777–92.

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society*, 2002(8):448–460.

Vytlacil, E. (2002). Independence, Monotonicity, and Latent Index Models: An Equivalence Result. *Econometrica*, 70(1):331–341.

Wagner, G., Frick, J., and Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. *Schmollers Jahrbuch: Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 127(1):139–169.

Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T., and Carstensen, C. (2011). Development of competencies across the life span. *Zeitschrift für Erziehungswissenschaft*, 14:67–86.

Weisser, A. (2005). 18. Juli 1961 – Entscheidung zur Gründung der Ruhr-Universität Bochum. Technical report, Internet-Portal Westfälische Geschichte, `http://www.westfaelische-geschichte.de/web495`.

Zimmermann, S., Artelt, C., and Weinert, S. (2014). The Assessment of Reading Speed in Adults and First-Year Students. Technical report, Leibniz Institute for Educational Trajectories (LIfBi) – National Educational Panel Study.

Appendix

Figures

Figure A1: Spatial variation of colleges across districts and over time

1958

1970

1980

1990
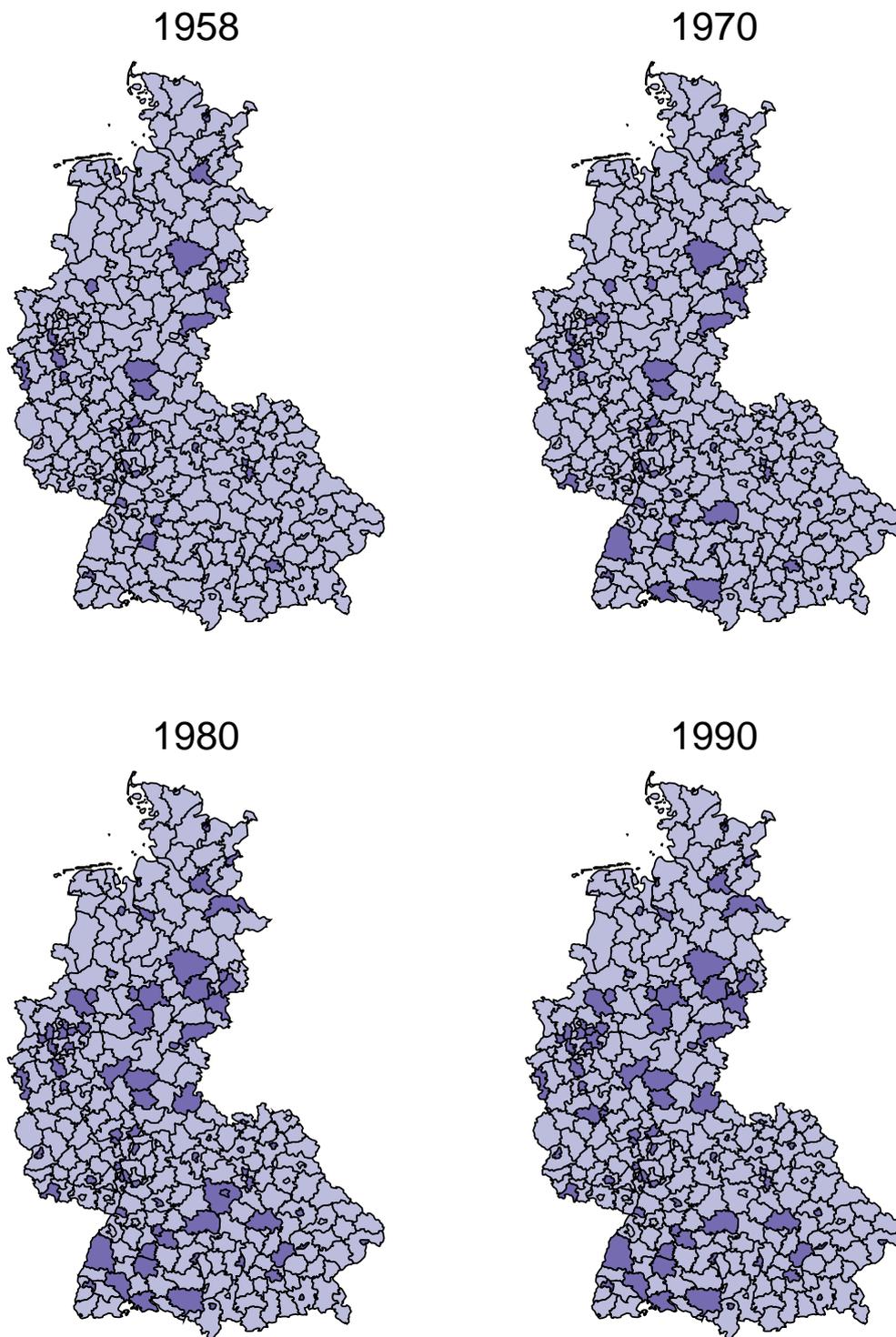
Notes: Own illustration based on the German Statistical Yearbooks 1959–1991 (German Federal Statistical Office, 1991). The maps show all 326 West German districts (*Kreise*, spatial units of 2009) but Berlin in the years 1958 (first year in the sample), 1970, 1980, and 1990 (last year in the sample). Districts usually cover a bigger city or some administratively connected villages. If a district has at least one college, the district is depicted darker. Only few districts have more than one college. For those districts the number of students is added up in the calculations but multiple colleges are not depicted separately in the maps.

Figure A2: Average distance to the closest college over time for districts with a college opening
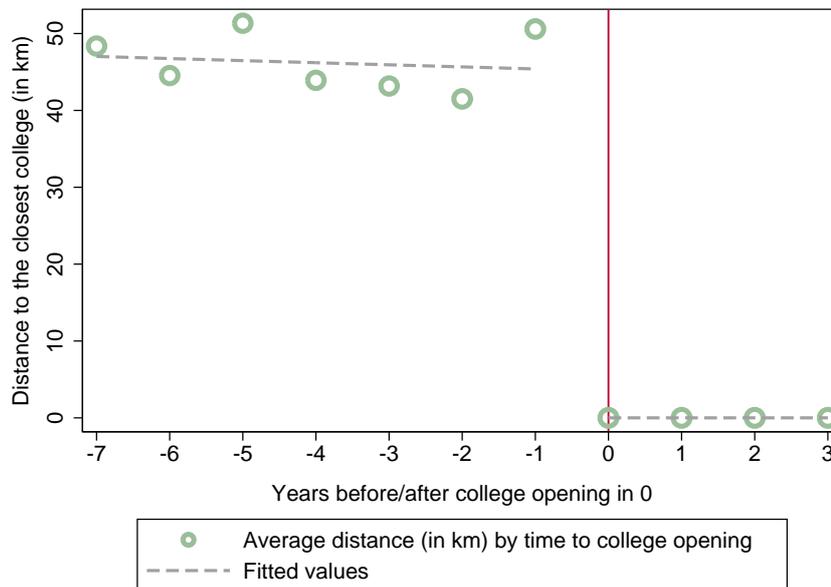
Notes: Own illustration. Information on colleges are taken from the German Statistical Yearbooks 1959–1991 (German Federal Statistical Office, 1991). The distances (in km) between the districts are calculated using district centroids. These distances are weighted by the number of individuals observed in the particular district-year cells in our estimation sample of the NEPS-Starting Cohort 6 data. The resulting average distances are depicted by green circles. Note that prior to time period 0, the average distance changes over time either due to sample composition or a college opening in a neighboring district. Only districts with a college opening are taken into account.

Tables

Table A1: Control variables and means by university degree

Variable	Definition	Respondents with univ. degree	w/o univ. degree
General information			
Female	=1 if respondent is female	41.90	53.61
Year of birth (FE)	Year of birth of the respondent	1959	1959
Migrational background	=1 if respondent was born abroad	0.52	0.10
No native speaker	=1 if mother tongue is not German	0.10	0.22
Mother still alive	=1 if mother is still alive in 2009/10	65.48	59.97
Father still alive	=1 if father is still alive in 2009/10	44.81	38.73
Pre-college living conditions			
Married before college	=1 if respondent got married before the year of the college decision or in the same year	2.05	2.42
Parent before college	=1 if respondent became a parent before the year of the college decision or in the same year	1.19	0.88
Siblings	Number of siblings	1.55	1.85
First born	=1 if respondent was the first born in the family	35.19	27.97
Age 15: lived by single parent	=1 if respondent was raised by single parent	4.43	6.25
Age 15: lived in patchwork family	=1 if respondent was raised in a patchwork family	1.43	2.79
Age 15: orphan	=1 if respondent was a orphan at the age of 15	0.81	2.08
Age 15: rural district	=1 if district at the age of 15 was rural	19.67	26.16
Age 15: mother employed	=1 if mother was employed at the respondent's age of 15	42.76	45.30
Age 15: mother never unemployed	=1 if mother was never unemployed until the respondent's age of 15	59.90	59.26
Age 15: father employed	=1 if father was employed at the respondent's age of 15	93.95	89.57
Age 15: father never unemployed	=1 if father was never unemployed until the respondent's age of 15	98.67	96.41
Military	=1 if respondent did military service	31.76	26.95
Pre-college health and education			
Final school grade: excellence	=1 if the overall grade of the highest school degree was excellent	3.05	1.03
Final school grade: good	=1 if the overall grade of the highest school degree was good	23.29	15.51
Final school grade: satisfactory	=1 if the overall grade of the highest school degree was satisfactory	14.48	18.35
Final school grade: sufficient or worse	=1 if the overall grade of the highest school degree was sufficient or worse	1.00	0.95
Repeated one grade	=1 if student needed to repeat one grade in elementary or secondary school	20.57	19.47
Repeated two or more grades	=1 if student needed to repeat two or more grades in elementary or secondary school	2.10	1.80
Parental characteristics (M: Mother, F: Father)			
M: year of birth (FE)	Year of birth of the respondent's mother	1931	1930
M: migrational background	=1 if mother was born abroad	4.67	4.34
M: at least inter. edu	=1 if mother has at least an intermediate secondary school degree	25.48	9.01
M: vocational training	=1 if mother's highest degree is vocational training	27.38	25.40
M: further job qualification	=1 if mother has further job qualification (e.g., *Meister* degree)	4.81	2.38
F: year of birth (FE)	Year of birth of the respondent's father	1927	1927

Continued on next page

Table A1 – *continued*

Variable	Definition	Respondents with univ. degree	Respondents w/o univ. degree
F: migrational background	=1 if father was born abroad	5.67	4.78
F: at least inter. edu	=1 if father has at least an intermediate secondary school degree	29.43	11.73
F: vocational training	=1 if father's highest degree is vocational training	23.52	34.22
F: further job qualification	=1 if father has further job qualification (e.g., *Meister* degree)	13.14	10.56
Number of observations (health satisfaction sample)		2,100	5,918

Notes: Information taken from NEPS–Starting Cohort 6. Mean values refer to the health satisfaction sample. In the case of binary variables, the mean gives the percentage of 1s. FE = variable values are included as fixed effects in the analysis. [a] Only available for males who did military eligibility test (2,359 observations).

Table A2: Linear and non-linear selection equations using the instruments separately

	(1)	(2)	(3)	(4)	(5)	(6)
	Cognitive ability component			Health measure		
	Read. speed	Read. comp.	Math liter.	PCS	MCS	Health satis.
Panel A: 2SLS first-stage results for college availability						
College availability	1.920***	1.851***	1.943***	1.985***	1.985***	1.858***
	(0.095)	(0.089)	(0.113)	(0.085)	(0.085)	(0.060)
F statistic instr.	406.64	436.68	294.59	541.64	541.64	953.90
Panel C: 2SLS first-stage results for BAfoeG eligibility						
BAfoeG eligibility	0.017***	0.018***	0.017***	0.013***	0.013***	0.015***
	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)
F statistic instr.	41.64	63.28	27.67	42.91	42.91	117.08
Panel C: Logit results for both instruments (non-linear version of Panel B, Table 3)						
College availability	1.249***	1.182***	1.266***	1.323***	1.323***	1.196***
	(0.073)	(0.068)	(0.085)	(0.066)	(0.066)	(0.048)
BAfoeG	0.019***	0.020***	0.021***	0.013***	0.013***	0.014***
	(0.003)	(0.002)	(0.004)	(0.002)	(0.002)	(0.001)
Panel D: Logit results for college availability						
College availability	1.333***	1.278***	1.357***	1.380***	1.380***	1.268***
	(0.074)	(0.069)	(0.086)	(0.067)	(0.067)	(0.049)
Panel E: Logit results for BAfoeG eligibility						
BAfoeG	0.027***	0.026***	0.029***	0.018***	0.018***	0.019***
	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)
Number of observations	3,559	4,116	2,587	4,304	4,304	8,018

Notes: The table reports marginal effects. Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Regressions also include year-of-birth and federal-state fixed effects.