Short- and long-term effects of vocational versus general schooling: A regression discontinuity approach*

Jan Sauermann[a] Anders Stenberg[b]

April 1, 2017

Abstract

In this paper, we estimate the relative earnings returns for Swedish upper secondary schooling programs with vocational and general contents for individuals born 1955-75. Using data on individuals' application choices, we address endogeneity of program choice by exploiting admission cutoffs. These cutoffs, which are based on pupils' grade point averages (GPA), allow us to estimate the returns to educational contents in a regression discontinuity (RD) framework. In addition, we control for students' choice order to proxy for students' comparative advantages. We apply this RD design to estimate the effects of schooling contents on annual earnings from 1978-2014. For men, we find that vocational schooling has an earnings advantage over general schooling up and until the age of 27, with estimates thereafter close to zero. For women, we observe a similar earnings advantage for vocational schooling up and until the age of 23. For the longer term, there are no coherent earnings differences for men. For women with vocational schooling, there is a renewed earnings advantage from age 43 and onwards. We find suggestive evidence that this can be explained by further education completed in adulthood.

JEL-codes: J24, J31, I24
Keywords: Human capital, vocational education, regression discontinuity design, comparative advantages

[a] Swedish Institute for Social Research (SOFI), Stockholm University, 10691 Stockholm, Sweden; e-mail: jan.sauermann@sofi.su.se; Research Centre for Education and the Labour Market (ROA), Maastricht University, The Netherlands; Institute for the Study of Labor (IZA), Bonn, Germany.
[b] Swedish Institute for Social Research (SOFI), Stockholm University, 10691 Stockholm, Sweden; e-mail: anders.stenberg@sofi.su.se; Linnaeus University, Växjö/Kalmar, Sweden; Institute for the Study of Labor (IZA), Bonn, Germany.

*The authors gratefully acknowledge financial support from the Swedish Research Council (Grant number Dnr 2011-2613; Anders Stenberg) and the Jan Wallanders och Tom Hedelius Stiftelse for financial support (Grant number I2011-0345:1; Jan Sauermann). We also thank Magnus Carlsson, Karin Edmark, Kristian Koerselman, Rasmus Landersø, Matthew Lindquist, Michael Maier, Ferran Elias Moreno, Dan-Olof Rooth, Bettina Siflinger, and seminar and conference participants at the University of Copenhagen, CVER Workshop, the Education, Skills, and Labor Market Outcomes workshop in Trondheim, ESPE 2016 Berlin, IFN Stockholm, Maastricht University, ZEW Mannheim, Rockwool Fonden, and SOFI Stockholm for valuable comments and suggestions.

1 Introduction

There are large differences in how vocational and general contents are prioritized and mixed in different educational systems. Despite an enormous literature on the returns to years of schooling, and despite the widely accepted view that educational contents are important, there are only a handful of studies that address the long-run impact of vocational relative to general contents on labor market outcomes (Malamud and Pop-Eleches, 2010; Brunello and Rocco, 2015; Dustmann et al., 2016; Golsteyn and Stenberg, 2016; Hanushek et al., 2016). For a given number of schooling years, vocational and general schooling are typically linked with the following short-term and long-term hypotheses concerning earnings and employment.[1] First, the aim of vocational education is to prepare for specific occupations. It should therefore lead to a better school-to-work-transition, and possibly to lower unemployment and/or higher labor earnings in the short run. Second, general education is assumed to generate knowledge that enhances the ability to learn. It could therefore create a skill multiplier effect that allows workers to adapt more readily to new technologies during their career, resulting in long run earnings advantages. In sum, these hypotheses suggest that as careers progress, the earnings of individuals with general education could catch up and even exceed the earnings of those with vocational education.

The aim of this article is to estimate the causal earnings effects of vocational schooling relative to general schooling by exploiting data from applications to upper secondary school. After nine years of compulsory schooling, individuals can apply for a range of 2-year upper secondary programs that differ in their degree to which they contain vocational and general contents. We employ a regression discontinuity (RD) design based on grade point average (GPA)-cutoffs, which decide whether an individual gets into a program, or not (see, e.g., Imbens and Lemieux, 2008; Abdulkadiroğlu et al., 2014; Kirkebøen et al.,

[1] It is often claimed, though not formalized into a model, that the relative earnings impacts of vocational and general educations are different in the short and long term. See, e.g., Shavit and Müller (1998); Rosenbaum (2001); Ryan (2001); Korpi et al. (2003); Krueger and Kumar (2004a,b); Wolter and Ryan (2011); Brunello and Rocco (2015); Malamud and Pop-Eleches (2010); Dustmann et al. (2016); Golsteyn and Stenberg (2016); Hanushek et al. (2016).

2016). In addition, we follow Kirkebøen et al. (2016) to control for choice combinations as a proxy for individuals' comparative advantages. Our estimates are thus identified by the discontinuity in the assignment to programs within groups of individuals with the same type of preferred and second-best choices. The analyses are based on Swedish register data of individuals enrolled between 1971 and 1991, who attended a schooling system which in large parts is still in place today. We estimate the effects on annual earnings between 1978 and 2014, when individuals in our sample are between 18 and 59 years old. This framework allows us to test the theoretical hypotheses regarding the short-term and the long-term relative effects of vocational and general educational contents.

The Swedish upper secondary system sees more than one third of a cohort choose a vocational track from age 16. It makes the diversity greater than in the US, the UK or southern Europe, but smaller than in Germany and many other European countries. The institutional background of this study is comparable to a hypothetical change in the US system if it were to be made more flexible. Albeit one year shorter than the US high school, $9 + 2$ years, the 2-year general programs are similar to the completion of high-school and the 2-year vocational programs may be seen as a potential counterfactual. The comparison primarily concerns individuals below the top quartile of the ability distribution.

Earlier studies estimating the effects of vocational and general educational contents have foremost considered short term outcomes. These indicate earnings advantages of vocational contents (Kemple and Scott-Clayton, 2004; Kemple and Willner, 2008). Regarding long-term evidence, studies analyzing reforms which expanded comprehensive schooling typically find positive effects on labor market outcomes, whereas reforms expanding vocational schooling have been linked with no statistically significant effects.[2] While these results are taken from different contexts, they may still suggest that earnings impacts of schooling differ depending on educational contents.[3]

[2]For effects on comprehensive schooling, see Angrist and Krueger (1991), Harmon and Walker (1995), Meghir and Palme (2005), and Aakvik et al. (2010), and for expansions of vocational schooling see Oosterbeek and Webbink (2007), Pischke and von Wachter (2008) and Hall (2012).

[3]Hall (2012) analyzes a pilot scheme in Sweden from 1987 to 1991, where 2-year vocational programs at upper secondary school in some regions were extended to three years. The 2-year vocational programs

Studies directly comparing the relative long term outcomes of vocational versus general educational contents, include two contributions which exploit a random element in the assignment to schooling programs (Malamud and Pop-Eleches, 2010; Dustmann et al., 2016), whereas three other studies are based on selection on observables strategies (Brunello and Rocco, 2015; Hanushek et al., 2016; Golsteyn and Stenberg, 2016).

Malamud and Pop-Eleches (2010) exploit an secondary school reform in Romania. Compared to older chorts, individuals enrolled in 1973 or later were required to study an additional two years of general education before entering vocational schools. The later cohorts therefore received more general, and less vocational education. They compare adjacent cohorts from census data (1992, 2002) and household survey data (1995, 2000), when individuals were 33-44 years old. An attractive feature of the study is that the framework allows for a comparison between individuals who could not choose between different regimes. The authors are therefore able to compare average outcomes of comparable individuals exposed to different educational contents, providing a very clear setting. Their findings indicate that the occupational composition was affected, with less manual workers and craftsmen in the later cohort, but they found no statistically significant differences in family income or wages. While this is convincing evidence, the case of Romania may not generalize to other countries e.g. if the compressed wage structure of the communist regime affected wage setting in the following decade. Also, a relative long-term advantage of general education may be that it enhances learning capabilities and increases the likelihood of receiving on-the-job training. According to OECD (2004), Romania had the lowest incidence of on-the-job training in Europe, on average seven hours per year.

For individuals born in West Germany 1961-1976, Dustmann et al. (2016) evaluate the long term labor market impact of enrolment into different tracks, one of which contains

are identical to those studied here, with about 2-4 hours of general subjects per week, and the majority of hours devoted to vocational in-class training (no apprenticeships). These are compared with the 3-year pilot programs which extended the number of hours of general subjects to 6-7 hours per week, and the third year included 60% apprenticeship training. The insignificant results reported may be interpreted as the effects of an additional year of vocational school. However, the design of the pilot scheme makes it difficult to set it in direct relation to the question of educational contents, since both general and vocational education was increased.

vocational education. Information on birth dates is used to compare individuals around the cutoff date of school cohorts. Their evidence shows that individuals in different tracks, contrary to expecation, do not differ in terms of educational achievement. Consequently, labor market outcomes are not significantly different for individuals when aged between 30 and 45. The results suggest that the built-in flexibility of the tracking system allowed individuals to correct initial choices, so that high (low) performing individuals switched to more (less) advanced tracks after the initial tracking.

Hanushek et al. (2016) evaluate an average impact on labor market outcomes using data from 11 countries included in the International Literacy Survey (IALS). Controlling for cognitive skills, country fixed effects, years of schooling and parental background, their findings indicate an initial advantage in employment probabilities for individuals with vocational education, which diminishes with age. These results imply a short-term advantage which gradually declines. Although cross-sectional earnings information are not necessarily a good proxy for life-cycle earnings (Card and Lemieux, 2001), their study provides evidence of a converging pattern in labor market outcomes. This finding is supported by Brunello and Rocco (2015), using UK cohort data. These studies provide important contributions in relation to the findings in Malamud and Pop-Eleches (2010), who estimated earnings for males aged at least 33 years old, which thereby could hide short-term differences in outcomes.

One limitiation of all of these studies is that their samples are either limited to men (Malamud and Pop-Eleches, 2010; Brunello and Rocco, 2015; Hanushek et al., 2016), or that they do not report results separately by gender Dustmann et al. (2016). The validity of the results may not be possible to generalize across gender groups since women tend to choose educational paths and working careers that are very different from those of men.

One study that does analyze men and women separately is Golsteyn and Stenberg (2016), using a selection on observables approach with data from Sweden. They find support for a short-term earnings advantage of vocational contents, for both men and women. In contrast to the other studies, they also find statistically significant estimates

in the long term, which support a relative earnings advantage of general education. Although there are substantial gender differences in program choice, these results hold for both men and women. Additional analyses indicate that neither result can be explained by further education or the presence of children. Golsteyn and Stenberg (2016) use a subset of the data of the present study, but with a very different approach to identify the estimated relative effects of vocational and general educational contents.[4] Given that the unconfoundedness assumption holds, i.e. that unobserved factors are controlled for by family background and GPA, the results in Golsteyn and Stenberg (2016) may be interpreted as average treatment effects (ATE). The present study, in contrast, aims at estimating the local average treatment effect (LATE) for individuals on the margin between different educational types. The estimated ATE from Golsteyn and Stenberg (2016) may differ from LATE if the unconfoundedness assumption does not hold, and/or if there are heterogeneous treatment effects. Both estimators are interesting, although they have different implications from a policy perspective. The ATE might imply that the type of education which enhances productivity should be given more space in the curricula. The LATE, on the other hand, indicates if and how one would like to allocate marginal individuals to tracks which enhance production, e.g. by increasing the number of slots for these tracks.

The main contribution of this study is to present causal evidence of the relative returns of vocational and general schooling, for the short term and the long term, and for both men and women. To the best of our knowledge, this is the first causal evidence based on uninterrupted earnings trajectories, covering a large part of the individuals' working lives, to test the alleged trade-off between school to work transitions and long-term outcomes. We find that both males and females who have followed vocational schooling, as opposed to general schooling, have an earnings advantage up to age 27 (men), and 23 (women). We provide evidence that these results are not due to differences in further education after upper secondary schooling. This part of the results are in line

[4]Golsteyn and Stenberg (2016) use samples of all individuals enrolled in vocational or general 2-year programs from 1971 to 1979, analyzing uninterrupted earnings data from age 18 until 52 (1978-2011).

with those of previous studies. In the long term, we find no statistically significant earnings differences between the two types of schooling for men. Women with vocational schooling experience a renewed earnings advantage from age 43 and onwards. We find suggestive evidence that this can be explained by further education completed in adulthood.

This paper is structured as follows. In the following section, we explain the institutional setting and application process in detail and presents the data used in this study. Section 3 describes our estimation strategy to identify the relative effects of educational contents. Section 4 presents results and Section 5 concludes.

2 Institutional setting, data and sample

This study relies on two types of data. First, we use data containing information on all applications for upper secondary schooling programs in Sweden from 1971 to 1991. These data are used to identify application choice, track assignments and GPA threshold values, also referred to as the cutoff, which determine acceptance to programs. Second, data on annual earnings from 1978 to 2014 are used to estimate the effects of assignment to program on labor market outcomes. Annual earnings include all income from labor, without transfers, and are expressed in 2014 values.

2.1 Upper secondary schooling and program types

To analyze the relative returns of vocational versus general schooling, we focus on enrollees in Swedish upper secondary school between 1971 and 1991. In the Swedish system, children attend compulsory (primary) school for nine years, starting from the year they turn seven years old. After the ninth grade, individuals can leave school or continue in upper secondary school, either in 2-year vocational programs, 2-year general programs or 3-year academic programs, i.e. from grade 10 to 11 or from grade 10 to 12. To compare educational contents, we focus on the 2-year programs only. These programs attracted

individuals from similar social backgrounds with relatively low participation in further education.

The decades preceding the 1970's was a period of intense schooling reforms. First, an important compulsory school reform was implemented between 1949 and 1962, extending the minimum number of schooling years from seven to nine (Meghir and Palme, 2005). During this period, there was an overall increase in the demand for upper secondary schooling. The share of a cohort enrolled in upper secondary schooling (*gymnasium* which then only included the academic tracks) increased between 1950 and 1960 from 4% to 20%. The vocational educations (*yrkesskola*) also expanded fast, and the number of slots was tripled in this period. In 1962, the government established 2-year general programs (*fackskola*), as a response to the increased demand for prolonged educations. These programs were less specific than the vocational programs but more specific than the academic tracks and the compulsory school.

The *integrated gymnasium*, launched in 1971, was the final step in a process of intense reforms. This is what we refer to as upper secondary school in this paper. It merged three different forms of schooling (*fackskola, yrkesskola* and the old *gymnasium*) into a cohesive institution which offered 14 different 2-year vocational programs (e.g., nursing, electronics or construction), three different 2-year general programs (social studies, business, engineering) and five general 3-year programs (social studies, business, human arts, engineering, science).[5] Curricula from the most popular vocational and general 2-year programs are presented in Table 1. There is a clear distinction between vocational and general programs in that the majority of hours at vocational programs are devoted to a particular profession. Such subjects constitute zero hours in the curricula of the general programs.

The completion of 3-year programs qualified individuals for university, whereas both vocational and general 2-year programs only provide eligibility for short college educations,

[5]For vocational schooling, 3-year programs were introduced via a pilot scheme in 1987 (cf. Hall, 2012), before being more broadly implemented from 1992 onwards (cf. Footnote 3). The 2-year general programs were then also abolished.

classified as tertiary level schooling, in fields such as physiotherapy, nursing and pre-school teaching. Another possibility to return school is to attend adult education at the level of upper secondary school (*komvux*), which municipalities are obliged by law to offer. Individuals may complete courses at *komvux* to qualify for college education (Stenberg, 2011). This would also be of interest for drop outs who wish to complement their schooling.

2.2 Assignment to programs

Throughout the period of program enrolment relevant for this study, individuals stated up to six ordered choices when applying for upper secondary programs during ninth grade. Individuals were in principle only allowed to apply in their region of residence, unless the preferred program was not offered in the home region.[6] There were between 100 and 120 high-school regions each year (*gymnasieregion*) with an average of 591 applicants per cohort. On average, individuals applied for 2.8 programs (cf. Figure 1).

For each program offered in a region in a specific year, individuals competed for slots with their GPA from ninth grade. The GPA is based on 10-12 subject grades set on a relative scale from one (lowest) to five (highest). Centrally corrected exams were used as guidelines to teachers' grading levels. The GPAs given by the teachers were supposed to be normally distributed with mean 3.0, but grade inflation makes it closer to 3.2 in a given year.[7]

The assignment to programs follows an iterative process. In a first step, individuals are assigned to the program of their first choice and are ranked according to their application GPA. If the number of applicants is smaller or equal to the number of slots available, all applicants are accepted. If the number of applicants exceeds the number of slots, a preliminary threshold is defined. All individuals with a GPA below this threshold

[6] About 10% applied for programs not offered in their region of residence.

[7] When handing in the first application to upper secondary schooling, individuals apply with their GPA obtained at the end of the autumn semester. This information is used by the administration to make a preliminary assignment, which is later refined in June when schools report the final GPA to the admission centers.

are moved to their second choice, whereas individuals above the threshold are preliminarily placed in this choice. After moving low GPA individuals to their second choice, individuals in each program-region combination are again ranked according to their GPA and assigned to their third choice if the GPA is lower than the adjusted GPA cutoff. This process is repeated until all options are exhausted. Individuals are subsequently sent a letter of acceptance, which they need to confirm, or they will lose their slot.

2.3 GPA cutoffs to identify causal effects of track choice

GPA thresholds, i.e. the minimum GPA which is required to get into a specific program, are not systematically collected but need to be constructed from the data at hand. In a perfectly clean allocation process there would be a sharp discontinuity, where everybody above the GPA threshold is admitted to the program, and everybody below the threshold is not. The algorithm used to determine the threshold GPA is based on a simple rule. GPA is defined and measured with increments of 0.1. For each decimal GPA unit k from $k = 1.0, 1.1, \ldots, 4.9, 5.0$, the GPA unit k with the greatest difference between the shares accepted ($GPA > k$) and the share non-accepted, with $GPA < k$, constitutes the threshold $k^* = GPA^{Th}$:

$$k^* = argmax(k) = \frac{\sum_{k=k+.1}^{5.0} I_{i,GPA=k}^{acc}}{\sum_{k=k+.1}^{5.0} I_{i,GPA=k}} - \frac{\sum_{k=1.0}^{k-.1} I_{i,GPA=k}^{non}}{\sum_{k=1.0}^{k-.1} I_{i,GPA=k}} \tag{1}$$

Thus, the first term of Equation(1) is the ratio between the number of individuals *accepted* above the hypothetical threshold k, $\sum_{k=k+.1}^{5.0} I_{i,GPA=k}^{acc}$, and the *total number* of individuals who applied with a GPA above k, $\sum_{k=k+.1}^{5.0} I_{i,GPA=k}$. If assignment perfectly follows this simple rule, everybody above k^* is accepted, and this share is equal to one. The second term of Equation (1) is the ratio of *non-accepted* individuals below the threshold k, $\sum_{k=1.0}^{k-.1} I_{i,GPA=k}^{non}$ and the *total number* of individuals who applied with a GPA below threshold k. In the perfect case, this share is zero.

To assess the effectiveness of the defined threshold, the distance between an individual's GPA and the GPA threshold can be written as

$$dist_i^{Th} = GPA_i - GPA^{Th} \tag{2}$$

Figure 2 shows the share of accepted individuals around the calculated GPA cutoff, which is normalized to zero. The figure shows that the discontinuity is fuzzy around the cutoff, but the share of individuals accepted being relatively small (large) to the left (right) of the cutoff. Thus, a small change in GPA around the cutoff is linked to a large change in the probability of being accepted to a program, or not. As the figure shows, there are individuals below the threshold who are enroled and individuals above the threshold who are not enroled. There are several reasons for this. First, headmasters have some discretion in assigning students to programs. Second, an individual who changes her mind after being accepted, is recorded as non-accepted. This is because data does not contain information on whether an individual received an acceptance letter or not, but rather who was present two weeks after the start of the semester. Third, our algorithm to find a cutoff value may contain measurement error.

2.4 Sample of this study

Applicants were accepted to either their first or their second choice in 94.8% of the cases. We focus on the first two choices individuals make, and define the first choice as the preferred field, and the second choice as the next best alternative. To compare similar individuals with assignment to different programs, we restrict the sample to individuals who have either chosen a vocational program as their first choice and a general program as a second choice, or vice versa. This allows us to compare individuals with similar combinations of choices.

To compare similar individuals with different program type assigments, we restrict the sample to individuals who have either chosen a vocational program as their first

choice and a general program as a second choice, or vice versa. This allows us to compare individuals with similar combinations of choices. Applicants were accepted to either their first or their second choice in 94.8% of the cases. We therefore apply a simple rule to define the preferred field as the first choice, and the next best alternative as the second choice. Table 2 shows the program combinations observed in our data. The most common application (58%) states the same type of education as first and second choice (general-general or vocational-vocational). In total, 7.9% choose program combinations which are either vocational-general (4.5%) or general-vocational (3.5%).[8] We also condition individuals to apply directly from compulsory school at age 16, not to have received minority gender bonus points, and that they were accepted to either their first or their second choice.[9] Out of 1,522,959 applicants spanning over 21 years, we retain 120,313 observations with program combinations, which are either vocational-general, or general-vocational.

For our RD design, we need to compare individuals at the margin of becoming accepted to their preferred choice. We therefore need to condition that there is a threshold defined for the individuals' applied first choice. Our algorithm described in the previous section defines thresholds, some of which are of low quality if the number of applicants for a track only slightly exceeds the number of slots, or if there is a high share of accepted below k^* or a high share of non-accepted above k^*. We remove a quarter of the thresholds with the lowest quality, i.e. where the values of k^* are below the 25th percentile, resulting in 72,608 remaining observations.

To give an idea about what kind of selection these marginal individuals constitute, Table 3 presents descriptive statistics for the total sample as well as a sample consisting of only participants in 2-year programs. Table 4 displays the averages of the same variables for the estimation sample by gender and program type.

[8] 27% do not state a second choice, presumably as they were certain of being accepted to their first choice.

[9] Online Appendix A.1 describes details of the institutional design over the period 1971-1991, and how it has affects the data restrictions which vary slightly across the period.

3 Estimation strategy

We are interested in estimating the effects of vocational schooling, relative to general schooling, on annual earnings. To this end, the following equation can be estimated:

$$y_{i,t} = \gamma_1 + \gamma_2 \cdot voc_i + \varepsilon_{i,t} \tag{3}$$

where $y_{i,t}$ denotes annual earnings of individual i in year t, and where voc_i is a dummy for being enroled in a vocational program ($voc_i = 1$), or a general program ($voc_i = 0$). An estimate of γ_2 from an OLS regression, however, may be biased due to unobserved characteristics which vary systematically with program assignment. To address this selection problem, we employ a regression discontinuity (RD) design where GPA thresholds are assumed to create a random element in whether an individual is accepted to a vocational program, or not. RD designs in which GPA thresholds are used to identify causal effects of education has been used to study a wide range of outcomes.[10]

Within the RD framework, the variable associated with a discontinuous probability of treatment is commonly referred to as the running variable. If we limit our interest to individuals who have a vocational program as their first choice, $dist_i^{Th}$ would be our running variable. In this study, however, we consider two kinds of applicants. One with a vocational program as the first choice and with a general program as the second choice, while other applicants have a general program as the first choice and a vocational program as the second choice.

In order to make the interpretation of the running variable x_i consistent, we normalize its value to indicate an increased probability of being assigned a vocational track.[11]

[10] Examples include school choice and school performance (Clark, 2010; Jackson, 2010; Duflo et al., 2011; Pop-Eleches and Urquiola, 2013; Abdulkadiroğlu et al., 2014), marriage market outcomes (Kaufmann et al., 2015), political participation (Solis, 2015), and annual earnings returns (Kirkebøen et al., 2016).

[11] For simplicity, we here supress subscripts t for year and j for program.

$$x_i = \begin{cases} dist_i^{Th}, & \text{if vocational is first choice.} \\ -dist_i^{Th} - 0.1, & \text{if general is first choice .} \end{cases} \tag{4}$$

Thus, the running variable x_i is equal to $dist_i^{Th}$ if a vocational program is the first choice. If general program is first choice, the running variable is $-dist_i^{Th} - 0.1$. The the probability of being assigned to a vocational program will then in all cases increase discontinously for values between -0.1 and 0.0. The key assumption is that the cutoff $x_i = 0$ is associated with a discrete jump in the probability of enrolment in a vocational program. Panel (a) of Figure 3 shows the share of individuals in vocational studies across different values of the running variable x_i.

One concern is that individuals may manipulate their GPA value around the cutoff. This could be the case if they are aware of the threshold value in advance, and only perform sufficiently to be accepted into the program of their preference. In this case, the GPA may be a misleading predictor of ability, and one might expect an overrepresentation of individuals just to the right of the cutoff. As a first indication of this non-manipulation assumption, Panel (b) of Figure 3 shows that the density of the distribution around the GPA threshold is smooth.[12]

The model we estimate is a two stage least squares, where the excluded instrument affecting assignment to type of program is the indicator variable $z_i = 1$ if $x_i \geq 0$, and zero otherwise.

$$y_{i,t} = \beta_1 + \beta_2 \cdot \widehat{voc_i} + \beta_3 \cdot W_i + f(x_i) + \theta_j + \lambda_t + \nu_{i,t} \tag{5}$$

$$voc_{i,t} = \alpha_1 + \alpha_2 \cdot z_i + \alpha_3 \cdot W_i + f(x_i) + \theta_j + \lambda_t + \varepsilon_{i,t} \tag{6}$$

Equation (5) is the second stage equation and Equation (6) is the first stage, and where $f(x_i)$ is a flexible function which allows for different slopes on each side of the cutoff

[12]In principle, one does not need continuity of the density at the cutoff, but the discontinuity makes it possible that the non-manipulation assumption is violated (Imbens and Lemieux, 2008).

and for each group of applicants, i.e. whether they have vocational or general as their first choice.

The RD design may be regarded as an instrumental variable (IV) approach. The parameter of main interest is our estimate of β_2 which then identifies the local average treatment effect (LATE) estimated at the cutoff $x_i = 0$. The matrix W_i includes controls for individual characteristics: indicator variables for being adopted, foreign born, whether father is foreign born or mother is foreign born, if father or mother is low skilled (no upper secondary school) or high skilled (post-secondary education). θ_j is an indicator variable taking the value one if the preferred choice is a general program and cohort-specific effects are indicated by λ_t.[13] To give individuals far from the cutoff less influence on our estimates, we employ triangular weights that decrease smoothly as the distance to the cutoff point increases (Imbens and Lemieux, 2008), using a default bandwidth of 1.0, which we vary in robustness checks. We run these regressions separately for each age, which means that λ_t also controls for year specific effects.

Sorting around the threshold would presume that individuals have some information on the value of the GPA threshold. Although students may have information on thresholds in previous years, the exact cutoff is unknown *ex ante*. For the period of observation, we find that thresholds vary from year to year by 0.28 GPA points.

For the credibility of the RD strategy, it is vital that the discontinuity in the running variable is not associated with discontinuous changes in other background characteristics. Such a discontinuity would be an indication of sorting around the cutoff value. Therefore, pre-determined covariates, which by definition are unaffected by treatment, should not display discontinuous changes around the cutoff value. Figure 5 shows shares of mothers and fathers defined as low skilled and high skilled respectively. Similarly, the GPA, which is underlying the assignment process should also be smooth around the cutoff. Figure 6 demonstrates that there is no discontinuous increase in GPA levels, providing additional support that there is no sorting around the threshold.

[13] We do not include parental earnings in W_i in Equation (5) and Equation (6) because it would generate a large loss of observations without much gain in explanatory power.

Another approach to test the unconfoundedness assumptions is to test the null hypothesis of a zero average treatment effect on pre-determined variables (Lee et al., 2004; Imbens and Lemieux, 2008). We have access to 16 different pre-determined variables which by definition are not affected by the treatment. If inserted as outcome variables in Equation (5), one would expect estimates of β_2 to be statistically insignificant. For each of these 16 variables, regressions are estimated without the matrix W_i as its variables are instead used one by one as outcomes.[14] This is done separately for males and females, and we also use eight different bandwidths from 1.00 to 0.30. In total, this yields 256 coefficients ($8 \cdot 2 \cdot 16 = 256$). These tests yield in total 11 (4.3%) coefficients that are statistically significant from zero at the five percent level.

4 Results

4.1 Main results

Our working hypotheses are that earnings of individuals with vocational education will be higher in short term but that differences are evened out later in life. As a first check, it is interesting to see if the discontinuity in enrolment probability yields any form of discontinuous change in our outcome variable. Panel (a) of Figure 7 shows average annual earnings of individuals when aged 20. One may indeed discern a discontinuity in earnings at the cutoff point, whereas the relation is almost flat at other values of the running variable. The corresponding averages for earnings at age 40 are presented in 7, Panel (b). The pattern is now quite different, with earnings visibly increasing with the running variable, but without a discontinuous jump in earnings at the cutoff. This rudimentary analysis suggests that there is a short term earnings advantage of those enrolled in vocational

[14]The 16 variables are indicator variables for being adopted, foreign born, whether father is foreign born or mother is foreign born, if father or mother is low skilled (no upper secondary school) or high skilled (post-secondary education), mother's and father's birthyear, mother's and father's annual earnings and annual log earnings when individual was 16 years old.

programs, whereas there are no apparent earnings differences between program types at age 40.

Figures 8 and 9 show the main estimation results for individuals at the margin between program types. From ages 19 to 58, each point represents an estimate from a separate regression of being enrolled in a vocational program compared with a general program. The vertical lines illustrate the 95% confidence intervals. Panels (a) of Figures 8 and 9 show the estimation results from an OLS model without control variables, labeled descriptive estimates. Panels (b) represent results from a reduced form model, where the cutoff indicator z_i replaces the indicator variable of program type in an OLS regression. The results from the model specified in Equation (5) are presented in Panels (c), where the parameters correspond to estimtes of β_2. Panels (d) display the same empirical estimates as in (c), but translated into percentages of the samples' average earnings.

The pattern emerging is very clear in one sense, that vocational education generates a short-term earnings advantage. For males, the descriptive estimates without controls display higher earnings for individuals enrolled in vocational studies, both in the short and in the long run. The long term differences disappear to a large degree as one adds control variables and employ the reduced form estimates, implying that the descriptive averages partly reflect selection. For the full model, the results pattern is similar to the reduced form estimates, but slightly larger in magnitude. The remaining short-term earnings advantage of vocational studies is above 5% for the first eight years, with p-values below or close to 0.05. In line with our earlier stated long term hypothesis, earnings tend to converge and there are no statistically significant differences between ages 30 and 40. However, significant differences to the advantage of vocational education re-emerge around age 40 and again after age 50. Thus, the long term pattern is not completely coherent. As we narrow down the bandwidth, bias is reduced at the cost of precision. Figures 10, panels (a) and (b), show the same full model estimates with the bandwidth of the triangular weights reduced from 1.0 to 0.50 and 0.30 respectively. The pattern in the results across age groups remains roughly consistent, but the increased standard errors

make estimates statistically insignificant except the short term advantage of vocational when using bandwidth 0.50 (confidence intervals are capped in the figures to retain the scaling).

For females, panel (a) in Figure 9 shows that the descriptive estimates without controls form a U-shaped pattern with higher earnings for those with vocational studies both initially and at the end of our observations window. The reduced form model estimates indicate that control variables have a moderating effect on the estimated earnings differences between age 25 and 40. Turning to the full model estimates, in panel (c), the U-shaped pattern remains with both a short term and a long term advantage of vocational education. The short-term earnings advantage is shorter for females compared with males, lasting until the age of 23. Figures 10, panels (c) and (d), indicates that the short term advantages is stable also when one reduces the bandwidth. However, the long term advantage disappears when the bandwidth is set to 0.30. This is not due to the weaker precision of the estimates, because one may also discern that point estimates become closer to zero as the bandwidth is reduced.

4.2 Mechanisms

It is important to note that estimates reflecting causality may reflect other mechanisms than educational contents. For instance, if initial educational choices are linked to differences in further education, our results may reflect years of schooling or, for the short-term, locking in effects of registrations in education. Panels (a) and (b) of Figure 11 show descriptive statistics of college registrations for men and women. The typical exams are in areas such as nursing or school teaching. Statistics Sweden classify exam types into eight categories, 80% of the women got their exam in the categories "teaching" (27%) or "health" (53%).

Figure 12 provide results from IV regression estimates of Equation (5) with accumulated college exams as the outcome variable. The proportions of accumulated college exams do not display any significant differences among males. However, for females, the

results differ depending on the bandwidth applied. Females enrolled in vocational education are found to be more likely to complete a college education if bandwidth is 1.0. However, this difference entirely disappears when the bandwidth is reduced to 0.30. This is consistent with the results observed in Figure 10, where the pattern across age-groups remained stable for males as the bandwidth was varied. For females, with default bandwidth at 1.0, there was a relative long term advantage of vocational education. However, it is no longer visible when the bandwidth is narrowed to 0.30. This indicates that further education plausibly drives the estimates when one applies a wider bandwidth.

5 Conclusion

In this paper, we exploit GPA-cutoffs to test whether vocational and general schooling differ in their impact on labor earnings. Using data from all applications to upper secondary schooling in Sweden from 1971 to 1991, our findings are consistent with the most common theoretical predictions. Participation in vocational schooling enhances short-term earnings but earnings later tend to converge. Specifically, men enrolled in vocational programs have significantly higher earnings, of around 5%, until the age of 27, whereas women with vocational schooling have an initially larger earnings advantage which drops quckly and disappears at age 23. Importantly, these effects do not seem to be mediated by differential participation in higher education.

The results might be interpreted in combination with estimated average effects reported in Golsteyn and Stenberg (2016), using a selection on observables strategy with partly the same sources of data as in this study. In contrast to the present study, and other studies, they found a persistent relative long term earnings advantage of general education. The difference in results between average effects and marginal effects may indicate substantial heterogeneity in the long term relative effects of educational contents. The RD design employed in this study strengthens the case for a causal interpretation of the results, but this comes at the cost of the validity of the results. One way around

the trade-off between validity and causality would be if one had access to uninterrupted earnings in a set-up where individuals are assigned to different regimes, with a set-up similar to Malamud and Pop-Eleches (2010). However, there is still no such study.

References

Aakvik, A., K. G. Salvanes, and K. Vaage (2010): "Measuring heterogeneity in the returns to education using an education reform," *European Economic Review*, 54, 483–500.

Abdulkadiroğlu, A., J. Angrist, and P. Pathak (2014): "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools," *Econometrica*, 82, 137–196.

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

Brunello, G. and L. Rocco (2015): "The Labour Market Effects of Academic and Vocational Education over the Life Cycle: Evidence from Two British Cohorts," IZA Discussion Papers 9275, Institute for the Study of Labor (IZA).

Card, D. and T. Lemieux (2001): "Can falling supply explain the rising return to college for younger men? A cohort-based analysis," *Quarterly Journal of Economics*, 116, 705–746.

Clark, D. (2010): "Selective Schools and Academic Achievement," *B.E. Journal of Economic Analysis & Policy*, 10, 1–40.

Duflo, E., P. Dupas, and M. Kremer (2011): "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," *American Economic Review*, 101, 1739–1774.

Dustmann, C., P. Puhani, and U. Schönberg (2016): "The Long-Term Effects of Early Track Choice," *Economic Journal*, forthcoming.

Golsteyn, B. H. and Stenberg (2016): "Earnings over the Life Course: General versus Vocational Education," unpublished manuscript, Stockholm University.

Hall, C. (2012): "The Effects of Reducing Tracking in Upper Secondary School: Evidence from a Large-Scale Pilot Scheme," *Journal of Human Resources*, 47, 237–269.

Hanushek, E. A., G. Schwerdt, L. Woessmann, and L. Zhang (2016): "General Education, Vocational Education, and Labor-Market Outcomes over the Life-Cycle," *Journal of Human Resources*, forthcoming.

Harmon, C. and I. Walker (1995): "Estimates of the Economic Return to Schooling for the United Kingdom," *American Economic Review*, 85, 1278–1286.

Imbens, G. W. and T. Lemieux (2008): "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, 142, 615–635.

Jackson, C. K. (2010): "Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignments in Trinidad and Tobago," *Economic Journal*, 120, 1399–1429.

Kaufmann, K., M. Messner, and A. Solis (2015): "Returns to Elite Higher Education in the Marriage Market: Evidence from Chile," .

Kemple, J. and J. Scott-Clayton (2004): *Career Academies: Impacts on Labor Market Outcomes and Educational Attainment*, New York: MDRC.

Kemple, J. and C. Willner (2008): *Career Academies: Long-Term Impacts on Labor Market Outcomes, Educational Attainment, and Transitions to Adulthood*, New York: MDRC.

Kirkebøen, L., E. Leuven, and M. Mogstad (2016): "Field of Study, Earnings, and Self-Selection," *Quarterly Journal of Economics*, forthcoming.

Korpi, T., P. De Graaf, J. Hendrickx, and R. Layte (2003): "Vocational Training and Career Employment Precariousness in Great Britain, the Netherlands and Sweden," *Acta Sociologica*, 46, 17–30.

Krueger, D. and K. B. Kumar (2004a): "Skill-Specific Rather than General Education: A Reason for US-Europe Growth Differences?" *Journal of Economic Growth*, 9, 167–207.

——— (2004b): "US-Europe differences in technology-driven growth: quantifying the role of education," *Journal of Monetary Economics*, 51, 161–190.

Lee, D. S., E. Moretti, and M. J. Butler (2004): "Do Voters Affect or Elect Policies? Evidence from the U.S. House," *Quarterly Journal of Economics*, 119, 807–859.

Malamud, O. and C. Pop-Eleches (2010): "General Education versus Vocational Training: Evidence from an Economy in Transition," *Review of Economics and Statistics*, 92, 43–60.

Meghir, C. and M. Palme (2005): "Educational Reform, Ability, and Family Background," *American Economic Review*, 95, 414–424.

Oosterbeek, H. and D. Webbink (2007): "Wage effects of an extra year of basic vocational education," *Economics of Education Review*, 26, 408–419.

Pischke, J.-S. and T. von Wachter (2008): "Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation," *Review of Economics and Statistics*, 90, 592–598.

Pop-Eleches, C. and M. Urquiola (2013): "Going to a Better School: Effects and Behavioral Responses," *American Economic Review*, 103, 1289–1324.

Rosenbaum, J. E. (2001): *Beyond College for All: Career Paths for the Forgotten Half*, New York: Russell Sage Foundation.

Ryan, P. (2001): "The School-to-Work Transition: A Cross-National Perspective," *Journal of Economic Literature*, 39, 34–92.

Shavit, Y. and W. Müller (1998): *From School to Work*, Oxford University Press.

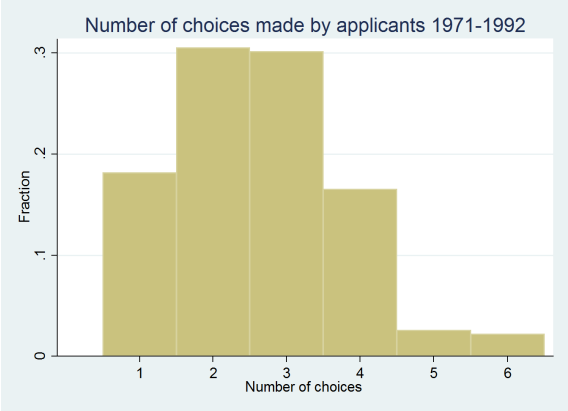
Solis, A. (2015): "Does Higher Education Cause Political Participation? Evidence From a Regression Discontinuity Design," Unpublished manuscript.

Stenberg, A. (2011): "Using longitudinal data to evaluate publicly provided formal education for low skilled," *Economics of Education Review*, 30, 1262–1280.

Wolter, S. and P. Ryan (2011): "Apprenticeship," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, S. Machin, and L. Woessmann, Amsterdam, North Holland: Elsevier, vol. 3, chap. 11, 521–576.

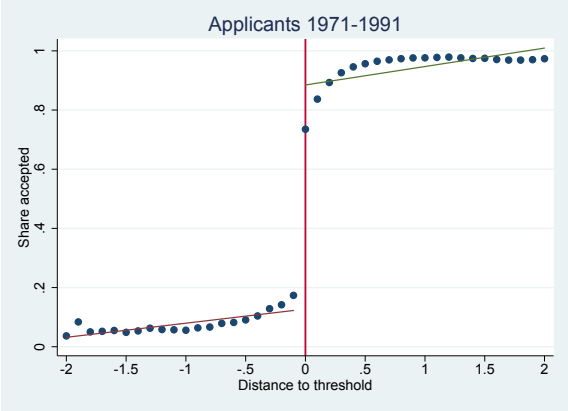
Figures

Figure 1: Distribution of individuals' choices



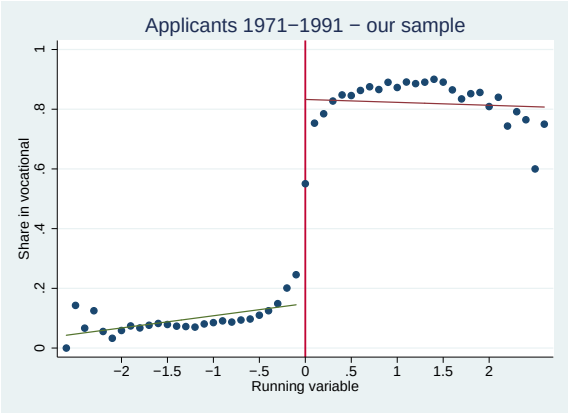
Note: The figure shows the distribution of the total number of choices across all students in our sample.

Figure 2: The discontinuity in probability of acceptance around the threshold



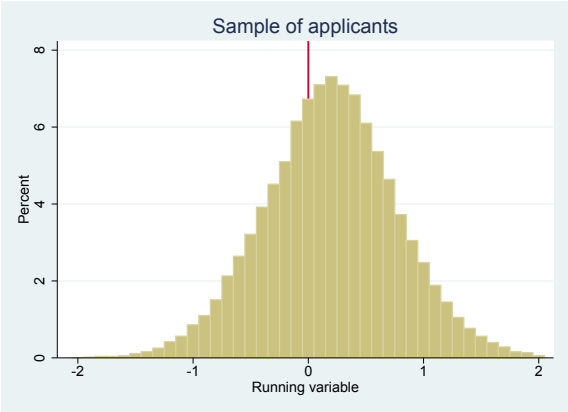
Note: The figure shows the share of accepted individuals for different values around the threshold. The solid lines show the fitted values separately for $GPA < k*$, and for $GPA \geq k*$, respectively.

Figure 3: The discontinuity in probability of acceptance to vocational program around the cutoff



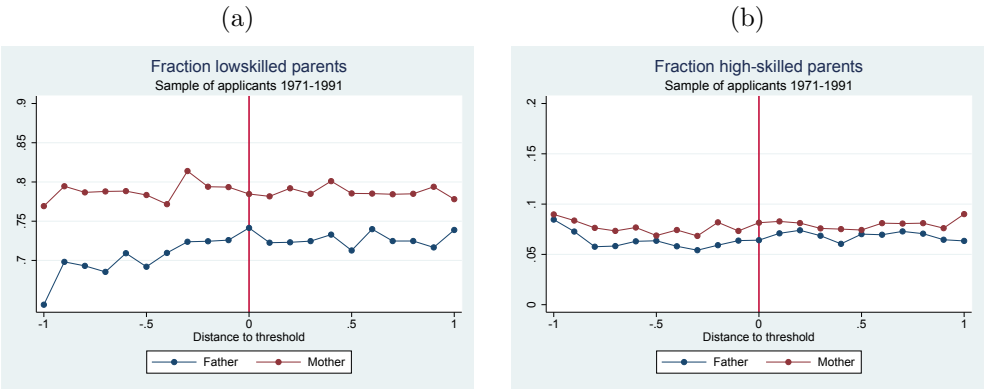
Note: The figure shows the share of individuals accepted to vocational programs for different values around the cutoff. The solid lines show the fitted values separately for $x_i < 0$, and for $x_i \geq 0$, respectively.

Figure 4: The distribution of individuals around the cutoff



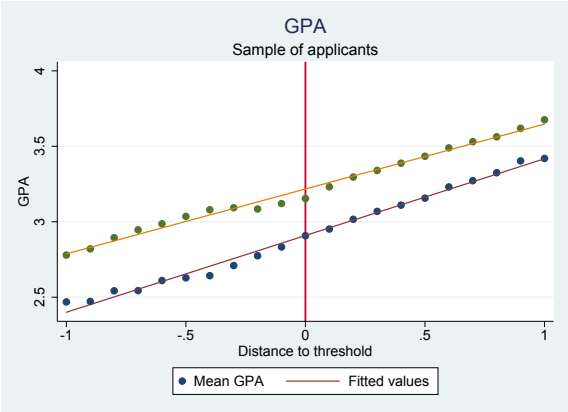
Note: The figure shows the distribution of individuals around threshold value for acceptance to the first choice.

Figure 5: The variation in parent's education around the cutoff.



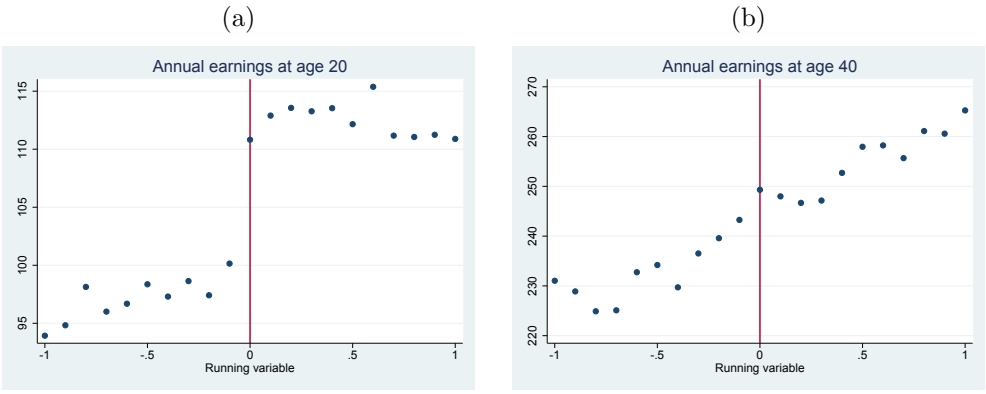
Note: The figures shows across the running variable share of mothers and fathers with no upper secondary school, lowskilled, or with some completed post secondary school, highskilled.

Figure 6: The variation in GPA levels around the cutoff.



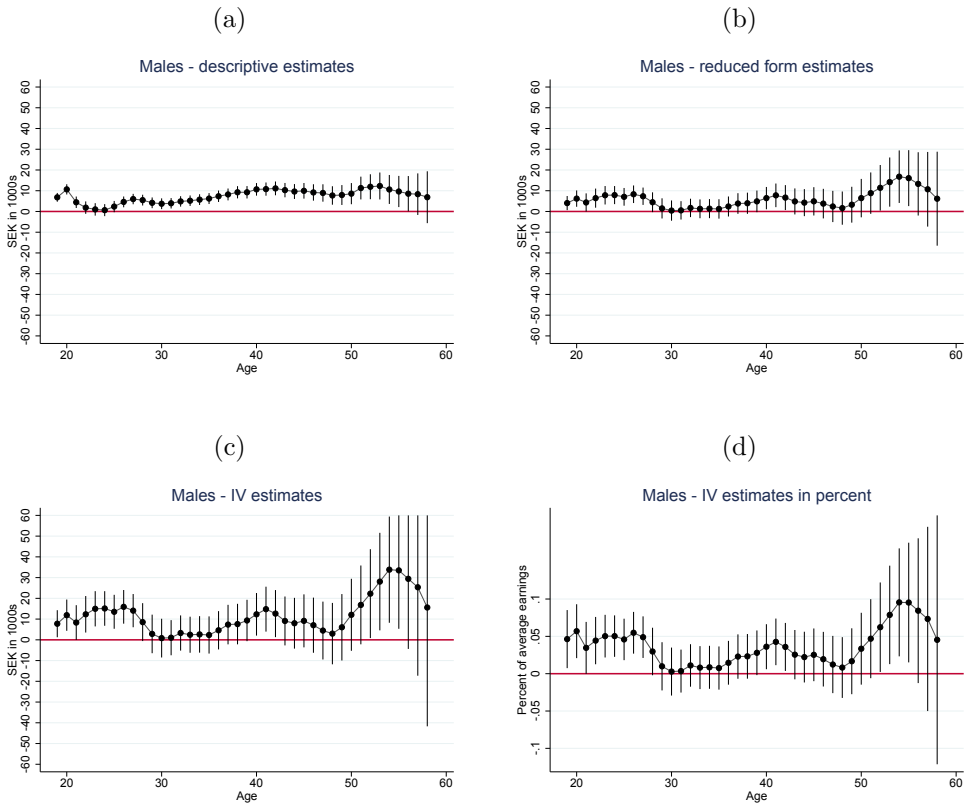
Note: The figure shows the mean GPA for males and females respectively around the cutoff.

Figure 7: Average annual earnings across the running variable



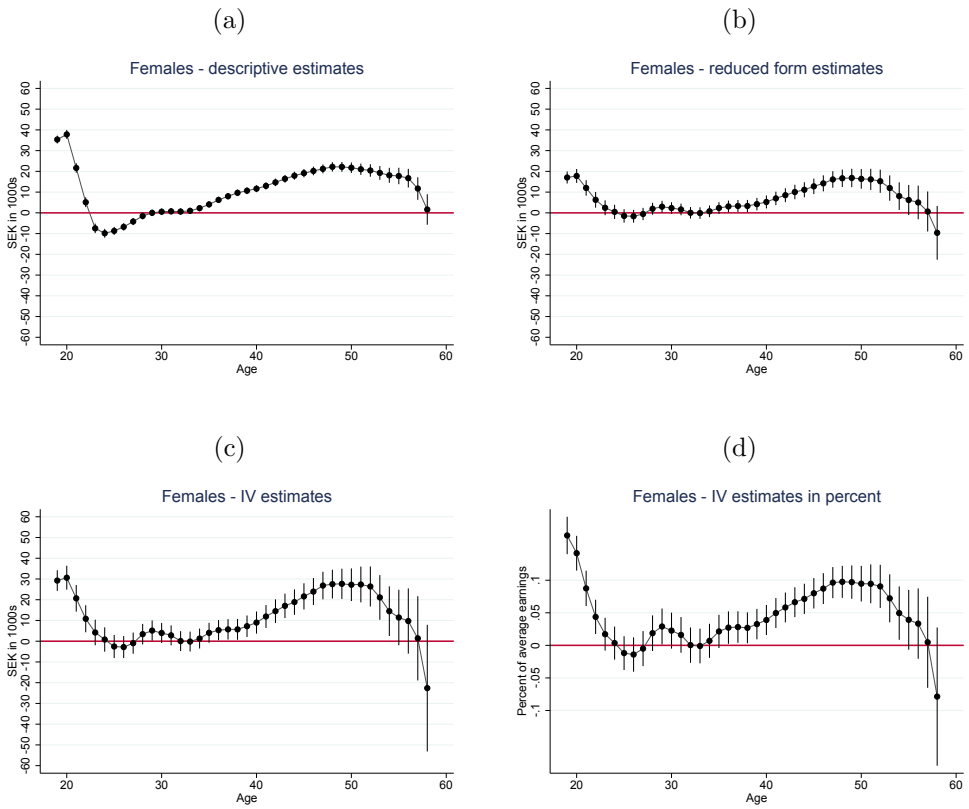
Note: The figures shows the average annual earnings for individuals at different values of the running variable around the cutoff.

Figure 8: Absolute earnings advantage of vocational versus general tracks (males)



Note: The figures show the relative earnings advantage of vocational schooling tracks over general schooling tracks for male individuals. Each point represents one estimate. Panel (a) is based on regressions without any controls; Panel (b) is based on regressions which include all observable attributes; Panel (c) is based on the full model; Panel (d) displays the full model estimates translated in percentage terms of absolute earnings.

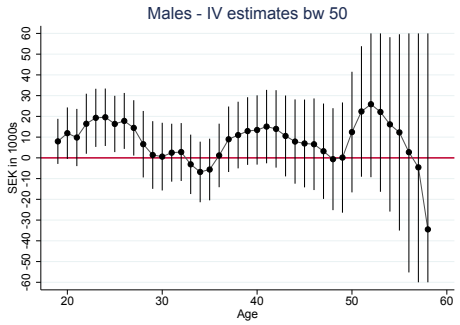
Figure 9: Absolute earnings advantage of vocational versus general tracks (females)



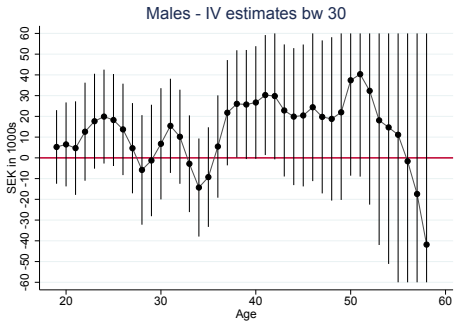
Note: The figures show the relative earnings advantage of vocational schooling tracks over general schooling tracks for male individuals. Each point represents one estimate. Panel (a) is based on regressions without any controls; Panel (b) is based on regressions which include all observable attributes; Panel (c) is based on the full model; Panel (d) displays the full model estimates translated in percentage terms of absolute earnings.

Figure 10: Absolute earnings advantage of vocational versus general tracks - varying bandwidth of triangular weights around threshold

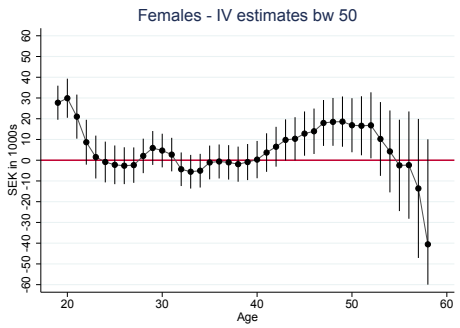
(a) Bandwidth 0.50 grade points



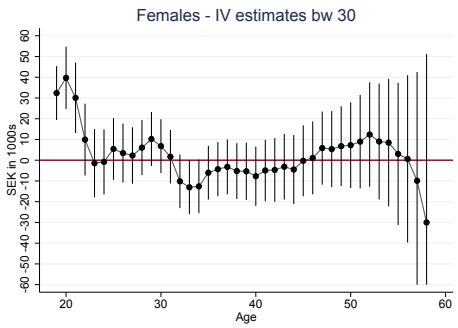
(b) Bandwidth 0.30 grade points



(c) Bandwidth 0.50 grade points



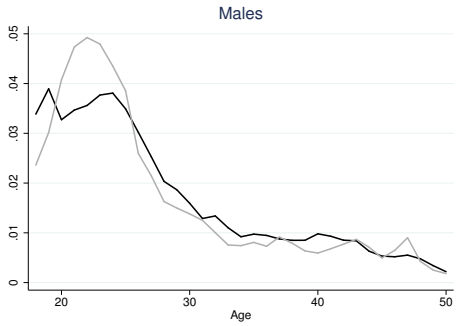
(d) Bandwidth 0.30 grade points



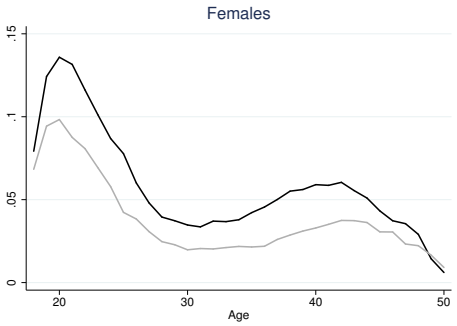
Note: The figures show the relative earnings advantage of vocational schooling tracks over general schooling tracks for male individuals. Each point represents one estimate. Samples in panels (a) and (c) are based on bandwidths less than 0.50 around the threshold; Panel (b) and (d) show the corresponding estimates with bandwidths set to less than 0.30 around the threshold.

Figure 11: Further education

(a) College registrations (male)



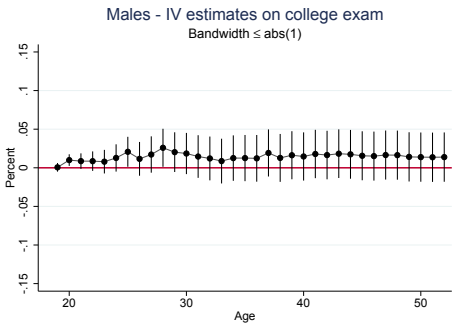
(b) College registrations (female)



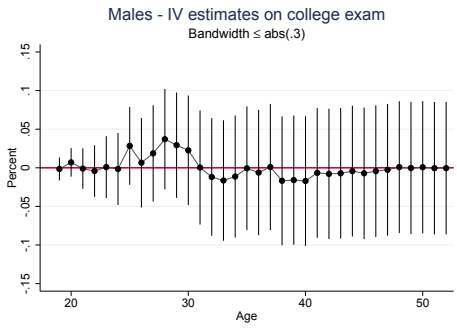
Note: Panels (a) and (b) show the share of individuals with college registrations.

Figure 12: Further education

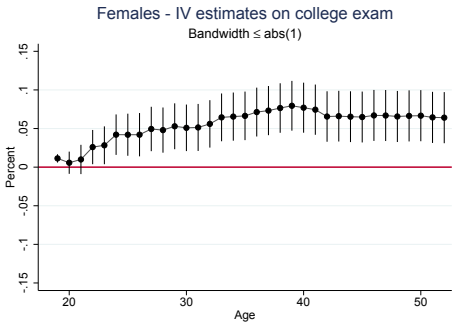
(a) Accumulated college exams (male)



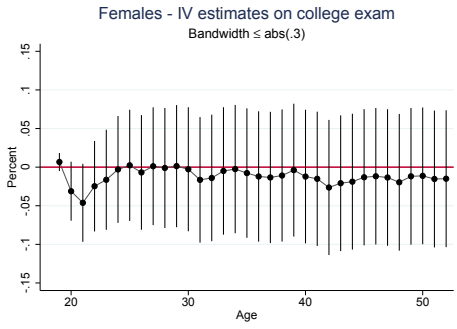
(b) Accumulated college exams (male)



(c) Accumulated college exams (female)



(d) Accumulated college exams (female)



Note: Panels show the estimated causal effect of vocational schooling on accumulated college exams, using different bandwidths.

Tables

Table 1: Number of hours per week in vocational and general programs in the curricula of the major 2-year upper secondary programs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	General programs			Vocational programs						
Type of subject	Social	Business	Technology	Consumer	Nursing	Office	Vehicle	Mech	Electronics	Construction
Non-vocational 1:										
Swedish	3.5	4	2	3.5	3.5	3.5	2	2	2	2
English	3	4	1.5							
Mathematics	3	1.5	4							
Social science	3	3	1							
History/Religion	3.5	1	1							
Natural science	6		5.25							
Business Admin.		8.5	0.5			5				
Sum non-voc. 1	22	22	15.25	3.5	3.5	8.5	2	2	2	2
Non-voc. 2:										
Optional	3	4.5		3	3	3	3	3	3	3
Technical subjects[1]			15.25							
Sum non-voc. 1& 2	25	26.5	30.5	6.5	6.5	11.5	5	5	5	5
Non-voc. 3:										
Worklife knowledge				1	1	1	1	1	1	1
Social studies	2.5			0.33						
Ergonomics			1							
Psychology				0.58	1.65					
Typing/Stenography	2	4.5				3.67				
Consumers[2]				7.2						
Social medicine[3]					2.8					
Sum non-voc. 1, 2 & 3	29.5	31	31.5	15.6	12	16.2	6	6	6	6
Voc. subjects:										
Professional practice				18.38	13.73	9	31	31	31	31
Nursing				0.83	6.38					
Childcare				1.17	4.33					
Office						10.5				
Sum voc. subjects	0	0	0	20.4	24.4	19.5	31	31	31	31
Other subjects[4]	5.5	4.5	4	2.5	2.8	2.5	2.5	2.5	2.5	2.5
Total weekly hours	35	35.5	35.5	38.5	39.2	38.2	39.5	39.5	39.5	39.5
Non-voc. 1	63%	62%	43%	9%	9%	22%	5%	5%	5%	5%
Non-voc. 1 & 2	71%	75%	86%	17%	17%	30%	13%	13%	13%	13%
Non-voc. 1 & 2 & 3	84%	87%	89%	41%	31%	42%	15%	15%	15%	15%
Voc.	0%	0%	0%	53%	62%	51%	78%	78%	78%	78%

Note: The curricula described concern the most popular 2-year upper secondary programs. Classroom subjects are divided into "non-vocational", "vocational" and "other" (physical education, music and drawing). The "non-vocational" subjects are in turn been divided into three categories: 1, 2 and 3, where the first group provides the most general knowledge and "non-vocational 3" includes subjects related to a track's intended professional activity (e.g., typing in business programs, social policy in social sciences programs).

[1] In total 24 subjects divided across four different educational paths, electro-technical, chemical-technical, machine-technical and construction.

[2] Consumer studies include Household economy, hygiene, family science, consumer studies, living environments, design.

[3] Nursing includes household economy, hygiene social medicine, anatomy.

[4] Music, drawing, physical education.

Table 2: Students' choice combinations

Choice combinations		(1) Full	(2) Estimation
Choice 1	Choice 2	sample	sample
Vocational	vocational	0.163	—
Vocational	general (2y)	0.053	0.592
Vocational	general (3y)	0.025	—
Vocational	office	0.016	—
General (2y)	vocational	0.037	0.408
General (2y)	general (2y)	0.039	—
General (2y)	general (3y)	0.031	—
General (2y)	office	0.018	—
General (3y)	vocational	0.028	—
General (3y)	general (2y)	0.069	—
General (3y)	general (3y)	0.316	—
General (3y)	office	0.011	—
Office	vocational	0.011	—
Office	general (2y)	0.009	—
Office	general (3y)	0.002	—
Office	office	0.001	—
Observations		745052	67092

Note: The figures show the shares of individuals with specific combinations of choice 1 and choice 2, for both the full sample of individuals, and for the estimation sample. The underlying samples are conditional on having two valid choices.

Table 3: Descriptive statistics by gender and program type

	(1)	(2)	(3)	(4)
	Vocational	General	Vocational	General
	Male		Female	
Distance to threshold	0.326	0.277	0.362	0.316
GPA 9th grade	3.157	3.105	3.591	3.317
Adopted (yes=1)	0.019	0.021	0.020	0.023
Oldest sibling (yes=1)	0.474	0.492	0.475	0.500
Father earnings	227.811	230.949	223.862	225.980
Father log earnings	5.318	5.322	5.303	5.308
Mother earnings	127.869	130.117	123.155	123.549
Mother log earnings	4.739	4.746	4.703	4.708
Father low skill (yes=1)	0.879	0.862	0.874	0.871
Father high skill (yes=1)	0.057	0.066	0.063	0.062
Mother low skill (yes=1)	0.866	0.846	0.866	0.868
Mother high skill (yes=1)	0.062	0.070	0.061	0.062
Foreign born (yes=1)	0.029	0.028	0.026	0.027
Father foreign (yes=1)	0.110	0.115	0.099	0.104
Mother foreign (yes=1)	0.108	0.108	0.095	0.099
Mother year of birth	1937	1938	1937	1936
Father year of birth	1934	1935	1935	1934
Log earnings age 20	4.453	4.326	4.570	4.330
Log earnings age 25	5.066	4.995	4.526	4.544
Log earnings age 30	5.269	5.206	4.559	4.541
Log earnings age 35	5.461	5.450	4.847	4.820
Observations	14256	10193	17176	25457

Note: The figures show the mean of observable variables for the estimation sample, stratified by gender and program type. Mothers' and fathers' education and annual earnings are measured at the age of 16.

Table 4: Descriptive mean statistics by gender and education type. Total sample and restricted samples.

	(1)	(2)	(3)	(4)	(5)	(6)
	Males			Females		
	Full	2-year	Rest. sample	Full	2 years	Rest. sample
Distance to threshold	0.449	0.319	0.306	0.552	0.419	0.334
GPA 9th grade	3.346	3.050	3.135	3.636	3.411	3.427
Adopted (yes=1)	0.016	0.017	0.020	0.020	0.021	0.022
Oldest sibling (yes=1)	0.505	0.475	0.481	0.512	0.488	0.490
Father earnings	253.446	216.281	229.197	257.716	223.715	225.060
Father log earnings	5.398	5.258	5.320	5.413	5.293	5.306
Mother earnings	137.613	123.185	128.865	139.109	123.800	123.378
Mother log earnings	4.805	4.707	4.742	4.816	4.709	4.706
Father low skill (yes=1)	0.792	0.892	0.872	0.785	0.871	0.872
Father high skill (yes=1)	0.125	0.049	0.061	0.131	0.064	0.063
Mother low skill (yes=1)	0.792	0.881	0.858	0.786	0.869	0.867
Mother high skill (yes=1)	0.113	0.053	0.066	0.116	0.061	0.062
Foreign born (yes=1)	0.029	0.029	0.029	0.030	0.028	0.026
Father foreign (yes=1)	0.114	0.112	0.112	0.114	0.108	0.102
Mother foreign (yes=1)	0.108	0.107	0.108	0.108	0.102	0.097
Mother year of birth	1939	1939	1937	1939	1938	1937
Father year of birth	1936	1937	1935	1936	1935	1934
Log earnings 20	4.131	4.438	4.400	4.254	4.441	4.434
Log earnings 25	4.903	5.047	5.036	4.533	4.526	4.537
Log earnings 30	5.327	5.258	5.243	4.704	4.571	4.548
Log earnings 35	5.572	5.462	5.456	4.971	4.856	4.831
Observations	377847	203106	24449	366987	169758	42633

Note: Columns (1) and (4) show the mean of observable variables for the full sample, Columns (2) and (5) for all 2-year programs, and Columns (3) and (6) for a restricted sample with running variable below .50 in absolute value. Mothers' and fathers' education and annual earnings are measured at the age of 16.