KU LEUVEN

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES
DPS16.25

OCTOBER 2016

Tracking and specialization of high schools: does school choice matter?

Olivier DE GROOTE and Koen DECLERCQ

Econometrics

Faculty of Economics
and Business

Tracking and specialization of high schools: Does school choice matter?

Olivier De Groote* Koen Declercq†

October 28, 2016

Abstract

We analyze the causal impact of choosing for an elite high school on the probability of obtaining a high school degree without study delay using a dataset for Belgium. While general schools offer study programs in all tracks, elite schools specialize and only offer programs in the academic track, preparing students for university education. If students underperform in the academic track, they can switch to a lower track to avoid study delay. For students in elite schools, switching also implies choosing for another school. We account for self-selection and heterogeneity in the treatment effect and derive a small and non-significant average effect. However, we find that there is substantial heterogeneity. Students with a high preference for elite schools experience the most negative effects, resulting in a significantly negative average treatment effect on the treated of -11.6 %points. We show that the same group is also unwilling to switch tracks after they entered an elite school. Despite the negative average effect, we also find heterogeneity among the group of treated students with a substantial fraction benefitting from an elite high school.

Keywords: school choice, early tracking, marginal treatment effects

JEL: C31, I28

* *Olivier De Groote*: KU Leuven and Ph.D. fellow of the Research Foundation Flanders (FWO), email: olivier.degroote@kuleuven.be. Corresponding author at KU Leuven: Department of Economics, Naamsestraat 69, 3000 Leuven, Belgium.

† *Koen Declercq*: KU Leuven, email: Koen.Declercq@kuleuven.be. We would like to thank Bart Cockx, Jan De Loecker, Kristof De Witte, Arnaud Maurel, Erwin Ooghe, Jo Van Biesebroeck, Frank Verboven, seminar participants at KU Leuven and Duke University and participants of the 2015 LEER workshop on "Efficiency in Education" in Leuven, the 2016 Belgian Day for Labor Economists in Antwerp and the 2016 EALE conference in Ghent. We would also like to thank the Flemish Ministry of Education for providing the dataset and Research Foundation Flanders (FWO) for financial support.

1 Introduction

Completing secondary education is important for later life outcomes. However, a substantial fraction of students does not obtain a high school degree or obtains this degree after some study delay. Among OECD countries, only 72% of young adults graduate within the theoretical duration of the program and 16% does not complete secondary education (OECD, 2014).

We investigate how the organizational structure of schools influences success in secondary education. In particular, we analyze how this structure interacts with the tracking of students in different programs. A tracking system consists of two stages. In the first stage, all students start at a common program. Afterwards, students are tracked into different programs. Tracks can often be distinguished by the way they prepare students for their life after high school like universities (academic track) or the labor market (technical and vocational track). These tracks often group students by their ability and preferences.

The supply of tracks within a specific school can affect study decisions if students prefer to stay in the same school. If schools offer programs in all tracks, tracking can occur within the same school (within-school tracking). If schools specialize in one track, students may be less likely to choose the optimal track if this track is not offered at this school. They then have to switch to another school and leave behind a familiar environment (between-school tracking). If students do not perform well, they can still switch to a lower track during secondary education. If more study options are available in the same school, it is possible that students are more likely to choose for a program that best matches with their interests and ability. Therefore, offering programs in several tracks within a school can improve study outcomes. On the other hand, specialization in one track can have benefits in attracting a more homogeneous peer group or specialized teachers. It could also be more efficient if the school infrastructure is very track-specific.

In this paper, we compare within- and between-school tracking within an early tracking system. More specifically, we analyze whether starting at a school that only offers programs in an academic track, influences the probability of graduating from high school within the theoretical duration of the program. In the rest of this paper we will refer to these schools as "elite schools".[1]

We apply our analysis to Flanders, the Dutch-speaking part of Belgium, where students

[1]Similar to our definition of elite schools, Clark and Del Bono (2014) and Guyon *et al.* (2012) define an elite school as a school that only offers academic programs. However, the definition of non-elite schools differs with our setting. In their contexts, non-elite schools only offer non-academic programs. In our setting, non-elite schools can be general schools and offer both academic and non-academic programs, or specialize in non-academic programs. Among the students attending non-elite schools, 82% attended a general school.

are tracked into different programs around the age of 13. The first year of secondary education consists of a comprehensive program and is open to all students that successfully completed elementary education. Students can also follow this program at the school of their choice as free school choice is legally enforced. In the following years, they specialize in a certain program, within a specific track. Programs can be categorized into four tracks. The academic track prepares students for academic programs (at least 4 years) at universities.[2] Other tracks prepare for professional programs (3 years) in tertiary education or for the labor market. As long as students perform well, they can choose among all programs. If they failed some courses, they might be excluded from certain programs in the next year. Alternatively, they may be required to repeat a year of study.

Our main finding is that there is considerable heterogeneity in the effect of treatment. We find that students who choose for elite schools are on average worse off. Students who did not start at an elite school would on average experience no effect. These effects can be explained by tracking decisions. We find that especially students with a high preference for elite schools, are less likely to switch to another school to enroll in another track. Additionally, there is substantial heterogeneity within the groups of treated and non-treated students. In both groups, some students experience a positive effect, while others experience a negative effect of treatment.

To establish our main findings, we face the following two problems: (1) endogeneity because of non-random self-selection into elite schools and (2) heterogeneity in the treatment effect. We address the first problem by using distance as an instrument for school choice. The policy of free school choice and the large availability of different schools in the neighborhood of students, make distance a reasonable instrument for school choice in Flanders. If the treatment effect differs between students, which turns out to be important in our setting, Heckman and Vytlacil (2005) show that the conventional two stages least squares (2SLS) estimator is difficult to interpret as, in general, it cannot be interpreted as an average treatment effect (ATE), average treatment effect on the treated (ATT), or average treatment effect on the non-treated (ATNT). We therefore follow Heckman and Vytlacil (2005) and allow for heterogeneous effects of attending an elite school by estimating marginal treatment effects (MTEs) in a nonparametric way. These effects show how elite schools affect students differently. We use the MTEs to compute the ATT: the average effect on students that actually started at an elite school. The large dataset we use, allows us to precisely estimate the ATT nonparametrically. However, in order to identify other treatment effects, such as

[2] Although there are almost no admission standards for higher education in Flanders, the tracks students followed in high school are very important predictors of success at these institutions (Declercq and Verboven, 2015).

the ATE and ATNT, we impose limited functional form and distributional assumptions in a parametric model. We also illustrate how much observable background characteristics can explain differences in treatment effects. Finally, we show how a factor structure, proposed by Aakvik *et al.* (2005), allows us to also identify distributional treatment effects, in particular the percentage of students who benefited and suffered from elite schools.

We obtain the following main findings. Although graduating within the theoretical duration of the program is more common for students starting at elite schools (77.3% versus only 70.1% in other schools), we find that this is not a causal effect but the result of self-selection by high-ability students. We also find significant heterogeneity in the treatment effect, reflected in an upward sloping MTE curve. This means that students with the highest preference for elite schools suffer most from it. We find an ATT of -11.6 %points, i.e. students who started at an elite school, experienced an 11.6 %point drop in their chances to obtain their degree on time. However, the ATNT, the counterfactual outcome of sending all other students to the elite school, is small and insignificant. We can therefore conclude that precisely those students that choose for elite schools are worse off, possibly because they are being pushed by their parents or they are prepared to accumulate study delay in order to graduate from an elite school. We also find that going to an elite school makes students less willing to switch to other tracks. This is especially the case for students with a high preference for elite schools. To test this hypothesis, we repeat the analysis for a different outcome variable: the probability to switch tracks. We find that elite schools make students less likely to switch tracks, especially for students with the highest preference for these schools. This confirms our hypothesis that elite schools influence track decisions, making it harder for students to switch to the right track.

We further investigate the heterogeneity in the treatment effect. We find more negative effects for male students and students with a disadvantaged background. Finally, we show that there is also heterogeneity within both groups of treated and non-treated students. We find that 8.4% of the treated students benefited from starting at an elite school, while 19.1% of the treated suffered from it. The results have important policy implications. Banning elite schools could increase the overall rate of students graduating on time from 72% to 75%. However, the substantial heterogeneity among treated and non-treated students implies that more gains are possible by better allocating students to both types of schools.

This paper contributes to the literature on the effect of school characteristics on study outcomes. Many studies analyze the causal impact of school characteristics on success in secondary education.[3] Similar to our study, Dustmann *et al.* (2016), Clark and Del Bono

[3]Evans and Schwab (1995), Neal (1997) and Altonji, Elder and Taber (2005) study the impact of catholic schools. Dobbie and Fryer (2011), Deming (2011) and Deming *et al.* (2014) study the impact of school

(2014) and Guyon *et al.* (2012) analyze the effects of attending a school that specializes in academic programs. Pop-Eleches and Urquiola (2013) and Abdulkadiroglu *et al.* (2014) also study the effect of elite schools but define these schools on the basis of the quality of peers. We contribute to this literature by investigating the degree of heterogeneity in the treatment effect and how it results in differences between treated and non-treated students.

Another strand of the literature analyzes the impact of early tracking, see for example Hanusek and Woessman (2006) and Pekkarinen *et al.* (2009). While these papers study the impact of the age of tracking, we contribute to this literature by comparing within- and between school tracking within an early tracking system.

Finally, we are one of the few papers that compares and directly applies the recent advances in the econometric literature on estimating treatment effects proposed by Heckman and Vytlacil (2005) and Aakvik *et al.* (2005). In our application, we are able to nonparametrically estimate the ATT and compare it with the results of a model that uses functional form and distributional assumptions. Imposing these assumptions gives us additional insights in the degree of heterogeneity of the treatment effect.[4]

The paper is organized as follows. We start by providing an overview of secondary education in Flanders and introduce our dataset. We proceed by discussing the empirical framework in section 3 and discuss the instrument in section 4. Section 5 and 6 respectively summarize and interpret the results. We then discuss some alternative outcome variables to strengthen the interpretation of the results in section 7. Finally, we provide a sensitivity analysis in section 8 and conclude in section 9.

2 Secondary education in Flanders

In this section we discuss the institutional context in Flanders and show some descriptive statistics about enrollment and study outcomes for students choosing elite and non-elite schools.

quality. Wiswall *et al.* (2014) analyze the impact of attending a STEM school and Angrist *et al.* (2013) study the effectiveness of charter schools.

[4] MTEs are usually only identified for a subset of individuals such that it is not possible to estimate ATE, ATT or ATNT nonparametrically. See for example Doyle (2007 and 2008) and Galasso and Schankerman (2015). Other papers rely mainly on parametric assumptions on the shape of the MTE curve or the underlying behavioral model to improve common support or obtain sufficiently precise estimates (Carneiro *et al.* (2011), Carneiro *et al.* (2016), Cornelissen *et al.* (2016b), Basu *et al.* (2007)).

2.1 Institutional overview

Flanders is the Dutch-speaking part of Belgium, located in the North. It consists of about 60% of the population of 11 million inhabitants, compared with 40% in the French-speaking part, which is located in the South and most of Brussels[5]. After finishing primary school, students enroll in secondary education, usually at the age of 12. Students can choose between all schools in Flanders since school choice is not geographically restricted and free school choice is law-enforced. Also capacity constraints are uncommon.[6] In practice, most students choose one of the closest alternatives. When starting at secondary education, almost all students start in a comprehensive program. A small fraction starts at a vocational-preparatory program. As these students often did not successfully complete primary education, we do not consider this group and only include students who started at the comprehensive program in our analysis.

Students follow a common program during the first year of secondary education. After the first year, they can choose between programs in four tracks.[7] The academic track prepares students for academic programs in higher education (mostly 4 or 5 year programs). The technical track and the artistic track prepare students for professional programs in higher education (mostly 3 year programs) or the labor market. Students can also choose for the vocational track, that prepares them directly for the labor market. Both students' preferences and study results in the first year determine the track. If they performed well, they can choose between all tracks. Students who fail on some courses can be excluded from certain programs or tracks. If they fail on too many important courses, they can also be required to repeat their year.[8] While mobility between tracks is generally possible, we observe almost only downward mobility, i.e. students going from the academic track to the

[5] A small minority of the Dutch-speaking part (about 10%) also lives in Brussels. There is also a small German-speaking part in Belgium, located in the East (about 0.6% of the population).

[6] Capacity constraints are becoming more common in some cities. The law however protects free school choice and prevents schools from cream skimming. If the school is capacity constrained, it must add pupils to a waiting list and if spots become available, it must respect the order of this list.

[7] Officially the distinction between tracks exists only from the third year on. However, at the start of the second year, pupils decide on elective courses that prepare for a particular track. We therefore classify tracks from the second year on.

[8] At the end of each year, students obtain a certificate, based on their study result. They obtain an A-certificate if they succeeded on all major courses. They can then move on to the next year and continue the same program. If they did not succeed for all courses, they might obtain a C-certificate which implies that they have to repeat all courses of the previous year. There is also a third possibility: a B-certificate. This implies that they can move on to the next year but only if they switch to a lower-ranked program. Alternatively, they can also decide to repeat the year instead.

technical or artistic track and from the technical or artistic track to the vocational track.[9] Note that this mobility is not always due to restrictions of continuing in the same track but often results from a choice by students or their parents because they want to avoid failing on too many courses and thereby risk accumulating study delay in the future.

The supply of programs differs between schools in Flanders. Some schools specialize and offer programs in only one track, while other schools offer programs in more tracks. Specialization offers some benefits that can help in obtaining better study outcomes like having more homogenous peer groups, specialized teachers and more efficient use of school infrastructure. However, as students' preferences are an important determinant of the chosen track, students' choices can be different in both types of schools. If schools offer programs in all tracks, tracking can occur within the same school. In schools that specialize in only one track, students have to switch schools when changing tracks. Tracking then occurs between schools. If they do not like switching to another school, students who start at a specialized school might not end up in the track that suits them best. It is possible that these students are then more likely to accumulate study delay or drop out than students who start at a school that offers programs in all tracks.

2.2 A first look at the data

We use a rich dataset provided by the Flemish Ministry of Education. We observe all students who started at the comprehensive program[10] in secondary education in the year 2003 until 2006[11] and follow them in secondary education. For every school year, we observe their school[12], program and study result. We also observe whether they graduated from secondary education and whether they graduated within six or more years of studying.

Table 1 shows summary statistics for all students who started at the comprehensive

[9] Only mobility from the vocational track to other tracks is not allowed but in general it is hard to move upwards from other tracks too because students do not have the pre-requirements for courses of the higher tracks.

[10] 85% of students enroll in the comprehensive program in secondary education. 15% directly start at a preparatory year for vocational education or in special education. We do not consider these students because they often did not successfully complete primary school.

[11] Information about the residence and socio-economic background is only available in the data from 2007. We remove pupils without information about socio-economic status or place of residence from the dataset. We also omit pupils who drop out from public education in Flanders before 2007.

[12] The administrative definition of a school does not always overlap with the actual school as it is perceived by parents and children. Large schools often have several administrative entities on the same address, while other schools use the same administrative entity for schools in very different locations. We therefore use the address of the campus where the pupil is located to define a school.

program of the first year of secondary education between 2003 and 2006. 26.6% of all students start at an elite school, a school that only offers programs in the academic track. Girls are slightly more likely to start at elite schools. Socio-economic status also determines school choice. Advantaged students are more likely to start at elite schools. The most notable effect is for the educational level of the mother. 38.0% of students whose mother has a degree in higher education start at an elite school, while this is only 14.4% if the mother has no degree in secondary education.

In the second panel of Table 1, we represent study choices of students who completed the first, comprehensive, year of secondary education. Almost all students who finished the comprehensive year in an elite school enroll in the academic track in the second year. Less students starting at non-elite schools choose for the academic track. This can be explained by two possible effects. First, if students do not like switching to another school, students at elite schools are more likely to enroll in the academic track because this is the only track offered at this school. Second, if students of higher ability are more likely to enroll in elite schools, proportionally more students will start at the academic track after their first year in an elite school.

Finally, we also represent study outcomes for both types of schools. We consider graduating from high school within the theoretical duration of the program (6 years) as our main outcome variable. Students who started at elite schools seem to perform better. 77.3% of the students who started in their first year at an elite school graduate from high school without study delay, while this is only 70.1% of the students who started at a non-elite school. We also consider an alternative outcome: graduating from high school within seven years of studying. A substantial fraction of the students graduate with one year of study delay. If graduating from elite schools has additional benefits, such as higher success in higher education, it could be beneficial to study one year longer. We observe that students who started at elite schools still perform better but the gap in graduation rates becomes smaller.

In general, we conclude that students choosing for elite schools have better study outcomes but enrollment in elite schools is not random. Therefore we need to control for self-selection to investigate the causal effect of an elite school on study success.

Table 1: Enrollment and study outcomes in secondary education

	Elite school	Non-elite school
First year enrollment		
All	26.6%	73.4%
Male	26.0%	74.0%
Female	27.3%	72.7%
Mother has post high school degree	38.0%	62.0%
Mother has high school degree	21.5%	78.5%
Mother has no high school degree	14.4%	85.6%
Dutch at home	26.5%	73.5%
No Dutch at home	28.6%	71.4%
High income (=not eligible for study grant)	28.6%	71.5%
Low income (=eligible for study grant)	18.7%	81.3%
Tracking after comprehensive year		
Academic track	93.4%	54.9%
Technical track	5.5%	36.6%
Artistic track	0.2%	0.9%
Vocational track	0.7%	7.5%
Study outcomes		
High school diploma within 6 years	77.3%	70.1%
High school diploma within 7 years	91.1%	86.6%

Note: Enrollment decisions and study outcomes are expressed as percentages of students enrolling in secondary education between 2003 and 2006. Tracking after the comprehensive year is expressed as a percentage of pupils who choose for courses that prepare for the academic track.

3 Empirical framework

We study the causal effect of starting at an elite school on graduating within the theoretical duration of the program. To address the potential self-selection of high ability students in elite schools, we use distance as an instrument for school choice. In this section, we first specify the potential outcomes and different treatment effects. If the treatment effect differs between students and depends on their willingness to select into treatment, Heckman and Vytlacil (2005) show that the standard 2SLS estimator in a traditional IV analysis cannot be interpreted as the ATT or ATE. We therefore follow two approaches that allow for a heterogeneous treatment effect. First, we discuss the nonparametric approach of Heckman and Vytlacil (2005), see also Cornelissen *et al.* (2016a) for an overview on how to apply this method. Finally, we show how parametric and distributional assumptions can overcome support problems of the nonparametric approach and how the factor structure of Aakvik *et al.* (2005) can give additional insights on the distribution of treatment effects.

3.1 Potential outcomes and treatment effects

Let Y_{i1} be the study outcome of student i in the case of treatment, starting at an elite school, and Y_{i0} the outcome of student i if he started at a non-elite school. Let D_i be equal to one if a student started at an elite school and zero otherwise. For each student i, we only observe the realized outcome in the observed state.

$$Y_i = D_i Y_{i1} + (1 - D_i) Y_{i0} \tag{1}$$

The potential study outcome in case of attending an elite school is

$$Y_{i1} = \mu_1(X_i, U_{i1}), \tag{2}$$

while the potential study outcome in case of no treatment is

$$Y_{i0} = \mu_0(X_i, U_{i0}), \tag{3}$$

where X_i is a vector of observed characteristics of students and U_{i0} and U_{i1} are unobserved random variables.

From this model, we can obtain the commonly used treatment parameters:

- Average Treatment Effect (ATE):

$$\alpha^{ATE}(x) = E[Y_1 - Y_0 | X = x]$$

- Average Treatment Effect on the Treated (ATT):

$$\alpha^{ATT}(x) = E[Y_1 - Y_0 | X = x, D = 1]$$

- Average Treatment Effect on the Non-Treated (ATNT):

$$\alpha^{ATNT}(x) = E[Y_1 - Y_0 | X = x, D = 0]$$

With x the realization of the random variable X. These treatment effects can then be averaged over the empirical distribution of X to obtain one average treatment effect.

3.2 Selection equation

Students choose an elite school if the following condition holds:

$$\begin{aligned} D_i &= 1(U_{iV} \leq \mu_V(X_i, Z_i)) \\ &= 1(F_{U_V}(U_{iV}) \leq F_{U_V}(\mu_V(X_i, Z_i)) \\ &= 1(U_{iD} \leq \Pr(D_i = 1 | X_{i,} Z_i)) \end{aligned} \tag{4}$$

with 1() an indicator function = 1 if the expression between brackets is true, $\mu_V(X_i, Z_i)$ an arbitrary function of observed characteristics X_i and Z_i that determine the utility of starting at an elite school. U_{iV} is the unobserved cost of treatment for student i and F_{U_V} its cumulative distribution function (cdf). The probability normalization in (4) then allows us to write selection into treatment as a function of the propensity score $\Pr(D = 1|X, Z)$ and an unobserved cost of treatment that is by construction uniformly distributed: $U_D \sim Uniform(0, 1)$, regardless of the unspecified distribution of U_V. U_D can also be interpreted as the quantiles of U_V.

In the rest of the paper we maintain the following assumptions on the instrument Z:

Condition 1 (Relevance) *Z is a random variable such that the propensity score $P(X, Z) \equiv \Pr(D = 1|X, Z)$ is a nontrivial function of Z.*

Condition 2 (Exogeneity (conditional on X)) *Z is uncorrelated with U_1 and U_0 after conditioning on X*

These two conditions are similar to the standard 2SLS assumptions. The first condition assures that the instrument is strong. The second condition assures that the instrument has no direct impact on study outcomes. Vytlacil (2002) also notes that the selection model in (4) is equivalent to the monotonicity condition of the LATE literature (Imbens and Angrist, 1994). In our application this means that all students should perceive distance to school as a cost and not as a benefit.

3.3 Estimation

A common treatment effect implies that the ATT is the same as the ATNT and that students do not select themselves on the basis of expected returns from schooling. In this case, one can estimate the treatment effect using 2SLS. If this assumption is not satisfied, Heckman and Vytlacil (2005) show that 2SLS does not estimate the ATT, ATE or ATNT and that this estimate is then difficult to interpret. To allow for heterogeneous effects, we follow the framework of Heckman and Vytlacil (2005) by using marginal treatment effects (MTE) as a building block for nonparametric estimation of ATT, ATE and ATNT. While our data allows for nonparametric estimation of the ATT, we have insufficient common support to identify ATNT or ATE. Furthermore we cannot identify how observable covariates can explain heterogeneity because common support that is conditional on observables, is more difficult to achieve. We therefore propose a parametric model, similar to Manski *et al.* (1992) with limited functional form and distributional assumptions. Finally, we show how distributional treatment effects can be identified, using identifying assumptions of a factor model, proposed by Aakvik e*t al.* (2005).

3.3.1 Nonparametric approach

Heckman and Vytlacil (2005) estimate MTEs to allow for a heterogeneous effect and use these estimates to compute the aforementioned treatment effects. The MTE is the effect of attending an elite school for students with observable characteristics $X = x$ and unobserved cost of treatment $U_D = u_D$:

$$\alpha^{MTE}(x,u_D) = E[Y_1 - Y_0|X = x, U_D = u_D]$$

Since the cost of treatment U_D is by construction uniformly distributed between 0 and 1, we can interpret the treatment effect at low values of u_D as the effect for students that have a low unobserved cost, i.e. a high unobserved preference for going to an elite school. The effect at high values is the effect for students with high unobserved costs, or low unobserved preferences. From the MTEs, we can derive the following average treatment effects (Heckman and Vytlacil, 2005):

$$\begin{aligned}
\alpha^{ATE}(x) &= \int_0^1 \alpha^{MTE}(x,u_D)du_D \\
\alpha^{ATT}(x) &= \frac{1}{E[P(X,Z|X=x)]}\int_0^1 \alpha^{MTE}(x,u_D)\Pr(P(X,Z) > u_D|X=x)du_D \qquad (5)\\
\alpha^{ATNT}(x) &= \frac{1}{E[1-P(X,Z|X=x)]}\int_0^1 \alpha^{MTE}(x,u_D)\Pr(P(X,Z) \leq u_D|X=x)du_D
\end{aligned}$$

Estimation of the MTE can proceed using local IV with $P(x,z) \equiv P(X = x, Z = z)$ as an instrument for Y in small neighborhoods around each value of $P(x, z)$ (Heckman and Vytlacil, 1999). This is because the MTE can also be interpreted as the treatment effect for students that are indifferent between elite and non-elite schools at different levels of the propensity score. To identify the entire $\alpha^{MTE}(x, u_D)$ function, we need sufficient observations of both treated and non-treated students at each value of $P(x, z)$:

Condition 3 (Common support) *For each $P(x, z)$, there should be treated and non-treated individuals.*

Since it is difficult to achieve common support, conditional on observable characteristics X, we ignore them in the nonparametric estimation. This implies that X gets absorbed by the unobservables in the model and therefore the exclusion restriction of the instrument is now stronger:

Condition 4 (Exogeneity (unconditional on X)) *Z is uncorrelated with X, U_1 and U_0*

This assumption makes Z independent of X and ignores the distinction between observed and unobserved costs of treatment. In section 4 we show that we can find $\alpha^{MTE}(u_D)$ for all $u_D < 0.6$. This allows us to interpret how costs of treatment relate to treatment outcomes for students with low costs, or high preferences, of going to an elite school. This is sufficient to estimate the ATT because the weights in the calculation of the ATT ($\frac{1}{E[P(Z)]} \Pr(P(Z) > u_D)$) are 0 for $u_D > 0.6$.

Nevertheless our data do not allow to nonparametrically estimate the ATNT and ATE. Nor does it tell us how observable characteristics contribute to the treatment effects. We therefore discuss a model that introduces limited functional form and distributional assumptions.

3.3.2 Parametric approach

To overcome the support conditions of nonparametric estimation, we follow Manski *et al.* (1992) and specify a simple, yet flexible model for going to an elite school and graduating on time. Most importantly it does not impose a Roy model structure, i.e. students do not need to select on gains of treatment. We use the following functional form assumptions:

$$\begin{aligned}
D_i^* &= \gamma Z_i + \delta X_i - U_{iV} \\
D_i &= 1 \text{ if } D_i^* \geq 0 \text{ and } D_i = 0 \text{ otherwise} \\
Y_{i0}^* &= \beta_0 X_i - U_{i0} \\
Y_{i0} &= 1 \text{ if } Y_{i0}^* \geq 0 \text{ and } Y_{i0} = 0 \text{ otherwise} \\
Y_{i1}^* &= \beta_1 X_i - U_{i1} \\
Y_{i1} &= 1 \text{ if } Y_{i1}^* \geq 0 \text{ and } Y_{i1} = 0 \text{ otherwise}
\end{aligned} \tag{6}$$

The model is then completed by assuming a distribution of U_1, U_0 and U_V. We assume that the error terms are jointly normal with a mean-zero vector and correlation matrix Ω:

$$\Omega = \begin{pmatrix} 1 & \rho_0 & \rho_1 \\ & 1 & \rho_{10} \\ & & 1 \end{pmatrix} \tag{7}$$

Because Y_1 and Y_0 are never observed simultaneously, the joint distribution of (U_1, U_0) and thus their correlation ρ_{10} is not identified. We can however estimate correlations between U_0 and U_V: ρ_0 and between U_1 and U_V: ρ_1. Since the joint distribution of each outcome equation and the selection equation is identified, we can calculate ATT, ATNT and ATE:

$$\begin{aligned}
\alpha^{ATE}(x) &= \Pr(Y_1 = 1|X = x) - \Pr(Y_0 = 1|X = x) \\
&= \Phi(\beta_1 x) - \Phi(\beta_0 x)
\end{aligned} \tag{8}$$

With Φ the cdf of a normal distribution. Similarly we find:

$$\begin{aligned}
\alpha^{ATT}(x) &= E_Z\left[\frac{\Phi_2(\beta_1 X, \gamma Z + \delta X, \rho_1) - \Phi_2(\beta_0 X, \gamma Z + \delta X, \rho_0)}{\Phi(\gamma Z + \delta X)}|X = x, D = 1\right] \\
\alpha^{ATNT}(x) &= E_Z\left[\frac{\Phi_2(\beta_1 X, -\gamma Z - \delta X, -\rho_1) - \Phi_2(\beta_0 X, -\gamma Z - \delta X, -\rho_0)}{\Phi(-\gamma Z - \delta X)}|X = x, D = 0\right]
\end{aligned} \tag{9}$$

With Φ_2 the cdf of a bivariate normal and E_Z the expected value over the empirical distribution of Z. Note that the average treatment parameters are identified for each possible set of covariates x. We can average over the empirical distribution of X (possibly conditional on treatment status) to recover the average effects α^{ATE}, α^{ATT} and α^{ATNT}. Furthermore, we can identify average marginal effects by investigating how the treatment effect differs when one covariate changes value. Note that all covariates are dummy variables, therefore we calculate the marginal effect of variable x_k as:

$$E_X\left[(\Phi(\beta_1 x_{x_k=1}) - \Phi(\beta_0 x_{x_k=1})) - (\Phi(\beta_1 x_{x_k=0}) - \Phi(\beta_0 x_{x_k=0}))\right]$$

With E_X the expected value over the empirical distribution of X, $x_{x_k=1}$ is the observed x vector with element x_k replaced by 1, and $x_{x_k=0}$ the observed x vector with element x_k replaced by 0.

3.3.3 Distributional treatment effects

While ρ_{10} was not needed for average effects (see (8) and (9)), it is required to know more about the distribution of treatment effects. In particular, we investigate the percentage of the students that would benefit or suffer from going to an elite school. The calculation of these distributional treatment effects requires additional structure on the covariances of the error terms since Y_1 and Y_0 are never observed simultaneously. Aakvik *et al.* (2005) provide a model to calculate distributional treatment effects when the outcome variable is discrete. In particular they discuss the effects on a dichotomous outcome that is generated by an underlying linear latent index. In this case, the individual treatment effect can only take three values. A student either benefits from going to an elite school, suffers or experiences no effect. While individual treatment effects remain unidentified, distributional treatment effects can be identified. One example is the percentage of students that would benefit from going to an elite school[13]:

$$\begin{aligned} E[\alpha &= 1|X = x] = \Pr(Y_0 = 0, Y_1 = 1|X = x) \\ &= \Pr(Y_1 = 1|X = x) - \Pr(Y_0 = 1, Y_1 = 1|X = x) \\ &= \Phi(X\beta_1) - \Phi_2(X\beta_0, X\beta_{1,}\rho_{10}) \end{aligned}$$

In order to calculate these distributional treatment effects, we need to identify ρ_{10}. Therefore Aakvik *et al.* (2005) follow a similar structure as in (6) but instead of estimating covariances between error terms, they impose the following factor structure:

$$\begin{aligned} U_{iV} &= -\theta_i + \varepsilon_{iD} \\ U_{i0} &= -\alpha_0\theta_i + \varepsilon_{i0} \\ U_{i1} &= -\alpha_1\theta_i + \varepsilon_{i1} \end{aligned} \tag{10}$$

with $\theta, \varepsilon_D, \varepsilon_0$ and ε_1 all independently distributed. I.e. correlation between error terms enters exclusively through one common factor θ_i.[14]

[13] See Aakvik *et al.* (2005) for other distributional treatment parameters. We also calculate the proportion of students that would benefit among the treated and among the non-treated.

[14] One can think of θ_i as unobserved ability or differences in school environment.

One particular case is where $\theta, \varepsilon_D, \varepsilon_0$ and ε_1 are all normally distributed. In this case, the model is a identical to the one we proposed in the previous subsection[15] and has the following correlation structure:

$$\rho_0 = \frac{\alpha_0}{\sqrt{2}\sqrt{1+\alpha_0^2}}$$
$$\rho_1 = \frac{\alpha_1}{\sqrt{2}\sqrt{1+\alpha_1^2}}$$

Note that the estimated correlations ρ_0 and ρ_1 are simply transformations of the factor loadings α_0 and α_1. More importantly, the previously unidentified correlation, ρ_{10}, is now identified as it only depends on the two factor loadings:

$$\rho_{10} = \frac{\alpha_0\alpha_1}{\sqrt{1+\alpha_0^2}\sqrt{1+\alpha_1^2}} = 2\rho_0\rho_1$$

Provided that there is a solution for α_0 and α_1, the factor structure identifies ρ_{10} in our model. This identity can then be used to calculate distributional treatment effects. Note that this factor structure was only needed to identify ρ_{10} and therefore has not impact on the estimates of ATT, ATNT, ATE or marginal effects.

To estimate this model, we proceed as follows: we first estimate (6) with error structure (7) using maximum likelihood[16] to estimate all parameters in (6) and ρ_0 and ρ_1. We then calculate ATT, ATNT, ATE and average marginal effects of observed characteristics. We subsequently test if the estimates of ρ_0 and ρ_1 have a solution for α_0 and α_1 and then calculate $\rho_{10} = 2\rho_0\rho_1$. This allows us to also calculate distributional effects.

4 Discussion of the instrument

We use the relative distance to an elite school as an instrument for school choice. The relative distance is the distance to the closest non-elite high school, subtracted by the distance to the closest elite high school[17]. The use of geographical variation as an instrument for school

[15] Note that normalizations are different. Instead of (implicitly) normalizing the variances of U_{iD}, U_{i0}, and U_{i1} to be 1, the model of Aakvik *et. al* (2005) implies variances of respectively 2, α_0^2+1 and α_1^2+1. I.e. our estimates should be multiplied by the square roots of these numbers to translate them to their model.

[16] We estimate the model in STATA using the user-written command 'switch_probit' (Lokshin and Sajaia, 2011).

[17] We observe the location of the students at the level of the statistical sector. In Belgium, each municipality is divided into several statistical sectors. Belgium consists of 19 781 statistical sectors (Vademecum Statistische sectoren, 2012). Statistical sectors have a surface of on average 1.54 km^2 and on average 539 inhabitants. To construct our instrument, we compute the distance from the center of the statistical sector to the exact location of the school.

choice has been proposed by Card (1995) and has also been used by other studies in the educational literature, see for example Barrow *et al.* (2015), Cullen *et al.* (2015) and Carneiro *et al.* (2016). In this section, we provide a discussion of the three assumptions of the instrument: relevance, exogeneity and common support. Note that also a fourth assumption, monotonicity, was implied by the selection model 4. As discussed in section 3.2, this implies that all students must perceive distance as a cost and not as a benefit.

4.1 Relevance

The first assumption implies that distance should have a strong impact on school choice. Table 2 shows the results of the first stage of a 2SLS regression. The relative distance has the expected positive sign and is highly significant, resulting in a high F-statistic of the exclusion restriction. Students (or their parents) are sensitive to distances to schools when making the choice to go to an elite high school as an additional kilometer to a non-elite high school (or an equal reduction in the distance to the elite school) makes them 3.4% points more likely to choose the elite school.

Table 2: First stage: choosing for an elite school

	(1)		(2)	
Variables	Coef.	St. error	Coef.	St. error
Relative distance	0.034*	(0.001)	0.032*	(0.001)
Mother no high school degree[a]			-0.215*	(0.004)
Mother high school degree[a]			-0.145*	(0.003)
No dutch at home			0.056*	(0.006)
Low income			-0.046*	(0.003)
Male			-0.016*	(0.002)
Constant	0.350*	(0.004)	0.460*	(0.005)
Observations	218,211		218,211	
R-squared	0.073		0.116	
F-stat excl. instr.	1697		1711	

Note: Standard errors are corrected for clustering within the statistical sector.

* statistical significance at 5% level.

[a] Base category = mother has a degree in higher education.

This result remains stable when controlling for observed student characteristics. This is what we would expect from a valid instrument. The control variables do already indicate that selection into elite schools is non-random as disadvantaged students are less likely to choose for elite schools. The only exception is students who do not speak Dutch at home as they are more likely to choose an elite school but they also have lower expected outcomes.[18]

4.2 Exogeneity

The second condition (exogeneity) implies that the instrument should not be correlated with unobservables that determine the outcome equation. Exogeneity of an instrument is an untestable assumption and is usually justified on institutional grounds. In our case, differences in distances cannot be correlated with unobservables in the success equation. This implies that parents are expected to ignore the school environment in their location choice. In a similar way, schools that offer better education must not locate themselves in areas with better students. First, we argue based on institutional grounds that distance to schools is an exogenous instrument for school choice in Flanders. Second, we assess how the estimation of the treatment effect would change if the instrument would be correlated with unobservables in the outcome equation.

While distances can be problematic in a lot of institutional contexts, we argue that this is not the case in Flanders.[19] School choice is free and most students live close to several schools (see Table A1 in appendix A). Students can choose between 827 schools in Flanders. For the median student, six schools are located within 5 km distance.[20] Students can also benefit from cheap public transportation or bike roads, the most common means of transportation for high school students. So while distances can be large enough for students to influence their school choice, it is unlikely that it would influence moving decisions. Moreover, schools are not expected to differ so much in their quality because of government imposed curricula and similar public financing. So while we still expect and investigate differences between schools, we do not expect them to be large enough to cover moving costs.

However, if there would still be some correlation between the instrument and the unobservables determining study success, it would be informative to assess the direction of the

[18] A possible explanation is that this variable groups pupils from very diverse migrant backgrounds, both low and high skilled, and is therefore difficult to interpret. While low skilled migration was very common in Flanders for industrial production during the 20th century, also high skilled migration is important because of Flanders' proximity to Brussels, the capital of the European Union.

[19] Altonji et al. (2005) show that distance is not a valid instrument to identify the effect of attending a catholic high school using data from the U.S.

[20] In section 8 we provide a sensitivity analysis where we restrict the sample to students living in areas with many schools and we obtain similar results.

potential bias and assess how sensitive our results are towards deviations from the exogeneity assumption. If the instrument Z is correlated with the unobservables in the outcome equation ε, the effect of elite schooling α will be biased. To evaluate the direction of this bias, note that the probability limit of the 2SLS estimator $\hat{\alpha}$ is:

$$p\lim \hat{\alpha} = \alpha + \frac{Corr(Z_i, \varepsilon_i)}{Corr(Z_i, D_i)} \frac{\sqrt{Var(\varepsilon_i)}}{\sqrt{Var(D_i)}} \tag{11}$$

So the sign of the bias is positive if $sign(Corr(Z_i, \varepsilon_i)) = sign(Corr(Z_i, D_i))$ and negative otherwise. We argue that it is more likely that the sign of the potential bias is positive, so that in case our instrument is endogenous, we would overestimate the benefits or underestimate the losses from treatment. First, we find that the denominator of the bias is positive as the instrument is positively correlated with the treatment indicator. Second, we expect $Corr(Z_i, \varepsilon_i)$ to be positive. This positive correlation arises if (1) parents of students with higher ability locate themselves in the neighborhood of elite schools or (2) elite schools locate themselves in the neighborhood of high-ability students.[21] As the MTE approach proceeds by using local IV, a similar intuition holds for the results in which we allow treatment effects to differ among individuals. Since one of our main results is that the ATT of going to an elite school is negative, it can thus be interpreted as a conservative estimate.

4.3 Common support

This last condition implies that for each value of the propensity score $P(z)$, there should be treated and non-treated individuals. We estimate the propensity score using a probit regression and plot the histogram of common support in Figure 1. The common support assumption is satisfied for values of $P(z)$ below 0.6. This allows us to identify the MTE for values of u_D below 0.6. For higher values of $P(z)$, there are almost no treated or non-treated students[22]. From (5) it is clear that this allows us to compute the ATT but not the other two treatment effects. This is because the weights in the calculation of the ATT ($\frac{1}{E[P(Z)]} \Pr(P(Z) > u_D)$) become 0 for $u_D > 0.6$.[23] Therefore we do not need estimates of the MTE for $u_D > 0.6$ and thus we do not need common support at $P(z) > 0.6$. We plot the weights in Figure A1 in Appendix A. The ATT attaches more weight to students with low costs of treatment (low u_D) as they are more likely to select into treatment. Note that

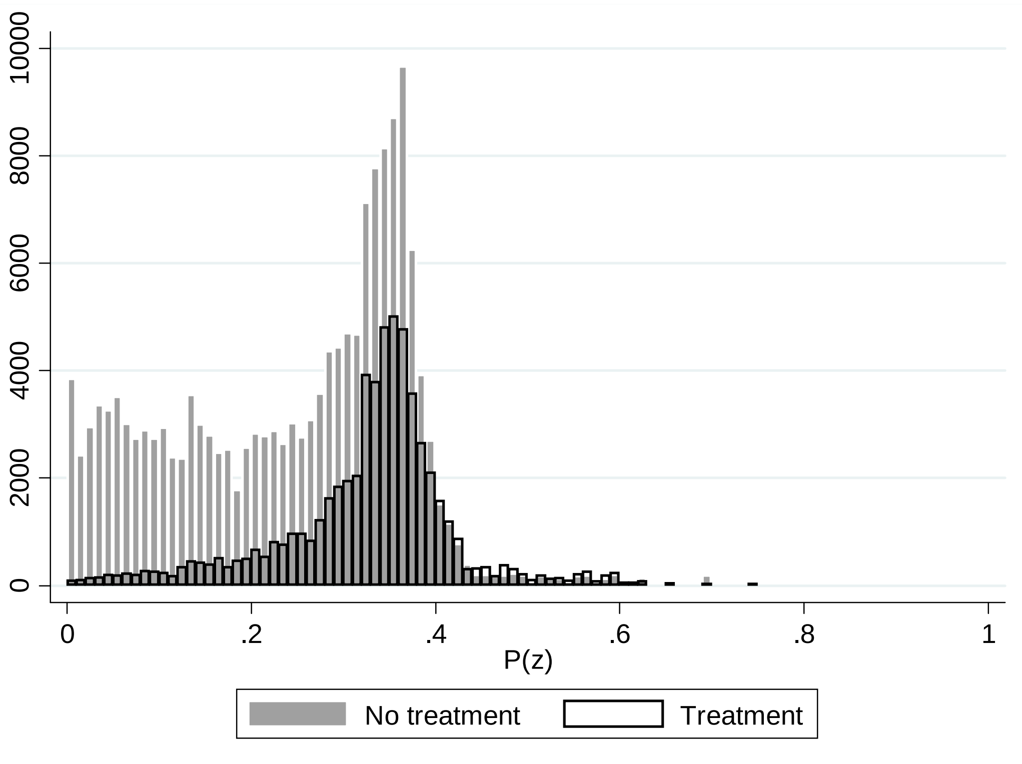
[21] Table A1 also shows that the differences in distance based on observable characteristics are small.

[22] Each percentile below 0.6 contains at least 50 treated and 50 non-treated students. For most percentiles above 0.6 this condition is not satisfied.

[23] In practice the weights are non-zero but small but we set them equal to zero for values of $P(z)$ larger than 0.6 (see Figure A1).

this support condition is no longer necessary in the parametric approach as the parametric and distributional assumptions allow for the identification of the entire distribution of potential outcomes and the selection probability (as well as the entire MTE as a function of unobservables).

Figure 1: Common support



Note: This figure represents the number of treated and non-treated students for each percentile of the propensity score.

5 Empirical results

We first compare the results of simple OLS and 2SLS regressions. Next, we account for both self-selection and heterogeneous treatment effects and derive the corresponding average treatment effects. We show how the average treatment effects vary with personal characteristics and compute the fraction of students that benefits or suffers from elite schools.

5.1 Traditional IV analysis

Table 3 represents the results of the OLS and 2SLS regressions. The first specification only includes our variable of interest, starting at an elite school. There is a positive effect on the probability of finishing high school without study delay. This effect decreases when controlling for background characteristics such as gender and socio-economic status. This decrease in the coefficient can be explained by selection on observed characteristics. If more advantaged students choose for elite schools, omitting background characteristics will lead to an upward bias of the treatment effect.

The OLS regressions do not take into account selection based on unobserved characteristics. If high ability students are more likely to attend elite schools, the effect of elite schooling will be overestimated. When we control for self-selection, we find a significantly negative effect. When adding control variables in the last specification, the 2SLS estimate remains almost the same, further suggesting that the instrument is exogenous without conditioning on observables.

Table 3: obtain a HS degree within 6 years

	OLS				2SLS			
	(1)		(2)		(3)		(4)	
Variables	Coef.	St. error	Coef.	St. error	Coef.	St. error	Coef.	St. error
Elite school	0.072*	(0.003)	0.037*	(0.003)	-0.091*	(0.014)	-0.108*	(0.012)
Mother no high school[a]			-0.174*	(0.003)			-0.206*	(0.005)
Mother high school[a]			-0.085*	(0.002)			-0.107*	(0.003)
No dutch at home			-0.192*	(0.006)			-0.179*	(0.006)
Low income			-0.042*	(0.003)			-0.049*	(0.003)
Male			-0.141*	(0.002)			-0.143*	(0.002)
Constant	0.701*	(0.002)	0.867*	(0.002)	0.745*	(0.004)	0.923*	(0.005)
Observations	218,211		218,211		218,211		218,211	

Note: Standard errors are corrected for clustering within the statistical sector.

* statistical significance at 5% level.

[a] Base category = mother has a degree in higher education.

5.2 Heterogeneous treatment effects

If treatment effects are the same for everyone, the 2SLS method estimates this effect to be -10.8% points. This average would be the same for treated and non-treated individuals. However, if treatment effects are heterogeneous and selection of schools by students is influenced by this heterogeneity, we do not estimate any of the relevant treatment parameters. We first discuss the results of the MTE estimates of the nonparametric model. We nonparametrically estimate the ATT using MTEs and compare this estimate to the corresponding estimate, resulting from the parametric model we imposed in section 3.3.2. We use this model to further investigate the heterogeneity of the treatment effect.

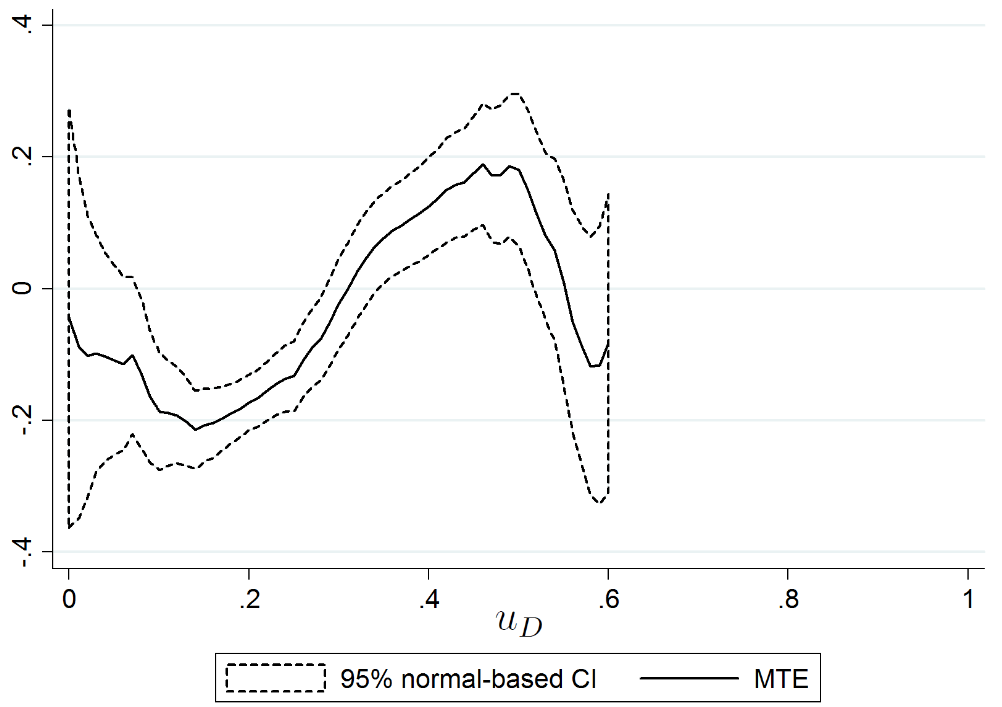
5.2.1 Marginal Treatment Effects

In Figure 2, we represent the estimation of the marginal treatment effects of starting at an elite school.[24] In the estimation of the MTEs, we do not include control variables, similar to specifications (1) and (3) in Table 3. This allows for easier interpretation of the results and is valid under the assumption that the instrument is exogenous. On the horizontal axis, we represent u_D, which can be interpreted as the cost of treatment. Students with a low u_D have a low cost or a high preference for elite schools and are therefore more likely to be treated.

Figure 2 shows that the treatment effect is not constant, but differs between individuals with a different cost of treatment. Therefore it is difficult to assess what treatment effect was identified by 2SLS in the previous subsection. The slope of the MTE is in itself of interest. If students would select on the basis of unobserved expected gains, we would expect a downward sloping MTE curve, because students with a low cost of treatment (for example high-ability students) would benefit more from the treatment. However, we find an increasing marginal treatment effect for a large part of the graph, especially in the area that is estimated most precisely. This implies that students who have a high preference for elite schools, experience a more negative treatment effect.

[24] We follow Doyle (2007) and use the bandwidth that minimizes the sum of squared errors between the local quadratic estimator and a fourth-degree polynomial model.

Figure 2: MTE of elite schools to graduate on time



Note: MTEs are calculated using a local polynomial regression of degree 2 with an Epanechnikov kernel (bandwidth 0.117). Standard errors are computed with a bootstrap procedure (250 replications) and clustered within the statistical sector. No results for $u_D > 0.60$ due to insufficient common support.

5.2.2 Average and distributional treatment effects

ATT, ATNT and ATE

Table 4 compares the estimates of the average treatment effects for the different models. Heckman and Vytlacil (2005) show that MTEs can be used to compute more interesting treatment effects such as the ATT, ATNT and ATE. In order to be able to compute these three different treatment effects, the common support assumption should be satisfied. In general this assumption is difficult to satisfy in nonparametric estimation. Nevertheless, as explained in section 4, we are able to identify an ATT of -0.116. On average, students in elite schools experienced an 11.6 %points decrease in their probability to obtain a degree. Note that the estimated ATT is close to the 2SLS estimate but this is merely coincidental.

Table 4: Treatment effects: obtaining a high school degree without study delay

	Traditional IV		Nonparametric approach		Parametric approach	
	Coef.	St. error	Coef.	St. error	Coef.	St. error
Treatment effects						
ATT	-0.108*	(0.012)	-0.116*	(0.024)	-0.107*	(0.010)
ATNT	-0.108*	(0.012)	*Not identified*[a]		0.015	(0.014)
ATE	-0.108*	(0.012)	*Not identified*[a]		-0.018	(0.011)

Note: Standard errors are corrected for clustering within the statistical sector and computed with a bootstrap procedure using 250 replications.

* statistical significance at 5% level.

[a] Not identified due to insufficient common support

Despite the extra structure imposed in the third column of Table 4, the estimate of the ATT (-10.7%points) is very similar to the MTE approach of Heckman and Vytlacil (-11.6%points).[25] The parametric model also allows us to compute the ATE and ATNT. We derive a non-significantly ATE of -2.2 %points and a positive, but non-significant ATNT of 0.7 %points. These estimates confirm our interpretation of the increasing MTE in Figure 2 in the previous section: the effect of starting at elite schools is more negative for those students with the highest preference, i.e. the students who actually choose these schools.

Average marginal effects of student characteristics

To gain more insights in the way the treatment effect differs among observable student characteristics, we report the average marginal effects of each background variable in Table 5. As these are all dummy variables, we look at the effect of a change from 0 to 1 for each variable. As these effects differ between individuals, we evaluate them for each student in the sample at their actual realizations of all other variables and report the mean effect.

All background characteristics that predicted worse study outcomes[26] also predict a more negative treatment effect. Most of these variables also have a negative effect on predicting self-selection into elite schools. Nevertheless, we found that the ATT is more negative than the ATNT, suggesting worse outcomes for those who selected into treatment. This can then

[25] The results of the parameters in (6) can be found in Table A2 in Appendix A. We also ran a model without observable characteristics X and obtained very similar results.

[26] This can be concluded from the signs of coefficients in Table A2 and is consistent with the results we found for the simple OLS and IV models in Table 3.

only be explained by the effect of unobservable characteristics that make students enter elite schools but also make them have worse outcomes of elite schools.

Table 5: Average marginal effects and distributional effects

	Coef.	St. error
Average marginal effects of student background		
Mother has no high school degree[a]	-0.016	(0.011)
Mother has high school degree[a]	-0.009	(0.007)
No Dutch at home	-0.023*	(0.011)
Low income	-0.018*	(0.007)
Male	-0.022*	(0.006)
Distributional treatment effects[b]		
Among all students		
% benefit from elite school	16.7*	(0.4)
% suffer from elite school	18.5*	(0.8)
Among students in elite school		
% benefit from elite school	8.4*	(0.7)
% suffer from elite school	19.1*	(0.3)
Among student in non-elite school		
% benefit from elite school	19.7*	(0.7)
% suffer from elite school	18.3*	(0.9)

Note: Standard errors are corrected for clustering within the statistical sector and computed with a bootstrap procedure using 250 replications. Distributional effects derived by accept-reject simulation of error terms with 100 draws for each student in each bootstrap sample.

* statistical significance at 5% level.

[a] Base category = mother has a degree in higher education.

[b] Requires identification of ρ_{10} using factor structure Aakvik *et al.* (2005)

Distributional treatment effects

The extra structure imposed by the factor model of Aakvik *et al.* (2005) is necessary to compute distributional treatment effects. Distributional treatment effects are another measure of the heterogeneity of the treatment effect and identify the fraction of students who experienced or would have experienced (in case of no treatment) a positive or negative treatment effect. The second panel of Table 5 shows the distributional treatment effects for the total population of students, the group of students who started at elite schools, and the group of students who started at non-elite schools. In total, 16.7% of the students benefits from starting at an elite school and 18.5% experiences a negative effect. Although we found a negative ATT in Table 4, the distributional treatment effects show that among the group of treated students, 8.4% still experienced a positive treatment effect. These students would not have graduated on time if they had started at a non-elite school. From the group of non-treated students, a larger fraction of 19.3% would have been better off at an elite school.

This heterogeneity in treatment effects implies that a policy aiming to increase the number of students graduating on time should optimally allocate students to elite and non-elite schools, rather than banning elite schools. A first-best policy that assigns students to the type of school in which they have the highest probability of success would increase the total percentage of students graduating on time from 72.0% to 91.6%, while a policy that bans elite schools entirely only increases it to 74.9%.[27] Note however that this first-best policy is infeasible as it is difficult to identify the students suffering from going to an elite school.

6 Interpretation of results

The combination of a negative ATT and a small and non-significant ATNT in Table 4 implies that the treatment effect is on average more negative for the students who actually started at an elite school. Students who did not choose for an elite school would on average experience no effect from the treatment. The upward sloping MTE curve in Figure 2 also shows that students with a high preference for elite schools (left side of the graph) experience a more negative treatment effect. However, if students would select on the basis of unobserved gains, we would expect a downward sloping MTE curve and ATT>ATNT.[28]

An elite school only offers academic programs, while non-elite schools offer more programs. Tracking decisions may therefore be different in these two types of schools. When students start at an elite school, they have to switch to another school if they do not want to

[27] The total graduation rate without study delay is the weighted average of the rate of an elite school (77.3%) and that of a non-elite school (70.1%): 26.6% x 77.3% + 73.4% x 70.1%=72.0%. If everyone would go to a non-elite school, students in elite schools would on average have a 10.7% points (=ATT) higher outcome: 26.6% x (77.3%+10.7%) + 73.4% x 70.1%=74.9%. If all students would start at the school in which they have the highest probability of success, success rates would increase to: 26.6% x (77.3%+19.1%) +73.4% x (70.1%+19.7%)=91.6%.

[28] Reverse selection on unobserved gains is not uncommon in the literature as it has also been established in Aakvik *et al.* (2005) on the effects of a rehabilitation program on employment and Cornelissen *et al.* (2016b) on the effects of universal child care on school outcomes.

follow the academic track. If switching to another school is costly, students in an elite school have an extra incentive to follow the academic track. Dustmann *et al.* (2016) also find that within an early tracking system, the possibility to change tracks over time is important to mitigate negative long-term effects of tracking. Table 1 shows that students who started at an elite school were more likely to choose for the academic track. The fact that some of these students would have been better off in another track, can explain their lower performance in elite schools. In the next section we empirically verify this hypothesis. Especially the students with a high preference for elite schools might be unwilling to switch to other schools during the course of their high school education, even if the academic track does not suit them well. Therefore, they will be more likely to accumulate study delay or risk to graduate without a degree, instead of switching schools. Note that this high preference does not necessarily need to apply to the student. They can also be pushed by their parents to go to and stay in an elite school.

An alternative explanation for a negative treatment effect is that choosing a better school also causes negative behavioral responses by parents and peers. Pop-Eleches and Urquiola (2013) find that parents of children who make it into elite schools often reduce their effort in helping them. The authors also find that relatively weak students in better-ranked schools feel more marginalized and insecure than similar students in other schools. While they still find a positive effect on educational outcomes, it is possible that these side-effects in our context tip the balance towards a negative effect.

A negative effect can also be explained by differences in grading standards and demands from students. Schools have some autonomy in deciding about grade retention and excluding students from certain programs. It is therefore possible that elite schools more often require weaker students to repeat a grade in order to achieve the required schooling level.

We find that there is substantial heterogeneity in the treatment effect. First, the average treatment effect differs between treated and non-treated students. Second, within both groups of treated and non-treated students, the treatment effect also differs between students. Some students benefit, while others suffer from treatment. Table 5 shows that observed student characteristics explain part of this heterogeneity. However, also unobserved differences between students and schools can result in heterogeneous treatment effects.

7 Alternative outcome variables

In this section we repeat the analysis for two additional outcome variables to give more insights into the interpretation of our results. All results are represented in Appendix B.

7.1 Graduating with at most one year of study delay

Our main results indicate that students starting at elite schools are on average less likely to graduate on time. If graduating from an elite school has positive effects on success in higher education or in the labor market, it might be beneficial to persist in elite schools even if it takes one year longer to graduate from high school. If students value graduating from elite schools more and are therefore willing to study one year longer, we would expect a less negative effect of elite schooling on the probability of obtaining a degree with at most one year of study delay.

We therefore repeat the analysis for this alternative outcome variable. In Figure A3, we still find an upward sloping marginal treatment effect. Table A3 summarizes the treatment effects derived from the three approaches. We still derive a significantly negative ATT of -0.046, implying that students who started at elite schools are on average less likely to obtain a high school degree within 7 years of studying. However, this effect is smaller compared with the previous outcome variable. This seems to suggest that some, but not all students are willing to graduate with one year of study delay in order to obtain a degree from an elite school.

7.2 Switching tracks

A possible explanation for the negative ATT is that within an elite school, it is more difficult for students to choose the track that corresponds best with their ability because only the academic track is offered within these schools. To test this hypothesis, we repeat the analysis with a different outcome variable: the probability to switch to a lower track, or downgrade, during high school. We define downgrading as a switch from either the academic track to another track or from the technical or arts track to the vocational track. Downgrading is common in the Flemish education system. 27% of students starting in elite schools downgrade at least once during secondary education. In non-elite schools, a larger fraction of 35% downgraded at least once. Downgrading can help students switch to a different track, without having to repeat a grade. However, for students who started at elite schools, this implies not only a switch of tracks but also a switch of schools. If there are school switching costs, going to an elite school makes students less willing to switch to other tracks.

The estimated MTE for the probability to downgrade is given in Figure A3. Note that this looks remarkably similar to the MTE for graduating without study delay (Figure 2). We derive a similar ATT of -0.086, i.e. elite schools decrease the probability of downgrading during secondary education. Moreover, especially those students with a low cost of going to an elite school (left side of the graph) are less likely to downgrade. These are the same students that were less likely to graduate on time, when choosing for elite schools. Their high preference for the elite school makes them prepared to stay in the academic track, even though they might fail and accumulate study delay.

8 Sensitivity analysis

In this section we assess how sensitive our results are to the chosen sample of students. First, we restrict the sample to students for which the exogeneity condition of the instrument is most likely to hold. Second, we change the control group (i.e. students in non-elite schools) and include only students in schools that offer an academic track. All results are represented in Appendix C.

8.1 Exogeneity of the instrument

Based on institutional grounds, we argued that distance is an exogenous instrument for school choice in Flanders. School choice is essentially free. For most students several schools are located within commuting distance. From our dataset, we computed that for the median student, six schools are located within 5 km distance. Therefore, it seems unlikely that parents would base their location decision on preferences for certain schools. If our instrument would not be exogenous, we expect this violation of the exogeneity assumption to be stronger in areas with less school alternatives. We therefore repeat the analysis and restrict the sample to students living in areas with many schools. In Appendix C, we assess this issue and repeat the analysis on a subsample of the data where we only include students who have at least four schooling options, located within 5 km distance. The MTE curve in Figure A4 still shows an upward sloping MTE. We find a similar negative ATT of -0.126. In the parametric model, we find a negative ATT of -0.087 (Table A7), and a positive ATNT of 0.044. The latter is slightly larger than in our base model and significant at the 5% level.

8.2 Composition of non-elite schools

There exist two different types of non-elite schools. The first type are general schools that offer programs in the academic track in combination with technical, artistic or vocational programs. Most first year students (60.3%) enroll at this type of schools. The second type of schools specialize in technical, artistic and/or vocational programs and do not offer academic programs. These schools only have a small share of first year enrollment (13.1%). As students who start at a non-elite school without programs in the academic track can be different from students who start at the first type of non-elite schools, we repeat the analysis and only include students who started at an elite school and students who started at a non-elite school that also offers the academic track. We obtain very similar results as in the main analysis. We again find that the treatment effect is heterogeneous and that students who started at elite schools suffered the most as suggested by the upward sloping MTE curve in Figure A5 and the negative ATT between -0.103 (nonparametric model) and -0.096 (parametric model) and a positive ATNT of 0.046 in Table A9.

9 Conclusion

We have studied how the organizational structure of schools influences success in an early-tracking system in secondary education. We compared within- and between-school tracking and analyzed whether students starting at elite schools, schools that specialize and only offer programs in the academic track, are more likely to obtain their high school degree without study delay. We controlled for self-selection of students and allowed for heterogeneous treatment effects. We applied our analysis to the region of Flanders.

We find that there is substantial heterogeneity in the treatment effect of choosing for an elite school, both within and between treatment groups. Although we find a small and non-significant ATE, students with a high preference for elite schools experience large negative effects. We derive an ATT of -11.6 %points, i.e. students starting at elite schools experienced on average an 11.6 %points decrease in their probability to graduate from high school without study delay. This negative effect can be explained by tracking decisions. We find that students are less likely to switch to another track when they started at an elite school because they then have to switch to another school. This interference of school choice with optimal tracking can explain the negative result.

The results have important implications for school assignment policies. The negative ATT implies that banning elite schools would increase the overall rate of students graduating on time. However, the additional result that within the groups of treated and non-treated students a substantial fraction benefits or suffers from elite schools, implies that more gains are possible with a policy that optimally allocates students to elite and non-elite schools. This first best policy might be impossible to achieve in practice as students who would benefit or suffer from attending an elite school are difficult to identify. Better provision of information could help students and their parents.

Further research is needed to consider the effects of elite schools on other outcomes like success in higher education or on the labor market. It is possible that the extra time in high school is worth the cost of going to an elite school if later life outcomes are also affected. The substantial amount of treatment heterogeneity also suggests that for many students, the elite school can be the best option. A potential explanation is heterogeneity in school quality. Further research could address the reasons behind this heterogeneity.

References

Aakvik, A., Heckman, J. and Vytlacil, E. (2005), *Estimating treatment effects for discrete outcomes when responses to treatment vary: an application to Norwegian vocational rehabilitation programs*, Journal of Econometrics 125, (1-2), 15-51

Abdulkadiroğlu, A., Angrist, J. and Pathak, P. (2014), *The Elite Illusion: Achievement Effects at Boston and New York Exam Schools*, Econometrica, 82 (1), 137-196

Altonji, J., Elder, T. and Taber, C. (2005), *Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools*, Journal of Political Economy, 113 (1), 151-184

Altonji, J., Elder, T. and Taber, C. (2005), *An evaluation of instrumental variable strategies for estimating the effects of catholic schooling*, The Journal of Human Resources, 40(4), 791-821

Angrist, J., Pathak, P. and Walters, C. (2013), *Explaining charter school effectiveness*, American Economic Journal: Applied Economics, 5 (4), 1-27

Barrow, L., Whitmore Schanzenbach, D. and Claessens, A. (2015), *The impact of Chicago's small high school inititative*, Journal of Urban Economics, 87, 100-113

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

Card, D. (1995). *Using geographic variation in college proximity to estimate the return to schooling*, Aspects of Labour Economics: Essays in Honour of John Vanderkamp, University of Toronto Press.

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

Carneiro, P., Lokshin, M. and Umapathi, N. (2016), *Average and Marginal Returns to Upper Secondary Schooling in Indonesia: Average and Marginal Returns to Schooling in Indonesia.* Journal of Applied Econometrics. *Forthcoming*

Clark, D. and Del Bono, E. (2014), *The long-run effects of attending an elite school: evidence from the UK*, IRES working paper series, 2014-05

Cornelissen, T., Dustmann, C., Raute, A. and Schönberg, U. (2016), *From LATE to MTE: Alternative methods for the evaluation of policy interventions*, Labour Economics, 41, 47-60.

Cornelissen, T., Dustmann, C., Raute, A. and Schönberg, U. (2016), *Who benefits from universal child care? Estimating marginal returns to early child care attendance*, Working paper.

Cullen, J., Jacob, B. and Levitt, S. (2005), *The impact of school choice on student outcomes: an analysis of the Chicago public schools*, Journal of Public Economics, 89 (5), 729-760

Declercq, K. and Verboven, F. (2015), *Socio-economic status and enrollment in higher education: do costs matter?,* Education Economics, 23 (5), 532-556

Deming, D. (2011), *Better schools, less crime?*, The Quarterly Journal of Economics, 126 (4), 2063-2115

Deming, D., Hastings, J., Kane, T. and Staiger, D. (2014), *School choice, school quality, and postsecondary attainment*, American Economic Review, 104 (3), 991-1013

Dobbie, W. and Fryer, R. (2011), *Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem children's zone*, American Economic Journal: Applied Economics, 3 (3), 158-187

Doyle, J. (2007), *Child protection and child outcomes: Measuring the effects of foster care*, American Economic Review, 97 (5), 1583-1610

Doyle, J. (2008), *Child protection and adult crime: Using investigator assignment to estimate causal effects of foster care*, Journal of Political Economy, 116 (4), 746-770

Dustmann, C., Puhani, P. and Schönberg, U. (2016), *The Long-term Effects of Early Track choice*, Economic Journal. *Forthcoming*

Evans, W. and Schwab, R. (1995), *Finishing high school and starting college: Do catholic schools make a difference?*, The Quarterly Journal of Economics, 110 (4), 941-974

Galasso, A. and Schankerman, M. (2015), *Patents and cumulative innovation: causal evidence from the courts*, The Quarterly Journal of Economics, 130 (1), 317-369

Guyon, N., Maurin, E. and McNally, S. (2012), *The effect of tracking students by ability into different schools: a natural experiment*, Journal of Human Resources, 47 (3), 684-721

Heckman, J. and Vytlacil, E. (1999), *Local instrumental variables and latent variable models for identifying and bounding treatment effects*, Proceedings of the National Academy of Sciences, 96 (8), 4730-4734

Heckman, J. and Vytlacil, E. (2005), *Structural equations, treatment effects, and econometric policy evaluation*, Econometrica, 73 (3), 669–738

Imbens, G. and Angrist, J. (1994), *Identification and Estimation of Local Average Treatment Effects*, Econometrica, 62 (2), 467-475

Lokshin, M. and Sajaia, Z. (2011), *Impact of interventions on discrete outcomes: Maximum likelihood estimation of the binary choice models with binary endogenous regressors.* The Stata Journal. 11 (3), 368-385

Manski, C., Sandefur G., McLanahan, S. and Powers, D. (1992), *Alternative Estimates of the Effect of Family Structure During Adolescence on High School Graduation.* Journal of the American Statistical Association 87 (417), 25-37.

Neal, D. (1997), *The effects of catholic secondary schooling on educational achievement*, Journal of Labor Economics, 15 (1), 98-123

OECD (2014), *How many students are expected to complete upper secondary education?*, Education at a Glance: Highlights, OECD Publishing

Pekkarinen, T., Uusitalo, R. and Kerr, S. (2009), *School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform*, Journal of Public Economics, 93 (7-8), 965-973

Pop-Eleches, C. and Urquiola, M. (2013), *Going to a Better School: Effects and Behavioral Responses*, American Economic Review, 103 (4), 1289-1324

Vademecum Statistische sectoren (2012), http://statbel.fgov.be/nl/binaries/Secteurs%20stat-NL_tcm325-174181.pdf, [23/10/2015]

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

Wiswall, M., Stiefel, L., Schwartz, A. and Boccardo, J. (2014), *Does attending a STEM high school improve student performance? Evidence from New York City*, Economics of Education Review, 40, 93-105

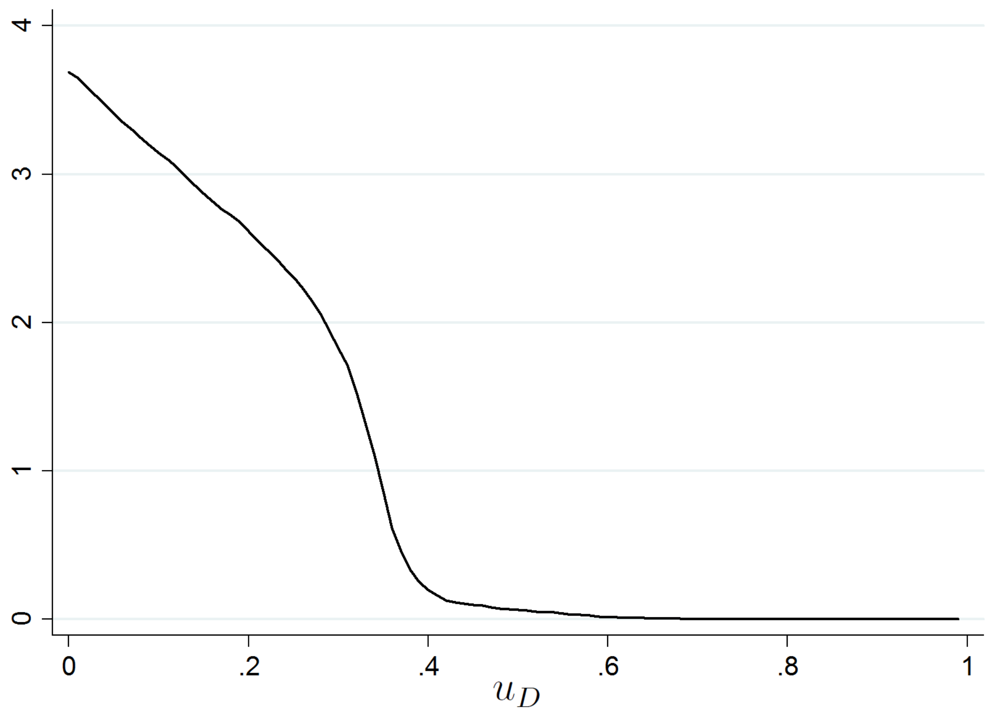
Appendix A: Additional tables and figures

Table A1: Distance to elite schools and pupils' background

	Relative distance
all	-2.6
male	-2.6
female	-2.5
mother has post high school degree	-2.3
mother has high school degree	-2.7
mother has no high school degree	-2.7
dutch at home	-2.6
no dutch at home	-1.8
high income	-2.5
low income	-2.7

Note: Relative distance is the distance to the closest non-elite school minus the distance to the closest elite school and is expressed in km.

Figure A1: Weights for the calculation of the ATT



Note: The weights for the calculation of the ATT given by equation (5). For values above 0.6, we estimated the weights to be smaller then 0.016, but we set them equal to 0 because the MTE is not identified in this region.

Table A2: Selection and outcome equations

	Selection equation		Outcome equations			
			elite school		non-elite school	
Variables	Coef.	St. error	Coef.	St. error	Coef.	St. error
Relative distance	0.158*	(0.004)				
Mother no high school[a]	-0.738*	(0.012)	-0.628*	(0.025)	-0.592*	(0.013)
Mother high school[a]	-0.453*	(0.009)	-0.342*	(0.017)	-0.320*	(0.009)
No dutch at home	0.187*	(0.020)	-0.527*	(0.027)	-0.469*	(0.018)
Low income	-0.174*	(0.010)	-0.183*	(0.018)	-0.132*	(0.009)
Male	-0.053*	(0.007)	-0.487*	(0.012)	-0.427*	(0.007)
Constant	0.019	(0.013)	1.226*	(0.030)	1.216*	(0.014)
ρ_1	0.036	(0.029)				
ρ_0	0.353*	(0.028)				
Log likelihood	-232,591					
Observations	218,211					

Note: Standard errors are corrected for clustering within the statistical sector. The results are obtained from a parametric model where we assume that the error terms are jointly normally distributed. We estimated the model with the user written "switch_probit" command (Lokshin and Sajaia, 2011).

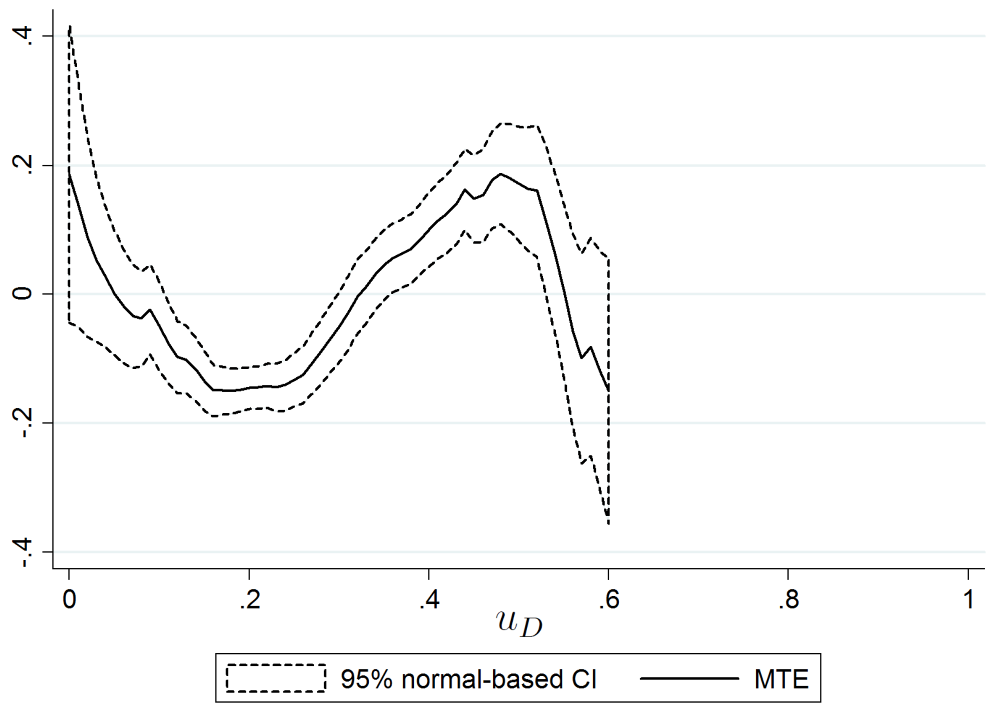
* statistical significance at 5% level.

[a] Base category = mother has a degree in higher education.

Appendix B: Alternative outcome variables

Graduating with at most one year of study delay

Figure A2: MTE of elite schools to graduate with at most one year of study delay



Note: MTEs are calculated using a local polynomial regression of degree 2 with an Epanechnikov kernel (bandwidth 0.109). Standard errors are are computed with a bootstrap procedure (250 replications) and clustered within the statistical sector. No results for $u_D > 0.60$ due to insufficient common support.

Table A3: Treatment effects (graduate with at most one year of study delay)

	Traditional IV		Nonparametric approach		Parametric approach	
	Coef.	St. error	Coef.	St. error	Coef.	St. error
Treatment effects						
ATT	-0.070*	(0.009)	-0.046*	(0.016)	-0.067*	(0.005)
ATNT	-0.070*	(0.009)	*Not identified*[a]		0.022*	(0.010)
ATE	-0.070*	(0.009)	*Not identified*[a]		-0.001	(0.008)

Note: Standard errors are corrected for clustering within the statistical sector and computed with a bootstrap procedure using 250 replications.

* statistical significance at 5% level.

[a] Not identified due to insufficient common support

Table A4: Additional results (graduate with at most one year of study delay)

	Coef.	St. error
Average marginal effects of student background		
Mother has no high school degree[a]	0.016	(0.011)
Mother has high school degree[a]	0.005	(0.005)
No Dutch at home	-0.030*	(0.010)
Low income	-0.012*	(0.005)
Male	-0.014*	(0.005)
Distributional treatment effects[b]		
Among all students		
% benefit from elite school	8.8*	(0.2)
% suffer from elite school	8.9*	(0.7)
Among students in elite school		
% benefit from elite school	1.9*	(0.4)
% suffer from elite school	8.6*	(0.2)
Among student in non-elite school		
% benefit from elite school	11.3*	(0.3)
% suffer from elite school	9.1*	(0.8)

Note: Standard errors are corrected for clustering within the statistical sector and computed with a bootstrap procedure using 250 replications. Distributional effects derived by accept-reject simulation of error terms with 100 draws for each student in each bootstrap sample.

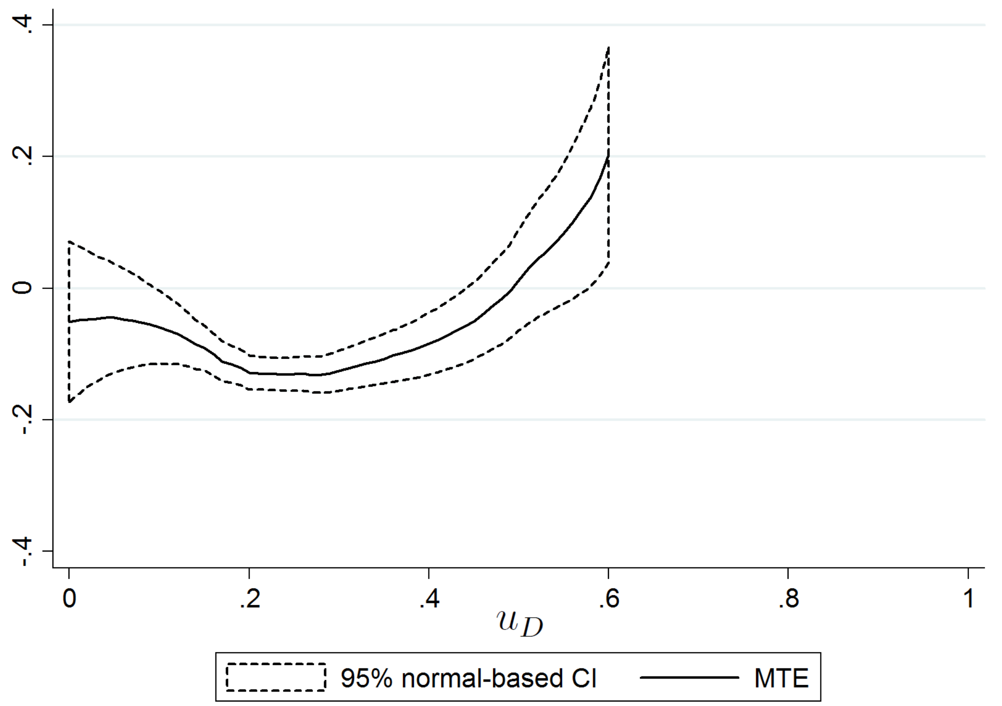
* statistical significance at 5% level.

[a] Base category = mother has a degree in higher education.

[b] Requires identification of ρ_{10} using factor structure Aakvik *et al.* (2005)

Switching track

Figure A3: MTE of elite schools to switch tracks



Note: MTEs are calculated using a local polynomial regression of degree 2 with an Epanechnikov kernel (bandwidth 0.189). Standard errors are are computed with a bootstrap procedure (250 replications) and clustered within the statistical sector. No results for $u_D > 0.60$ due to insufficient common support.

Table A5: Treatment effects (switch tracks)

	Traditional IV		Nonparametric approach		Parametric approach	
	Coef.	St. error	Coef.	St. error	Coef.	St. error
Treatment effects						
ATT	-0.067*	(0.011)	-0.086*	(0.018)	-0.087*	(0.012)
ATNT	-0.067*	(0.011)	*Not identified*[a]		-0.007	(0.014)
ATE	-0.067*	(0.011)	*Not identified*[a]		-0.028*	(0.011)

Note: Standard errors are corrected for clustering within the statistical sector and computed with a bootstrap procedure using 250 replications.

* statistical significance at 5% level.

[a] Not identified due to insufficient common support

Table A6: Additional results (switch tracks)

	Coef.	St. error
Average marginal effects of student background		
Mother has no high school degree[a]	0.108*	(0.011)
Mother has high school degree[a]	0.060*	(0.008)
No Dutch at home	-0.048*	(0.010)
Low income	0.038*	(0.007)
Male	0.064*	(0.005)
Distributional treatment effects[b]		
Among all students		
% benefit from elite school	20.3*	(0.7)
% suffer from elite school	23.1*	(0.5)
Among students in elite school		
% benefit from elite school	16.6*	(0.3)
% suffer from elite school	25.4*	(0.8)
Among student in non-elite school		
% benefit from elite school	21.6*	(1.0)
% suffer from elite school	22.3*	(0.5)

Note: Standard errors are corrected for clustering within the statistical sector and computed with a bootstrap procedure using 250 replications. Distributional effects derived by accept-reject simulation of error terms with 100 draws for each student in each bootstrap sample.

* statistical significance at 5% level.

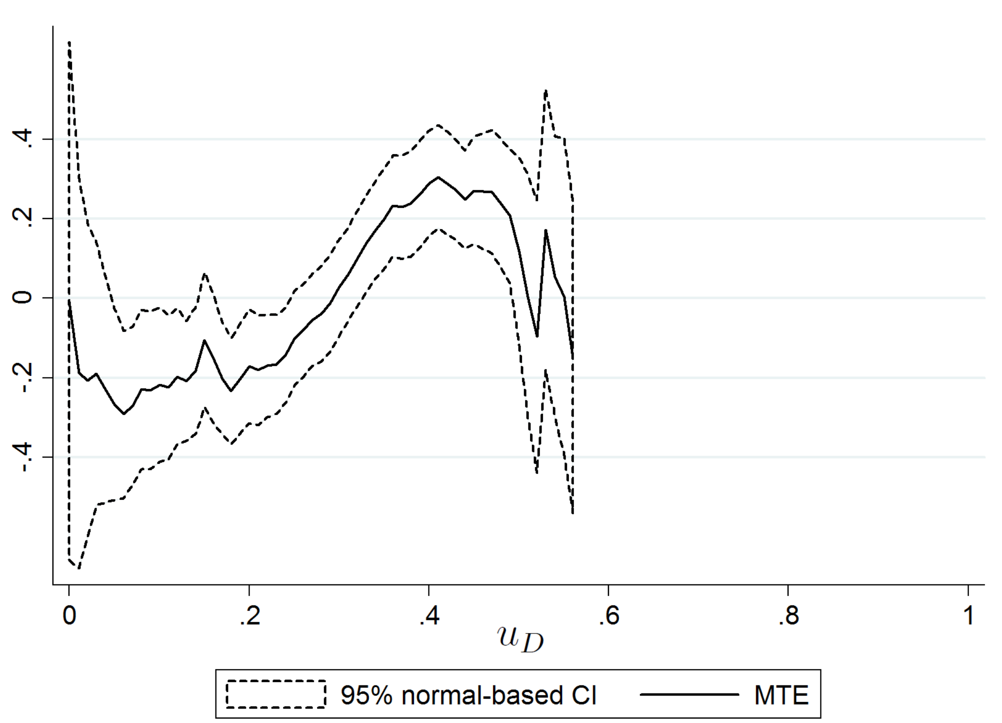
[a] Base category = mother has a degree in higher education.

[b] Requires identification of ρ_{10} using factor structure Aakvik *et al.* (2005)

Appendix C: Sensitivity analysis

Exogeneity of the instrument

Figure A4: MTE of elite schools (exogeneity of the instrument)



Note: MTEs are calculated using a local polynomial regression of degree 2 with an Epanechnikov kernel (bandwith 0.084). Standard errors are are computed with a bootstrap procedure (250 replications) and clustered within the statistical sector. No results for $u_D > 0.56$ due to insufficient common support.

Table A7: Treatment effects (exogeneity of the instrument)

	Traditional IV		Nonparametric approach		Parametric approach	
	Coef.	St. error	Coef.	St. error	Coef.	St. error
Treatment effects						
ATT	-0.085*	(0.015)	-0.126*	(0.030)	-0.087*	(0.014)
ATNT	-0.085*	(0.015)	*Not identified*[a]		0.044*	(0.018)
ATE	-0.085*	(0.015)	*Not identified*[a]		0.007	(0.014)

Note: Standard errors are corrected for clustering within the statistical sector and computed with a bootstrap procedure using 250 replications.

* statistical significance at 5% level.

[a] Not identified due to insufficient common support

Table A8: Additional results (exogeneity of the instrument)

	Coef.	St. error
Average marginal effects of student background		
Mother has no high school degree[b]	-0.005	(0.013)
Mother has high school degree[b]	-0.001	(0.008)
No Dutch at home	-0.009	(0.013)
Low income	-0.020*	(0.008)
Male	-0.015*	(0.006)
Distributional treatment effects[c]		
Among all students		
% benefit from elite school	18.6*	(0.5)
% suffer from elite school	18.0*	(0.8)
Among students in elite school		
% benefit from elite school	10.8*	(0.9)
% suffer from elite school	19.5*	(0.4)
Among student in non-elite school		
% benefit from elite school	21.7*	(0.7)
% suffer from elite school	17.4*	(0.8)

Note: Standard errors are corrected for clustering within the statistical sector and computed with a bootstrap procedure using 250 replications. Distributional effects derived by accept-reject simulation of error terms with 100 draws for each student in each bootstrap sample.

* statistical significance at 5% level.

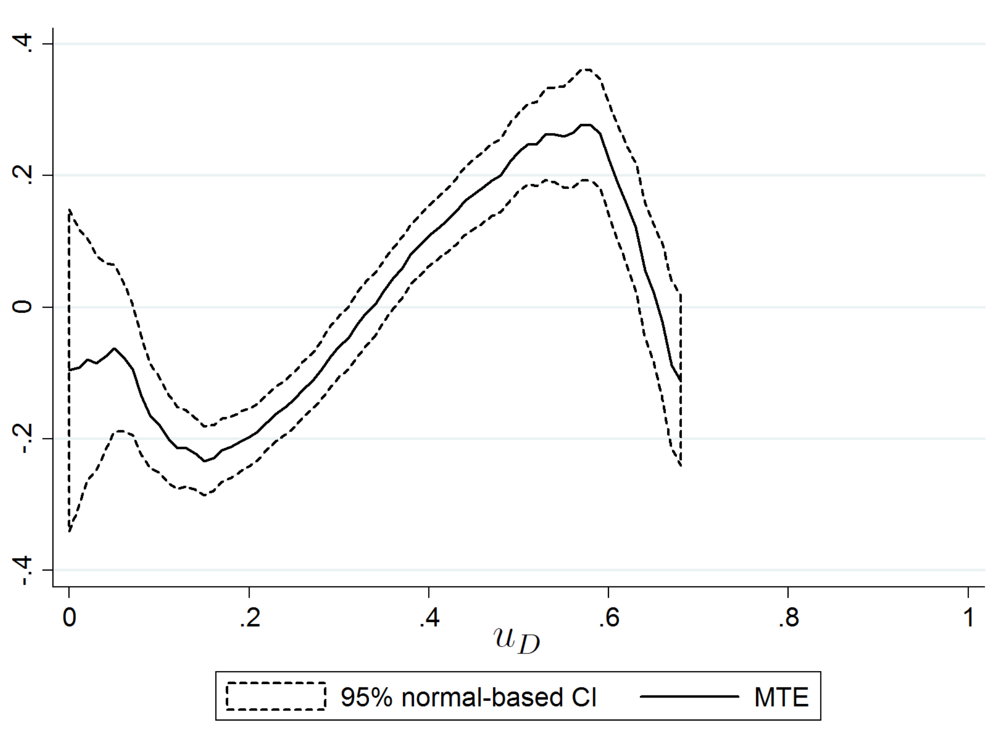
[a] An advantaged student is defined as a student with all dummy variables equal to 0, this means a female, advantaged student that started high school at the age of 12.

[b] Base category = mother has a degree in higher education.

[c] Requires identification of ρ_{10} using factor structure Aakvik *et al.* (2005)

Composition of non-elite schools

Figure A5: MTE of elite schools (composition of non-elite schools)



Note: MTEs are calculated using a local polynomial regression of degree 2 with an Epanechnikov kernel (bandwidth 0.145). Standard errors are computed with a bootstrap procedure (250 replications) and clustered within the statistical sector. No results for $u_D > 0.68$ due to insufficient common support.

Table A9: Treatment effects (composition of non-elite schools)

	Traditional IV		Nonparametric approach		Parametric approach	
	Coef.	St. error	Coef.	St. error	Coef.	St. error
Treatment effects						
ATT	-0.098*	(0.011)	-0.103*	(0.020)	-0.096*	(0.009)
ATNT	-0.098*	(0.011)	*Not identified*[a]		0.046*	(0.011)
ATE	-0.098*	(0.011)	*Not identified*[a]		0.002	(0.008)

Note: Standard errors are corrected for clustering within the statistical sector and computed with a bootstrap procedure using 250 replications.

* statistical significance at 5% level.

[a] Not identified due to insufficient common support

Table A10: Additional results (composition of non-elite schools)

	Coef.	St. error
Average marginal effects of student background		
Mother has no high school degree[b]	0.010	(0.009)
Mother has high school degree[b]	0.006	(0.006)
No Dutch at home	-0.029*	(0.013)
Low income	-0.005	(0.007)
Male	-0.010*	(0.005)
Distributional treatment effects[c]		
Among all students		
% benefit from elite school	17.3*	(0.4)
% suffer from elite school	18.9*	(0.6)
Among students in elite school		
% benefit from elite school	9.3*	(0.6)
% suffer from elite school	18.9*	(0.3)
Among student in non-elite school		
% benefit from elite school	20.8*	(0.5)
% suffer from elite school	16.2*	(0.6)

Note: Standard errors are corrected for clustering within the statistical sector and computed with a bootstrap procedure using 250 replications. Distributional effects derived by accept-reject simulation of error terms with 100 draws for each student in each bootstrap sample.

* statistical significance at 5% level.

[a] An advantaged student is defined as a student with all dummy variables equal to 0, this means a female, advantaged student that started high school at the age of 12.

[b] Base category = mother has a degree in higher education.

[c] Requires identification of ρ_{10} using factor structure Aakvik *et al.* (2005)