KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

FACULTY OF ECONOMICS AND BUSINESS

Determinants of secondary school preferences in Flanders

The case of Ghent

MASTER OF ECONOMICS

R0463704

Thesis submitted to obtain the degree of Master

Willem Konings

Major General Economics

Promoter: Prof. Dr. Erwin Ooghe
Assistant: Thomas Wouters

Academic year 2016-2017

KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

FACULTY OF ECONOMICS AND BUSINESS

Determinants of secondary school preferences in Flanders

The case of Ghent

This thesis looks at determinants of secondary school preferences in Ghent (East Flanders), using a conditional logit model. We found that school quality is likely to be the most important positive determinant of preferences. Furthermore, specific preferences for school composition according to the socio-economic status of the student (SES-status) exist. Finally, we found that the distance to the schools is an important factor, whose importance increases for disadvantaged students. In the following step, the analysis was conducted separately for three tracks existing in Flemish secondary school education (general secondary education (ASO), technical secondary education (TSO) and vocational secondary education (BSO)). We also compared the results from the smaller sample consisting of the ASO track participants to the results from the whole sample. The results are rather suggestive due to approximations, but indications of higher importance of distance and school quality were found for the ASO-track, higher importance of school composition for the BSO/TSO-track.

MASTER OF ECONOMICS

Willem Konings

R0463704

Thesis submitted to obtain
the degree of Master

Major General Economics

Promoter: Prof. Dr. Erwin Ooghe
Assistant: Thomas Wouters

Academic year 2016-2017

Table of contents

1 Preface

Writing a master thesis is not straightforward and it is also time-consuming process. Therefore, I would like to thank some people for their support, especially during the second semester of the academic year.

Firstly, I would like to thank my promotor and co-promotor. They were always prepared to help me while I was doing research. They also motivated me to regularly work on my thesis during the second semester and I believe I have become a better researcher now and a more disciplined person in general.

Secondly, I would like to thank everyone else who supported me. Some people proofread my thesis which I really appreciated, others made substantive suggestions for the analysis I performed.

Finally, I would like to thank family and friends, all the people whom I'm closely connected with. I had a wonderful academic year. Indirectly, this was invaluable for my work.

2 Introduction

This thesis examines determinants of school preferences in Flanders using a conditional logit model. These preferences are defined by looking at choices of secondary schools. The analysis is especially focussing on secondary education in Ghent, while including the distance to the school, school quality and composition. This section provides an introduction into the topic and its importance. Furthermore, the previous literature on the subject is discussed before explaining the specific contribution of this thesis to the discussion. Finally, we look at possible constraints for the analysis.

2.1 The importance of school preferences in Flanders

Socio-economically speaking, the school landscape in Flanders is rather segregated, so the topic is highly relevant to investigate[1]. Segregation itself can be both intrinsically and instrumentally bad for a society. From this follow clear policy implications. This is further discussed in the following section.

Constraints on school choice as seen in the USA or England increase the importance of the topic However, this choice is not strongly constrained in Flanders. Only minor constraints such as on school capacity exist[2]. School choice is a term often used in the international literature to describe ways to improve the process of school enrolment, enlarge the amount of schools from which to choose, … A later section will also further elaborate on this.

2.1.1 Segregation and parental preferences for school composition

Important research about preferences for school composition and its determinants is provided by Wouters & Groenez (2014). The results from their analysis suggest

---

[1] Correspondence with professor Ooghe (University of Leuven).

[2] We will explicitly talk about school preferences instead of school choice most of the time due to its higher importance for our analysis.

that increasing school segregation can partly be explained by higher residential segregation and partly by parental preferences for school composition. If the number of pupils with low-educated parents and the number of pupils who do not speak Dutch at home rises in a certain school, it becomes more likely that pupils from parents with a different socio-economic status (SES-status)[3] switch to another school.

However, the main question is whether segregation is intrinsically and/or instrumentally bad for society[4]. Here, two arguments against segregation are commonly mentioned (see Wouters & Groenez, 2015, for a recent overview of the literature).

The first argument takes a democratic or citizenship approach. Integration is regarded as a condition for a well-functioning society, shared identity and solidarity. Students should be prepared to function in a diverse society and therefore there exists an important role for the educational environment (Wouters & Groenez, 2015).

A second argument deals with learning outcomes. They seem to depend on the individual's background and school characteristics such as school size as well as group effects both in schools and in neighbourhoods. Cognitive effects and behaviour effects are directly impacted by these group- or peer-effects. Finally, there is also an indirect role of group composition: schools with a large variety of their pupil's SES have more middle-class parents who are closely connected to the school. This helps to supervene on the quality of education as well. Therefore, these schools have better teachers and higher expectations about their students because of the diverse school composition (Wouters & Groenez, 2015).

Seen from the first perspective, segregation has always downsides. Whereas the second perspective attaches a negative value to segregation if it hinders efficient learning outcomes and educational opportunities. Previous empirical research has

---

[3] Note that this does not necessarily mean "higher".

[4] Intrinsically means "in itself", instrumentally here focusses whether segregation could be a condition to achieve some other good.

shown that cognitive group effects are relatively small[5], while behaviour effects seem to be more important (Wouters & Groenez, 2015).

Further, literature has showed that segregation has some advantages as well. For instance, van Kempen & Özüekren (1998) have argued that "the existence, nurturing and development of social contacts (…) can be seen as an extremely useful aspect of spatial segregation and concentration" (van Kempen & Özüekren, 1998, p. 1635). However, segregation is mostly argued as an avoidable outcome which cannot be encouraged in the majority of situations (Wouters & Groenez, 2015). Therefore, it is pleaded against a segregated school landscape in this thesis.

We will now look more specifically at the subject of determinants of school preferences in existing literature.

2.2 Results showing up in literature: determinants of school preferences

2.2.1 A general overview of the literature

Some important studies about determinants for school preferences include Schneider & Buckley (2002), Rothstein (2006), Black (1999), Ball, Bowe, & Gewirtz (2006), Hastings & Weinstein (2008) Burgess, Greaves, Vignoles, & Wilson (2011) and Burgess, Greaves, Vignoles, & Wilson (2015). For Flanders, a recent study on the topic includes Wouters & Groenez (2014). A related study on school preferences in Flanders is Pinxten, Fraine, Noortgate, Damme, & Anumendem (2012).

A large number of results in the literature about determinants of school preferences, primary education, for example Burgess et al. (2015), and secondary education, e.g. Lankford & Wyckoff (1992)[6] , appear similar, although this is not always the case. Burgess et al. (2015) find that parents highly value a school's academic performance using a conditional logit model. However, parents also

[5] This certainly true with respect to the effects of individual background.

[6] Dealing with both types.

value socio-economic composition and distance as determinants of school preferences. Another conclusion of the study is that more advantaged parents attach more value to academic performance than parents with a lower SES. Hastings and Weinstein (2008) find comparable results in their study. It argues that proximity and academic achievement are important determinants using a mixed logit demand model. It also states that preferences for schools are heterogeneous across different SES-groups. Low-income families attach much less value on academic quality. Another important study is the one of Schneider and Buckley (2002), which uses an online school database in Washington for their research. According to these authors, characteristics of parents matter for search behaviour. They conclude that parents base their preferred school for a substantial part on demographic characteristics of students of a school. A further study pointing towards the importance of social class on school preferences is from Ball et al. (2006). However, this study has been criticized later in the British Journal of Sociology of Education (in 1997).[7] The study of Ball et al. (2006) investigates 137 interviews with parents over a 39-month period, asking detailed questions about factors they find important in schools. The sample of their study is derived from London. According to Tooley (1997), some features of this city might make the results of Ball et al. (2006) hard to generalize to another context (e.g. nationwide).

A study by Rothstein (2006) presents a rather different view on the importance of school quality. The study finds no significant valuations of school effectiveness for parents (it is not strongly evidence-based with respect to their choice process). Rothstein (2006) uses a more indirect approach compared to the previously mentioned studies. It looks at income of families and a measure of effectiveness of schools. The question here is whether parents choose neighbourhoods for this effectiveness of local schools. The author finds little evidence for this. However, Black (1999) studies elementary education using a very similar approach and does find significant effects. Housing prices seem to be positively correlated with school

[7] At the time of writing, Ball et al. (2006) was an influential work on school choice reforms in England and Wales, but its findings are challenged on methodological grounds. They use a qualitative method to generalize their findings for instance, but they have no clear support to do this (Tooley, 1997).

quality. The finding presented is robust to several sensitivity checks used in the study.

A number of studies have also focused on the effects of information given to parents. They focus on the relationship between this information for different social groups and their school preferences. For instance, Schneider & Buckley (2002) find that unconstrained choice might negatively affect the outcome in the distribution of students. They use an Internet-based tool about a city's schools to investigate parental preferences about schools. This information is revealed through information search patterns and the authors compare these findings with the existing literature, which is largely based on telephone interviews. If parents used the Internet-based tool, they seemed to put more importance on school composition as a determinant for which school to choose from. This implies that school preferences might be driven by student demographics (race and social class) rather than academics, which in turn gives no pressure to schools to improve quality. Therefore, the authors are in favour of “controlled choice”: a form of school choice that imposes regulations, to have a balance between equity and efficiency. In line with this, Hastings and Weinstein (2008) look at the effect of providing direct information about school test scores on school preferences. Their results suggest that school choice will most effectively increase academic achievement for disadvantaged students, if parents have easy access to test score information and if there are good options from which to choose. The results of this study are different from Schneider and Buckley (2002) as a public-school-choice plan is investigated here. In the study of Schneider and Buckley (2002), no such governmental guidance is given to parents.

A more general conclusion from Musset (2012) is as follows: “Choice programmes can be perceived as leading to a general improvement in the quality of education, and fostering efficiency and innovation. On the other hand, school choice critics suggest that school choice can exacerbate inequities, as it increases sorting of students between schools based on their socio-economic status, their ethnicity and their ability, and quality can become increasingly unequal between schools” (Musset, 2012, p. 4). This suggests that indeed school composition can be an important determinant for school preferences. However, other findings suggest that quality of education does play a role as well. Schools might have an incentive to improve the quality of their education because parents are willing to send their child to a school with high educational standards.

2.2.2 Theoretical and empirical results in the Flemish context

Two recent studies about school preferences in Flanders are the study of Wouters & Groenez (2014) and the related study of Pinxten et al. (2012).

In line with Wouters & Groenez (2015), Wouters & Groenez (2014) find that small differences in preferences for group composition can be sufficient to have segregation in schools. To come to this conclusion, they use theoretical models[8] based on Schelling and Tiebout, which are social interaction models[9]. Pinxten et al. (2012) investigates option choice[10] in secondary school, thereby looking at student-level (for example gender, achievement, ability and socioeconomic status) and school-level (this means gender composition, maths composition, and SES composition) determinants. It performs a multinomial multilevel analysis and finds that option choice can be better explained by student-level rather than school-level predictors. This study might be related to our thesis, as different options from which to choose in secondary school can influence school preferences of parents.

Further studies about education in Flanders exist as well, a substantial part of them sociological. A bulk of literature has dealt with social inequalities, social class and tracks in secondary education, school quality and its importance for school preferences and determinants for learning outcomes and wellbeing in education. (for example Tan, 1998; Groenez, Van den Brande, & Nicaise, 2004; Sierens, 2006).

Tan (1998) finds that social and cultural barriers are two of the main reasons for not choosing general secondary education amongst students. Similarly, Groenez et al. (2004) show that a lower SES-status of students' parents causes learning disadvantages and influences the study orientation. Another study trying to explain

---

[8] We should note that the models used by this study are developed in the context of the US.

[9] . However, this is a more general conclusion not specifically pointing to Flanders, but which can be applied there anyway.

[10] This means an option within a choice track like general secondary education.

choice tracks in secondary school is Sierens (2006), which states that there is a causal relationship between free school choice and ethnical segregation. Perceptions about occupational prestige and the role of gender, for instance, can play a role in study orientation. The study also finds that students go to schools where the SES-composition is "similar" to the one of their home situation (e.g. highly educated parents want their children to go to schools where the other students' parents have finished higher education as well).

As already seen very briefly, there exist more comprehensive works partly dealing with the situation in Belgium (Flanders) as well. An example we gave was Musset (2012).

2.3 What this research adds to the existing literature

For several reasons, this thesis can add to the current understandings about determinants of school preferences in Flanders.

Firstly, we dispose of a relatively rich administrative dataset spanning more than ten school years of observations. This enables us to perform different sorts of rather accurate regressions, using a conditional logit model. This framework is different from other studies about school preferences that currently exist in Flanders. With our dataset[11] we have for example the opportunity to look at the impact of quality of education on school preferences in an interesting way[12]. Another option with our data is to control for the ability of a student in her/his school choice process. We dispose of data about "acquired certificates" in secondary education[13].

Secondly, research about the topic of (secondary) school preferences in Flanders is relatively scarce compared to some other countries. Two such examples are the

---

[11] This is discussed in further detail below (when looking at the advantages of the data).

[12] See footnote 4.

[13] Translated in Dutch: "behaalde attesten".

US or England, on which a substantial part of the literature is built. This is especially true for secondary education in the Flemish context.

Investigating secondary education differs from investigating primary education: we should take into consideration different choice tracks (art education, general secondary education and vocational secondary education). This might influence school preferences of parents: a school with more choice tracks can potentially attract more students. Some choice tracks can also have higher demand than others in secondary school. There might also exist a correlation between amount of choice tracks offered by a school/the specific choice tracks and its quality of education.[14] To sum up, the determinants of secondary school preferences could be different from those of primary education. The effect of a specific determinant like quality, distance or school composition can also be different. For example, distance might matter less for secondary school preferences, as younger children are more dependent on their parents to go to school.

We will later observe that comparing school preferences across tracks is not so straightforward. We already mention that due to data limitations the importance of school quality across tracks was very difficult to compare.[15]

2.4 Possible constraints for the analysis

In Belgium, free school choice exists. There are no catchment areas. However, there are still capacity constraints. Belgium is even one of the OECD-countries where freedom in school choice is highest among public schools (Musset, 2012).[16] School choice is on a “first come, first served” basis. However, from 2003 onwards schools should give priority to siblings from children already in the school and children from a disadvantaged background. Since 2008, schools can even use geographic criteria if demand exceeds capacity for a school (Musset, 2012). To

[14] Other assumptions about tracks will be discussed later.

[15] We will further elaborate on this later.

[16] In Belgium there are government-funded schools and there also exists free education (Vlaamse Overheid, 2017). Most students attend private government-funded schools (Musset, 2012).

continue, central registration registers, from 2009 onwards in Ghent, later in the rest of the country, and the system of double quota, from 2013 onwards in local consultation platform areas ("LOP-gebieden"), have been introduced.[17] The objective of both measures were to reduce social inequalities in education. However, we believe these constraints will not importantly affect our results[18].

Another factor potentially influencing parental choice behaviour is the "middle school institution". Aim of this institution is to offer students more equal educational opportunities. The first two years of secondary education are the same for every student in these institutions (comprehensively organised). This is done to stimulate talent development by students and influencing later choices of students in the third year of secondary education (when they decide which track to follow) (Vanderhoeven, 2004). For example, parents of low SES-status might have a higher preference for institutions offering vocational and/or technical secondary education. This is an outcome this measure can try to encounter: students might base their track choice more on their ability if they have followed the first two years of secondary education in a middle school institution.

To conclude, we believe that the percentage of disadvantaged students might be higher in Ghent than the average for Flanders. In 2012, 40% of students attending secondary education in Flanders were disadvantaged (De Morgen, 2017). Looking at the same year in our dataset of first-year enrollers, this percentage is 58,56%. This is considerably higher. We take this into account while performing our analysis and try to translate our results to the Flemish context.

Section three will now come up with hypotheses: more detailed research questions we will investigate and their expected results. Section four deals with a basic

---

[17] With the system of double quota, the aim is to get an equal distribution of so-called "indicator students" and "non-indicator students". The first group meets criteria of deprivation (e.g. mother has not finished higher education) (Vlaamse Overheid, 2017). The general objective was to boost the allocative efficiency of students across schools while trying to deal with social segregation (Vlaamse Overheid, 2017). With the central registration registers, the Belgian government aimed at reducing inefficiencies like "camping" before a school to get accepted. One should also note that Flanders and Wallonia do not have the same educational policy, although there is considerable overlap (e.g. in terms of funding).

[18] We will further elaborate on this later.

economic and econometric model, presented to justify our approach of analysis. Section five comes up with the data we will use based on the previous framework. Finally, our results are presented and discussed thereafter (section six and seven).

3 Hypotheses

Building on a vast part of the literature, this section will present estimates on the effects of some important determinants of school preferences. We will look at quality of education, socio-economic composition and distance and discuss for each of these determinants their predicted effects. These determinants are taken because most of the data contains variables belonging to either one of those categories. As seen earlier, a substantial part of the literature on school preferences is specifically pointing to the importance of those three determinants as well.

3.1 Quality of education

Looking first at quality of education, two relatively closely linked studies, Burgess et al. (2015) and Hastings & Weinstein (2008) found significant positive effects on school choice. Musset (2012) also came to this conclusion building on the literature at that time. Although some research nuanced its importance (e.g. Rothstein, 2016; Schneider & Buckley, 2012; Wouters & Groenez, 2014), we will state the following hypothesis:

Hypothesis 1a (quality of education):

School preferences are positively correlated with the quality of education.

However, there are differences across SES-groups as we saw earlier. For instance, Burgess et al. (2015) pointed to the importance of belonging to a group with a lower SES on attaching value to quality of education. This leads us to the following hypothesis:

Hypothesis 1b (quality of education):

Parents with a lower SES-status attach relatively less importance to quality of education than parents with a higher SES-status.

In our conditional logit model, effects can only be interpreted relative with respect to other determinants of school preferences. This would imply, for example, that parents with a higher SES-status value quality of education more with respect to distance than parents with a lower SES-status, but other determinants can be seen with respect to this as well.

3.2 School composition

A bulk of literature referred to the importance of socio-economic composition for school preferences (e.g. Schneider & Buckley, 2002; Ball et al., 2006: Musset, 2012; Wouters & Groenez, 2014). We will therefore hypothesize that sorting across schools according to SES-status takes place. This leads us to:

Hypothesis 2a (school composition):

Students who are disadvantaged in some respect will more likely go to a school institution with a high percentage of similar students (i.e. also disadvantaged in the same respect).

Some research in Belgium (Flanders) also investigated the question whether parents with a lower SES-status have preferences for certain choice tracks in secondary school (see literature review). It has generally concluded that it plays a role: there is a positive correlation between parents with a higher SES-status and choice for general secondary education (ASO). Tan (1998) finds that social and cultural barriers are two of the main reasons for this. The study shows that pupils of low-educated parents are more likely to go to vocational secondary education. However, some other factors might play a role as well. One of these factors could be the role of teacher recommendations by choosing a choice track in the transition from primary to secondary education (Boone & Houtte, 2012). There is also a possible role for the ability of students as to explain why general secondary education is not chosen. This is something we can partially try to control for by means of “acquired certificate”, which we already briefly mentioned. Taken all these constraints into account we will state the following:

Hypothesis 2b (school composition):

Students of parents with a lower SES-status opt less for general secondary education.

3.3 Distance

On proximity to the school, the literature seems to agree that distance has significant negative effects on school preferences (e.g. Burgess et al., 2015: Hastings & Weinstein, 2008). We will therefore formulate the following hypothesis:

Hypothesis 3a (distance):

Distance is negatively correlated with school preferences.

However, some literature also noted that this effect becomes smaller if the school is further away (e.g. Burgess et al., 2015), leading to:

Hypothesis 3b (distance):

If a school is further away, the negative effect of distance becomes significantly less pronounced.

3.4 Omitted variables

As omitted variable bias is commonly inherent to every econometric analysis, we will end this section with a concise discussion about potential omitted variables.

Apart from the three determinants we will mostly consider in our analysis, Burgess et al. (2015) and Schneider & Buckley (2002) already pointed to the importance of other factors. Parents might value some specific characteristics of a school like IT-infrastructure, school size (Opdenakker & Van Damme, 2007), having facilities to stay during the week or knowledge that a substantial number of children in the (close) neighbourhood go to the same school[19]. To continue, the specific location of a school can play a role (e.g. near busy traffic or with more quiet surroundings like a green environment). Preferences of some parents might also be non-standard, some students need specific needs (for example those in normal secondary education who are physically/mentally constrained) (Burgess, Greaves, Vignoles, & Wilson, 2015).

[19] Maybe for some other reasons, of course.

To illustrate, school infrastructure can be positively correlated with a recommendation by the inspectorate (e.g. a positive advice). If parents are not aware of the recommendation on a school, but value school infrastructure, the coefficient on “positive advice” might be upwardly biased if a measure of school infrastructure is not included in our regressions.

We continue our analysis by giving a brief overview on the economic and econometric model applied for investigating our research questions.

4 Model

4.1 Economic model

Parents choose the school that maximises their utility. For the set-up of the economic model, the approach of a utility-maximising consumer is followed (McFadden & McFadden, 1977). There are several schools which we can index by $s = 1, \ldots n$. Suppose a parent i derives utility from a school s, call it U_{is}. This utility depends on characteristics of the school, parent and pupil. For a specific individual, we have: $U_{is} = V_{is} + \varepsilon_{is}$. Here is V_{is} the so-called deterministic component of utility depending on the chosen parent and the preferred school. ε_{is} represents a random component (with the same subscript). The deterministic components are about parent and school characteristics, we can write U_{is} as follows:

$$U_{is} = X_{is}\beta + W_i \propto + W_i X_{is}\gamma + \varepsilon_{is} \; with \; s = 1 , \ldots , n$$

The first important element is X_{is}, a vector of school attributes. Element β is a vector of coefficients corresponding to school attributes (also $\propto$ and γ represent a vector of coefficients on their respective variables).

The second term, W_i , shows a vector of household characteristics constant across choices. There is an interaction with the vector of school attributes included to have different preferences for school attributes for different family types (interaction effects). To give one example, parents with a lower socioeconomic status may value academic quality as less important than parents with a higher socioeconomic status (Burgess et al., 2015). A further discussion on interaction effects will appear in a later section where we present our regressions and their results.

The third term is random component ε_{is} which includes for instance unobserved characteristics on which parents base their school preferences. An example might be extra services the school offers like the possibility to stay during the week (a boarding school) (Burgess et al., 2015, pp. 1265-1266).

The probability that a school s is for parent i its most preferred school is captured by:

$P_{is} = \Pr(U_{is} > U_{it}) \, \forall t \neq s$ (Burgess et al., 2015, pp. 1266).

As seen earlier, examples can be given where a parent cannot take her/his first choice (e.g. capacity constraints). In this case the equation will not hold; there might be a school t which yields a lower utility level, but a higher choice probability.

4.2 Econometric model

As already noted, we will perform an econometric analysis to look at determinants of school preferences. Therefor we use a conditional logit model. Here, independent variables may vary across alternatives and individuals, i.e., the vector of school attributes varies across the choices as well as between families (Burgess et al., 2015). The model is generally presented as follows, using the approach of McFadden and assuming error terms with a standard Type 1 extreme value distribution[20]:

$P_{is} = \frac{e^{X_{is}\beta + W_i \propto + W_i X_{is}\gamma + \varepsilon_{is}}}{\sum_{l=1}^{n} e^{X_{il}\beta + W_i \propto + W_i X_{il}\gamma + \varepsilon_{il}}} \ with\ s = 1, \dots, n$ (Burgess et al., 2015, pp.1269).

4.2.1 Why we choose this model[21]

This section ends by briefly discussing the advantages and downsides/limitations of the econometric model.

To begin, the conditional logit model can control for unobserved student characteristics. Another clear advantage is that regressors vary across alternatives and individuals. This is useful as we indeed observe different vectors of school attributes across school institutions. The "individuals" in this approach are the parents and it's rather intuitive that different parents have different school preferences.

To continue, we use a logit model instead of a probit as it is more elegant to use. In the logistic model, tails are fatter than in the probit model (this means the variance is bigger). However, assuming a standard Type 1 extreme value

[20] E.g. assigning smallest and largest extremes to the unobserved variables and probability densities.

[21] Based on lecture notes of Econometrics in 2015 (Dhaene, 2015) and Micro-econometric Models in 2016 (Kesternich, 2016).

distribution for the residuals has the useful property that the difference of two such random variables is logistically distributed. This means that the property can be directly applied to a logit model as it has direct application for the random components of the alternatives. This is not true for the probit model, where the difference between two random components is assumed to be normally distributed. The multinomial probit model is also numerically more demanding to compute than a multinomial logit model.

Potential downsides of our approach might be due to the independence of irrelevant alternative property (IIA property). It states that the comparison of choice probabilities of two alternatives does not require information on other irrelevant alternatives. However, this can be problematic as some choice options are indeed correlated. A model like nested logit assumes a hierarchical structure among choices and can deal with this problem. It allows correlation of unobservables within nests, but not between nests to relax the IIA property. Another model that can relax the IIA property is the random parameters logit. This model allows parameters to vary between individuals, let's say by assuming a normal distribution in the population.

Another rather small downside of our approach is that we can only compare coefficients of the model relative with respect to each other. However, we will present different ways of making interpretation of the model easier.

Aware of this information, we choose the conditional logit model. We will later present a model where the track choice for the student is given, as an alternative to the nested logit model.

We will now proceed with the data-section. Here we present the data on which we will apply the economic and econometric model.

5 Data

5.1 Descriptive statistics

5.1.1 The administrative dataset

This thesis will look at determinants of secondary school preferences for the district of Ghent. For what concerns its determinants, we will focus specifically on distance, school composition and school quality.

We will go through the most important aspects of the administrative data, obtained from the Flemish government. The appendix will contain details on the variables considered as less important.

The dataset itself consists of 117 different school institutions. We should note that one school can have different institutions, which means that the number of institutions is not equal to all schools. This also comprises school institutions that none of the students investigated attend. We therefor drop them from our analysis. They might be of less relevance for these students[22]. We investigate 76 different secondary school institutions where each of them is at least once chosen in our dataset.

For what concerns students, we have a dataset that includes everyone who goes to the first year of secondary school, so-called first-year enrollers. This dataset has 37,560 observations and looks for the years 2002 until 2014 at school preferences of first-year enrollers (cross-sectional data). Students from this dataset have a NIS code ranging from 44001 until 44081. This is an administrative code indicating a geographic place, here it deals with the 21 municipalities in the district of Ghent.[23]

[22] We believe they are rather exceptionally chosen for students of the districts we will investigate here. These institutions are however for the sake of completeness included in one of our datasets.

[23] Note that 41.24% of observations come from the NIS code 44021 (city of Ghent).

Distance, our first determinant of interest for school preferences, is calculated based on the sector centroid of a statistical sector in Belgium[24] of a pupil and the exact coordinate of a school. We have a total of 19,781 observations for the statistical sectors. We might assume that some sectors are not relevant for calculating a distance between their centroid and the chosen school of the pupil. As we do not want to include statistical sectors of regions that are not relevant, we used 604 observations. The only ones investigated are those capturing the statistical sector code to which at least one student belongs in our dataset of first-year enrollers.

The second determinant of interest is school composition. Variables like “receives school allowance”[25], “mother did not finish higher education” or “student does not speak Dutch at home” can be linked to the socio-economic status of the parents. If a student has at least one of the above characteristics, she/he is considered disadvantaged. The percentage of pupils having one of these characteristics is available for every school in our dataset as well as a more general measure (“percentage of disadvantaged students”). These percentages will serve as indicators of school composition and will be used as possible determinants of school preferences.

The third important independent variable of interest is school quality. Therefore, we dispose of data from an inspection analysis of schools. The data present the school years 2012-2013 until 2015-2016 and deal with both ratings of quality of primary and secondary education. We have a total of 2,747 observations. The variable of specific interest in this data is “advice”.

A school can either get a positive advice, a restricted positive advice or a negative advice on the quality of education (Vlaamse Overheid, 2017). A positive advice means that the school or institution works adequately in terms of quality and that

---

[24] This is a geographic area, which is defined depending on population densities. 19,781 such areas exist in Belgium.

[25] This does not necessarily mean people are “poorer”: this depends on the optimality of the taxation system. We will however not comment upon this here. We assume the taxation system works relatively optimal.

structures exist to further supervene on and even improve the quality[26]. The school or the institution remains officially approved (Vlaamse Overheid, 2017). A restricted positive advice signifies that the school or the institution needs to work on shortcomings within a certain time. It remains officially approved, but a follow-up analysis will take place to decide whether the school receives a positive advice (no significant shortcomings anymore) or a negative advice (see further). If a school or institution receives a negative advice from an inspection analysis, it should work on some important shortcomings. This negative advice leads to the start of a procedure on withdrawing official recognition. However, schools or institutions can come up with an improvement plan within two months. If accepted, the procedure is withdrawn for at least one and at most three school years. In the last three months of this period, a new follow-up analysis takes place by the inspectorate to judge the school or the institution and the school/institution receives a new rating (Vlaamse Overheid, 2017).

5.1.2 Distance, school composition, school quality, acquired certificates and tracks

We will end our descriptive statistics with a summary of the variables "distance", "percentage of disadvantaged students" (in school institutions) and its components and "advice". We will also briefly present some other variables we consider as interesting[27].

Variable	**Mean**	**Standard deviation**
Distance	4.63	3.74

[26] This might deal with supervision of teachers by the head of the school institution, for example.

[27] E.g. for the analysis or because they reveal relevant information about learning outcomes in secondary school.

Percentage of disadvantaged students	0.71	0.25
Percentage of students who do not speak Dutch at home	0.16	0.22
Percentage of students receiving school allowance	0.29	0.21
Percentage of students of which the mother has not finished higher education	0.65	0.28

Table 1: summary statistics about distance and school composition

First, we present information on the determinants "distance" and "school composition" such as they appear in our final dataset of first-year enrollers.[28] "School composition" consists of three variables as we noted before, but also the more general indicator of school composition "percentage of disadvantaged students" is included. We further observe that the mean travelling distance for a student is 4.63 kilometres, with 3.74 kilometres deviation appears. This is a lot of variance on this mean.

Second, looking at school composition, we see a very high mean of the percentage of disadvantaged pupils in an average secondary school (71%). Contributing most to this high number is "mother has not finished higher education". However, we see that these percentages have relatively high standard errors, which can be even

[28] These statistics might be slightly different from the picture where all secondary schools are included. This is because we only look at those schools which are observed in our choice set of first-year enrollers.

more than half of the mean (e.g. for “school allowance”). This suggests, socio-economically speaking, a relatively differentiated landscape of secondary schools. We show it in figure 1. For example, in 11.13% of all school institutions investigated, all students are disadvantaged. Approximately 51% of all school institutions have more than 75% disadvantaged students. Note that we didn’t control for number of students in figure 1. We however calculated that if a school institution has 75% or more disadvantaged students in its population, the average number of students per institution drops by approximately 55.[29]

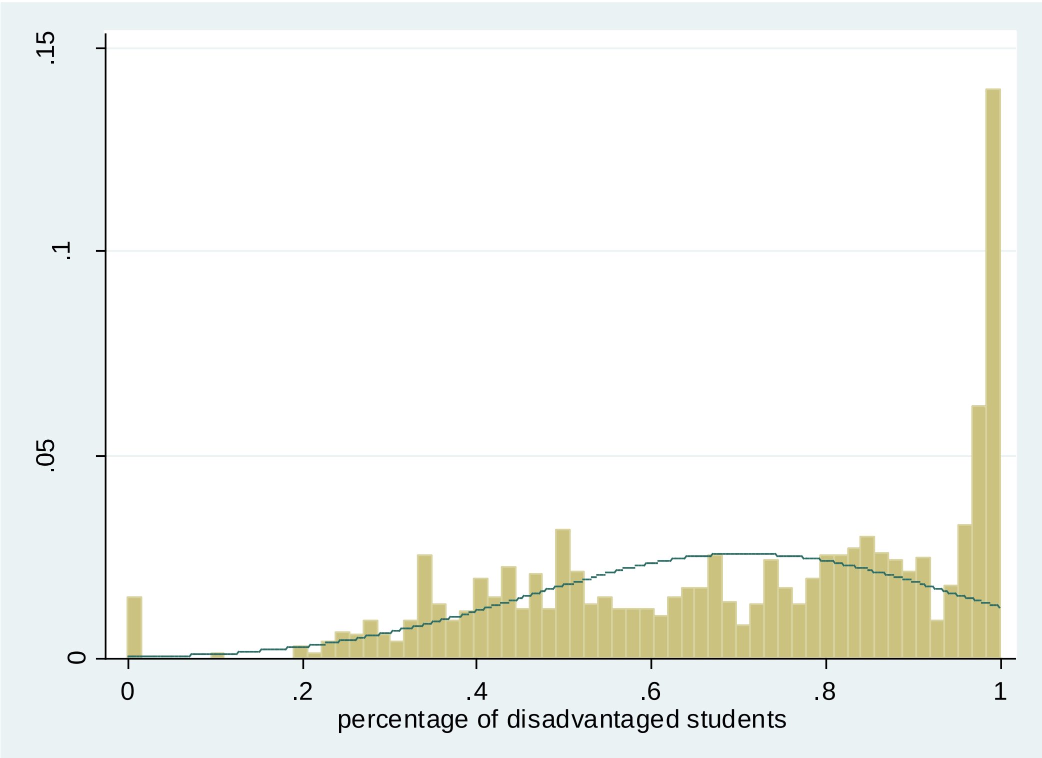

Figure 1: histogram of “percentage of disadvantaged students”

Advice	**Percent**	**Cum.**
Positive advice	34.88	34.88
Restricted positive advice	65.11	99.99
Negative advice	0.01	100

[29] Compared to the number of students in an average school institution.

Table 2: summary statistics about “advice”

As a third important determinant of school preferences, we look at “advice”, which gives a recommendation of a school based on an inspection analysis. The table above indicates that almost 2/3 of all secondary school institutions are given a “restricted positive advice” and a bit more than 1/3 of them received a “positive advice”. “Negative advice” is almost never observed in our dataset.[30] For sake of ease, we will include “positive advice”[31] in our regressions and this will be seen mostly in relation with “negative-advice-schools”. We will create three variables: one indicating a positive advice, one indicating a negative advice and one indicating “missing observations”. The latter variable is useful as we could not cover all schools and all school years with our data[32]. It will serve as a control variable in our regressions. As we have almost none “negative-advice-schools”, our indicator of “negative advice” comprises both “restricted positive advice” and “negative advice”. If a school receives one of both recommendations, it is considered to have a negative advice in our regressions.

Although this method is not standard in the literature, like sample selection models, we observed that the variables indicating missing observations and “negative advice” are insignificant, which is positive at first sight.

The educational institutions investigated with respect to “advice” included all normal secondary institutions but also part-time artistic education and part-time vocational education. Obviously, ratings for primary education were not considered for the analysis. Some other forms of education were dropped from the data as well, like “adult education” because of their irrelevance.

The fourth variable of interest is “acquired certificate”. This provides information on learning outcomes in education per school year. However, it is relatively rigid information only dealing with performances on a range of school courses. Wellbeing of students is for example not included here. Therefore, we used a dataset of students followed over several years (all the years of secondary

[30] See footnote 9.

[31] As well as a variable indicating missing observations.

[32] Only 20 out of 76 school institutions were covered by a non-missing value.

education are included and not only the first year, panel data). It ranges from 2001 until 2014. In short, 80.12% of all certificates indicate that the student passed the year, which is the “desired” outcome. Therefore, almost 20% of the data deal with observations where the student did not fully pass the year (i.e., she/he did not pass all the courses and is excluded from one or more choice tracks next year or she/he needs to redo the year). This is an undesired outcome, and it might be influenced by the SES-status of parents for example, but also by choosing a school of lower quality, ability of the student, …

The last descriptive statistic we discuss deals with tracks in secondary education. It is obtained from the same dataset as where we found “acquired certificate”. We show it as we will use it later, amongst others in our “results” section. Our results first show that “GSO” is mostly investigated. This includes the first two years of secondary education and comprises 35% of observations. This seems clear intuitively. From the third year onwards, students can really choose between different tracks like art education (KSO), vocational education (BSO), technical education (TSO) or general education (ASO). We observe ASO as being most popular with almost 31% of observations in this dataset. Thereafter TSO and BSO appear approximately equally important after ASO, dealing both with around 15% of the data, although TSO is of slightly more importance. Finally, a small majority of the data deals with art education (KSO), i.e. a bit more than 2%. We will not look at the track KSO in our second part of regressions, where we look at determinants of preferences across tracks for the sake of simplicity. In our dataset, students have considerably less choice for this form of education, which would also give less precise estimates. We will assume that determinants of preferences for KSO are most closely connected to those of ASO. We argue this statement based on the type of courses an average option within KSO offers. We believe these courses have a more theoretical focus than the courses of an average option within a TSO- or BSO-track. Therefore, KSO is more closely connected to ASO than to TSO or BSO.

5.2 Strengths, weaknesses and limitations of the data

5.2.1 Strengths

We will conclude the data-section by assessing all the variables we have. To begin, our data consist of 37,560 students, which is relatively rich and consists of all students we could possibly investigate because we use administrative data. For example, Burgess et al. (2015) investigates 42,047 pupils in their dataset, Hastings and Weinstein (2008) around 16,400. We also use information on a relatively big city where needs for secondary education are maybe more important and more value is attached to them than for smaller cities/towns[33].

To continue, we have an additional dataset for pupils throughout the years, which can be merged to the dataset of first-year enrollers[34]. This provides us additional information on tracks in secondary school and acquired certificates. This also allows us to perform a clustering of standard errors as "school where the student has followed her/his primary education" is a variable in this dataset. We will come back to this later. The idea behind this clustering is that within group i ,some correlation exists between the observations, which we want to encounter (Nichols & Schaffer, 2007). This means for our analysis that the "school where the student has followed its primary education" can influence students' perceptions for secondary school preferences and it might be correlated within this variable. Kezdi (2003) found that 50 clusters, where their size is approximately equal, is often enough to perform an accurate regression with the option. As we have more than 100 clusters, we feel ourselves confident of using it at first sight.[35]

Further, our measure of quality of education, "advice" includes several aspects of a school which contribute to educational attainment, like quality of teaching, school composition, state of the infrastructure, …[36] Our data on school composition are

---

[33] During our research, we found also confirmation for this belief. A study by De Cremer (2010) provided evidence.

[34] See appendix.

[35] We will come back to this in the "discussion" section.

[36] The first two mentioned here come also back in Burgess et al. (2015).

also rather detailed. For the measure of quality, we will nevertheless discuss some downsides/limitations as well.

Finally, we assume that parents are aware of information on distance, school composition and quality of education. We also believe the measures we propose for these determinants are also used by parents. They are clearly understandable, although perhaps not fully accessible, e.g. the precise figures on school composition. Parents might base their choices on proxies for the determinants. This might be information on quality of education or school composition from other parents in the neighbourhood.

5.2.2 Downsides

As already mentioned, “advice” as an independent variable might have downsides/limitations using it as a proxy of quality of education. Unlike school composition, consisting of many different variables, advice is just one very broad measure containing no detailed aspects, which can refer to quality of education (like state of infrastructure). Furthermore, Nelson & Ehren, (2014), a literature review about the impact of inspection on school improvement, and Penninckx (2017), specifically for Flanders, have dealt with this topic. Parents might base their preferred school partly on quality, referring to an inspection analysis. It needs to be mentioned that a positive advice on a school can come with negative side-effects, such as more stress amongst teachers and window-dressing by schools. One should be clearly aware of this when judging a school based on an inspection analysis. We should also note that information for “advice” is only available for a few school years.

5.2.3 Limitations

Finally, the data has limitations: there were sometimes duplicates of “advice” on a school in a certain year. This was perhaps because of a follow-up analysis later that year. This is encountered by dropping the best observation as the analysis in Stata is based on school preferences of a student in a certain year. “Building a reputation is a process of many years, although it can worsen rapidly after a bad advice”, is the intuition behind this. The approach thus attaches more value to “negative” observations.

As another limitation, confidentiality reasons provided for the calculations of distance a slightly inaccurate view. Centroids of statistical sectors to which students belong are used, not the exact coordinate. Nevertheless, we think this does not importantly affect our results. We made very small changes to the coordinates of our statistical sectors and observed that the variable on distance remained negative and highly significant with negligible changes to its coefficient/the coefficient on other variables[37].

The last significant limitation is that we do not dispose of a measure of wellbeing of stakeholders in our data[38]. A growing amount of literature has dealt with this subject pointing to it as an important indicator of quality of education (e.g. Holfve-Sabel, 2014; Petegem, Aelterman, Keer, & Rosseel, 2008). Negative side-effects of the performance society consist of stress, depressions and burn-outs among students (Petegem et al., 2008). Psychosocial and environmental problems could be decreased with higher wellbeing among students (Murray-Harvey & Slee, 2007). A concrete measure of this is the wellbeing inventory of secondary education (WISE, a questionnaire for students on their wellbeing) and the Questionnaire on Teacher Interaction (QTI) (Petegem et al., 2008). A possibility is for both measures to complement a measure on school performances by looking at scores on a standardized test or looking at a recommendation on a school by the measure “advice”.

[37] See appendix. This (slightly) inaccurate approach is more commonly used in the literature on school preferences. Burgess et al. (2015) use for instance “distance rank” to approximate exact travelling distances.

[38] Think of students, teachers, parents and so forth.

6 Results

This section will present the most important results from the econometric analysis. We will first start with a general analysis, where we gradually build our model by adding more variables/specifications of a certain variable. Thereafter we look at determinants of preferences across the tracks ASO, TSO and BSO.

6.1 General model

6.1.1 Some simple conditional logit regressions

<u>First regression: basic model with three determinants</u>

First, a simple conditional logit regression is performed which only includes three determinants of school preferences. The belief is that these are the most important determinants which will therefore contribute most to explaining school preferences. The independent variables are "positive advice", compared to "negative advice" and "missing", "distance" and "percentage of disadvantaged students":

simple conditional logit regression

	(1) Model 1
school choice	
positive advice	0.831***
	(0.0663)
distance	-0.400***
	(0.00209)
percentage of disadvantaged students	-1.629***
	(0.0207)
Observations	2341607
Pseudo R-squared	0.253

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

In general, we see strong and significant effects. A more qualitative education, positive advice instead of negative advice or "missing", compensates for an increase in travelling distance of approximately two kilometres[39]. The table also shows that if the percentage of disadvantaged people in a certain school goes up by 25%, this can be compensated if the school is a kilometre closer. A further remark is on the pseudo R-squared, approximately 0.253. This means 25.3% of variation in school preferences is explained. This is not a very good fit yet. If the school year was 2012-2013, we linked the result for advice to 2013. Parents thus base their school preferences on the inspection analysis of the previous year. In this case, information about quality in school year 2012-2013 is linked to school year 2013-2014. This gave us a better fit than linking it to 2012. The intuition is that parents have forward-looking expectations about the quality of education in a certain school, based on other sources as argued before. The advice that is received on a school in a certain year can be "predicted".

A last important note is on the number of observations: 2,341,607: it is the product of the number of school institutions and the number of students. It is less than 37,560 (number of students) ×76 (number of school institutions) = 2,854,560, as on some school years information on school composition, coordinates of the school, … was missing[40].

We will now compare this basic model with an approach where we use the logarithm of distance instead of distance.

<u>Second regression: first regression with logarithm of distance instead</u>

Performing the same regression again with the logarithm of distance instead of distance generates the following output ("Model 2"):

---

[39] Coefficients in a conditional logit model should always be interpreted with respect to each other. We calculated this number as follows: the coefficient on advice is approximately 0.8, the one on distance -0.4. If the school is two kilometres further away (2*(-0.4)), a positive advice compensates for this negative effect (-0.8+0.8=0)

[40] Although we felt able to construct some numbers on variables ourselves (based on other school years). An example is on the coordinates of a school institution. We however feel ourselves confident with using only 2,341,607 observations as it is still relatively much.

simple conditional logit regression

	(1) Model 1	(2) Model 2
school choice		
positive advice	0.831***	0.811***
	(0.0663)	(0.0668)
distance	-0.400***	
	(0.00209)	
percentage of disadvantaged students	-1.629***	-1.643***
	(0.0207)	(0.0206)
logarithm of distance		-1.736***
		(0.00788)
Observations	2341607	2341607
Pseudo R-squared	0.253	0.215

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

We see that the pseudo R-squared is approximately 0.215, slightly lower than in the previous regression. The variable “positive advice” is furthermore exerting slightly less effects, for example compared to “percentage of disadvantaged students”. The interpretation of “log of distance” is as follows: if a school is approximately 1% further away, this can be compensated by 1% less disadvantaged students in a certain school.

In what follows distance will be used instead of the logarithm of distance when performing regressions because of the better fit.

We will add some more variables to the basic regression model/change the basic regression model. Four output tables are presented and their results are discussed one after another. The first new model is also a relatively simple conditional logit regression, we will call it “extension 1: school composition”. Thereafter, we present some rather advanced extensions on the three basic models, in a new subsection. We built them gradually. We present all basic models and their previous regressions in one model and we do the same for the rather advanced extensions for comparison.

Different school composition variables (simple conditional logit regression)

extension 1: school composition

	(1) Model 1	(2) Model 2	(3) Model 3
school choice			
positive advice	0.831***	0.811***	0.876***
	(0.0663)	(0.0668)	(0.0664)
distance	-0.400***		-0.394***
	(0.00209)		(0.00211)
percentage of disadvantaged students	-1.629***	-1.643***	
	(0.0207)	(0.0206)	
logarithm of distance		-1.736***	
		(0.00788)	
% mother did not finish higher education			-1.032***
			(0.0256)
% not speaking Dutch at home			-0.267***
			(0.0377)
% school allowance			-1.014***
			(0.0391)
Observations	2341607	2341607	2341607
Pseudo R-squared	0.253	0.215	0.256

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

This third regression ("Model 3") looks more specifically at the different aspects of school composition. We note that "% mother did not finish higher education" and "% school allowance" exert strong negative effects on school preferences. "% not speaking Dutch at home" is negatively correlated with school preferences, but only with modest effects. The rest of the coefficients and the pseudo R-squared have only slightly changed compared to the previous model. These slight modifications are not surprising, as we did not really add "new" variables to our model. They were more detailed specifications of one variable.

6.1.2 Rather advanced extensions

The first type of variables we added were interaction variables. We looked at interactions on school composition and school quality according to the SES-status of the parents. We also included an interaction of distance with the SES-status. For the first extension, we only looked at the general indicator of "percentage of disadvantaged students". We thus added interactions with a low SES-status (the general variable) and school composition, "positive advice" and "distance" ("Model 4").

Thereafter, we divided "percentage of disadvantaged students" in its components ("Model 5"). For the other variables, we still interacted "positive advice" and "distance" like in "Model 4", with "percentage of disadvantaged students". As the interaction variable of the SES-status with "percentage of disadvantaged students" had a rather high coefficient, we now made three interactions for this variable. We divided the variable indicating a low SES-status in its components and interacted each of these components with the different components of school composition. To illustrate, we ask ourselves the question whether a student of which the mother has not finished higher education, is more likely to go to a school with a high percentage of students where the mother has no degree of higher education. We ask ourselves the same question for the indicator variables "does not speak Dutch at home" and "school allowance".

We will show all these interaction variables schematically here for the sake of simplicity:

-**Interaction quality and SES-status**: if the student has a disadvantaged background, does she/he values quality of education differently as a determinant of school preferences (compared to another determinant in the model)?

-**interaction distance and SES-status**: if a student has a disadvantaged background, does she/he values distance differently as a determinant of school preferences (compared to another determinant)?

-**interaction school composition and SES-status**: if a student has a disadvantaged background, does she/he value "percentage of disadvantaged students" differently (compared to another determinant)?

-**interaction 'education'**: if the mother of a student has no degree of higher education, is "percentage of students of which the mother has not finished higher education" valued differently (compared to another determinant)?

-**interaction 'language'**: if a student does not speak Dutch at home, does she/he value "percentage of students that does not speak Dutch at home" differently (compared to another determinant)?

-**interaction 'school allowance'**: if a student receives school allowance does she/he value "percentage of students that receive school allowance" differently (compared to another determinant)?

The last extension will conclude by presenting the second extension and adding the squared term of "distance". It is explicitly stated here that no interaction with the SES-status of the student is made. The effect of the squared term in the model will be rather small. Therefore, it is less important to include an interaction.

<u>First extension: interaction effects with the general variable "% of disadvantaged students"</u>

first extension

	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
school choice				
positive advice	0.831***	0.811***	0.876***	0.998***
	(0.0663)	(0.0668)	(0.0664)	(0.0961)
distance	-0.400***		-0.394***	-0.390***
	(0.00209)		(0.00211)	(0.00313)
percentage of disadvantaged students	-1.629***	-1.643***		-3.223***
	(0.0207)	(0.0206)		(0.0339)
logarithm of distance		-1.736***		
		(0.00788)		
% mother did not finish higher education			-1.032***	
			(0.0256)	
% not speaking Dutch at home			-0.267***	
			(0.0377)	
% school allowance			-1.014***	
			(0.0391)	
interaction quality & SES-status				-0.277*
				(0.133)
interaction distance & SES-status				-0.0256***
				(0.00426)
interaction school composition & SES-status				2.686***
				(0.0438)
Observations	2341607	2341607	2341607	2341607
Pseudo R-squared	0.253	0.215	0.256	0.266

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

To begin with, a strong interaction of the (low) SES-status of a student with the percentage of disadvantaged students in the school composition is observed. Although there is a strong effect, even disadvantaged students don't seem to value school composition positively. The interaction of low SES-status with quality is also relatively pronounced and significant at the 5 % significance level. We obtain a better fit now of 0.266. The coefficient on "positive advice" has slightly increased and the negative correlation of "percentage of disadvantaged students" is more pronounced now than in "Model 1" and "Model 2". They seem to be upwardly biased in the first two models by not including these interaction effects on school composition. This bias is true as well as for the coefficients on “positive advice” and “distance”. They also changed by adding their respective interaction effects. We will now divide “percentage of disadvantaged students” in its components.

Second extension: school composition variables, interaction effects with components of school composition second extension

	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4	(5) Model 5
school choice					
positive advice	0.831***	0.811***	0.876***	0.998***	1.136***
	(0.0663)	(0.0668)	(0.0664)	(0.0961)	(0.0965)
distance	-0.400***		-0.394***	-0.390***	-0.380***
	(0.00209)		(0.00211)	(0.00313)	(0.00312)
percentage of disadvantaged students	-1.629***	-1.643***		-3.223***	
	(0.0207)	(0.0206)		(0.0339)	
logarithm of distance		-1.736***			
		(0.00788)			
% mother did not finish higher education			-1.032***		-2.385***
			(0.0256)		(0.0349)
% not speaking Dutch at home			-0.267***		-1.368***
			(0.0377)		(0.0488)
% school allowance			-1.014***		-1.567***
			(0.0391)		(0.0460)
interaction quality & SES-status				-0.277*	-0.337*
				(0.133)	(0.133)
interaction distance & SES-status				-0.0256***	-0.0204***
				(0.00426)	(0.00420)
interaction school composition & SES-status				2.686***	
				(0.0438)	
interaction 'education of mother'					2.594***
					(0.0437)
interaction 'language'					2.963***
					(0.0709)
interaction 'school allowance'					1.802***
					(0.0667)
Observations	2341607	2341607	2341607	2341607	2341607
Pseudo R-squared	0.253	0.215	0.256	0.266	0.282

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

We observe in the table[41] that most of the interaction variables do have strong and significant effects. There seems to be a preference for a school in terms of school composition according to SES-status of parents of the pupil. The respective school composition variables now all become positively correlated with school preferences if their respective interaction effect plays a role as well. This interaction is strongest for the language spoken at home and the educational level of the mother. Important to note is that we obtain a slightly better pseudo R-squared (of 0.282).

We will now come to our final rather advanced extension. We already explained the logic behind including the variable “distance squared” in our hypotheses.

<u>Third extension: the second extension with distance squared included</u>

[41] We compressed the table due to the extra variables.

final extension

	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4	(5) Model 5	(6) Model 6
school choice						
positive advice	0.831***	0.811***	0.876***	0.998***	1.136***	1.156***
	(0.0663)	(0.0668)	(0.0664)	(0.0961)	(0.0965)	(0.0966)
distance	-0.400***		-0.394***	-0.390***	-0.380***	-0.497***
	(0.00209)		(0.00211)	(0.00313)	(0.00312)	(0.00533)
percentage of disadvantaged students	-1.629***	-1.643***		-3.223***		
	(0.0207)	(0.0206)		(0.0339)		
logarithm of distance		-1.736***				
		(0.00788)				
% mother did not finish higher education			-1.032***		-2.385***	-2.412***
			(0.0256)		(0.0349)	(0.0349)
% not speaking Dutch at home			-0.267***		-1.368***	-1.362***
			(0.0377)		(0.0488)	(0.0488)
% school allowance			-1.014***		-1.567***	-1.554***
			(0.0391)		(0.0460)	(0.0461)
interaction quality & SES-status				-0.277*	-0.337*	-0.341*
				(0.133)	(0.133)	(0.133)
interaction distance & SES-status				-0.0256***	-0.0204***	-0.0155***
				(0.00426)	(0.00420)	(0.00382)
interaction school composition & SES-status				2.686***		
				(0.0438)		
interaction 'education of mother'					2.594***	2.605***
					(0.0437)	(0.0438)
interaction 'language'					2.963***	2.935***
					(0.0709)	(0.0710)
interaction 'school allowance'					1.802***	1.805***
					(0.0667)	(0.0668)
distance squared						0.00705***
						(0.000253)
Observations	2341607	2341607	2341607	2341607	2341607	2341607
Pseudo R-squared	0.253	0.215	0.256	0.266	0.282	0.284

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

"Model 6" shows us what we predicted: the squared term of distance is slightly positively correlated with school preferences and highly significant. The coefficients on the other independent variables only slightly changed as well as our pseudo R-squared, which is now 0.284.

6.1.3 Final remarks on the results

Bigger remarks

Compared to our “Model 1”, we mention that the coefficient on “positive advice” has increased by 39.11%, e.g. by adding the interaction term on quality. The coefficient of distance *an sich* has also increased, by adding the squared term perhaps, and the coefficients for the variables of school composition have changed as well. However, we should interpret the coefficients relative with respect to each other. These effects might be partly cancelled out by changes in coefficients of other variables. Nevertheless, the variables do have significant correlations with school preferences.

Most interaction effects we added do have strong and significant effects. The most pronounced interactions exist for “language” and “education of mother”, although “school allowance” exerts strong effects as well. All these three interaction effects change the school composition variables in positive determinants of secondary school preferences if the student is disadvantaged. The coefficient of SES-status with quality is relatively large but only significant at the 5% significance level. We also observed slightly significant negative correlation between distance and SES-status.

Our interactions and the squared term of distance added altogether 3 percentage points to the fit of our model. We were clearly almost at the limit of explanatory power with the three variables of the basic model.

To continue, our last regression confirmed our beliefs about the squared term of distance, although it added only 0.2 percentage points to the pseudo R-squared of the model.

We will provide an example of how to interpret the importance of each determinant of preferences for students with a different SES-background. Looking at “Model 4”, we calculated that a disadvantaged student has a coefficient of 0.721 for “positive advice” and -0.756 for “distance” (the normal effect and the interaction effects combined). For a student who is not disadvantaged, the respective coefficients are 0.998 and -0.350. We observe the coefficient of “positive advice” to be approximately 1.92 times the coefficient of “distance” in absolute value for the disadvantaged student and approximately 2.85 times the coefficient of “distance” for the student who is not disadvantaged. This suggest that quality of education is valued higher with respect to distance for a student who is not disadvantaged. We can make other comparisons, for example with “percentage of disadvantaged students”, as well. Doing this, we could not conclude the higher importance of quality in general for students who are not disadvantaged.

Smaller remarks

We should note that for the coefficients on school composition it is not easy to draw conclusions. A school that is more often chosen might have lower percentages on the components of our variables of school composition or the general term “percentage of disadvantaged students”. This would then indicate reverse causality. The negative coefficients on these variables would then mean for example that smaller schools have more often disadvantaged students. We also noted that is indeed the case: in a former section this finding was presented. Smaller schools might be less able to cope with the “problems” concerning school composition. However, the variables of school composition might indicate preferences as well. To illustrate, it is possible that students with a disadvantaged background prefer to go to smaller school institutions, although both smaller and larger school institutions adopt a similar method to obtain a mixed school composition. We would then assume of course that each of these methods works equally well.

Looking at the final regression of this part, it needs to be mentioned again that constraints on “advice” exist. Even for the few school years where we could use the variable, not all schools were covered.

Although we mentioned the possible clustering of standard errors, we did not make use of the option. It does more harm than good to the regressions, as there are a number of cases where this variable has a missing value for a certain student. This yields a lower number of cases to investigate which positively influences standard errors on coefficients of independent variables. For this reason, we will also not make use of the option in the second part of our regressions, the final regression across tracks. We provide in our appendix the details for this option for the final regression.

6.2 Determinants of preferences across different tracks

6.2.1 Our approach

Before looking at the final regression of the previous part across tracks, we will explain what we have done to distinguish choice sets across ASO, TSO and BSO.

To begin, the assumption is made that tracks for students are chosen in the first year, although entering takes place in the third year of secondary education.

As we could not easily deduce offered tracks for different school institutions, their deduction is based on the observed track choice for a student, here in the third year of secondary education. Doing this, tracks on institutions are summarized throughout the years, assuming they won't change in cause of time.

A number of schools, called "middle schools", offered only the first two years of secondary education (GSO, 27 institutions). Sometimes these middle schools were artificially divided in different institutions where one or more of them offered specific tracks like ASO, TSO or BSO, besides one offering only GSO. If this were the case, the middle school was considered to offer this/these specific track/specific tracks. Therefore, a GSO-school institution could be assumed an "ASO-institution" or an institution offering BSO and TSO for example.

On other middle schools, this artificial distribution was not applicable, i.e. institutions for the respective school where an automatic flow into the former could be found. We approached this by looking at track preferences for students who had followed their first year in these respective schools. If from a certain middle school institution most students went to ASO in later years, the school institution was considered an ASO-institution. We used this approach in cases where more than 70% of students were following ASO in later years. If TSO and BSO combined were responsible for more than 70% of later choices, we considered the institution to offer these two tracks. We looked here at the combination because most institutions in our dataset offer BSO if they offer TSO and the other way around. If we could not find a specific trend with respect to the track, we assumed the school to offer all tracks[42]. The intuition here was that parents would choose the institution with the intention to follow a specific track in an institution nearby.

The information on students for the third year of secondary education could be found in a dataset following students over several years. To control for ability, we selected students who achieved both in the first and the second year of secondary education good learning outcomes. This was possible by looking at acquired certificates for the first and the second year: desired outcomes were the only ones we included[43]. From the third year onwards, we considered all students still in the dataset after using this approach. The already mentioned study by Pinxten et al. (2012) noted that earlier achievement has high predicting power by choosing an option choice in secondary education.[44]

We merged the data on students in their third year to the dataset of investigation. School institutions were divided according to the different tracks they offer. If TSO and BSO were offered for example, the institution was twice presented in the new dataset. A student would now have a match with the institution, if there would be a match with the track as well if there were different tracks offered. In case no information on the chosen track was obtained from the used dataset for the student and the school institution only offered one track, we assumed there to be a match as well, so that the institution was indeed chosen for the specific track offered.

This approach yielded a total of 31,765 observed school preferences from the original 37,560. Unfortunately, we were less lucky with this approach concerning the variable "positive advice'. As already mentioned, a restricted amount of observations is available: only 20 school institutions are covered. For the tracks TSO and BSO, students' choices only indicated missing observations on the variable "advice"[45]. 3,587 students choose the track BSO while 3,934 students choose the track TSO with a missing value on the variable. Apparently, we have

---

[42] Besides artistic education, which we will not (directly) cover.

[43] See earlier.

[44] Note that this is in fact a choice within a track, like a choice within general secondary education. An example is a languages-oriented option (ASO).

[45] Although there were TSO and BSO schools where the variable was not missing, of course.

too much missing observations for these track choices amongst our students. We can however still try to compare ASO with the whole sample, i.e. all tracks included. Later is explained why. We will also make the comparison for the other determinants of school preferences between the tracks, which are distance and school composition. To justify this approach, the appendix will present a regression table where we dropped all the middle schools and made no approximations, yielding rather similar results.

First, a table is presented to give an overview on the chosen tracks for the students in the newly created dataset. This table is only an approximation of the real situation, of course.

Track choice	**Percent**	**Cum.**
General secondary education (ASO)	64.59	64.59
Technical secondary education (TSO)	18.99	83.58
Vocational secondary education (BSO)	14.60	98.18
Artistic education (KSO)	1.82	100.00

Table 3: summary statistics about track choice

We observe our approach to be quite precise, relatively speaking, with respect to most tracks, although we have a high representation of the track ASO. 17,720 students attend in their first year of secondary education the GSO-track. Most of these students seem to continue in an ASO-track later.

We will now compare determinants of school preferences across tracks for distance and school composition. For the sake of ease, abbreviations for the interaction effects are still used, listed again before proceeding this analysis[46]:

[46] If we used them in the model of course.

-**interaction distance and SES-status**: if a student has a disadvantaged background, does she/he values distance differently as a determinant of school preferences (compared to another determinant in the model)

-**interaction 'education'**: if the mother of a student has no degree of higher education, is "percentage of students of students of which the mother has not finished higher education" valued differently (compared to another determinant)?

-**interaction 'language'**: if a student does not speak Dutch at home, does she/he value "percentage of students that does not speak Dutch at home" differently (compared to another determinant)?

-**interaction 'school allowance'**: if a student receives school allowance does she/he value "percentage of students that receive school allowance" differently (compared to another determinant)?

6.2.2 Determinants of preferences across tracks

Results and interpretation[47]

To test whether the differences in the coefficients are significant, we performed a regression where we first included the dummies "TSO" and "BSO". Thereafter, we interacted each of these dummies with all the variables of the previous regression and presented it next to all the "normal" variables of the model, i.e. the variables without the interaction. The idea is to investigate whether the coefficients of the interaction effects for the TSO- or BSO-track are statistically significant different from the ASO-track. If so, the specific determinants for school preferences with respect to each other might be significantly different across the tracks.

We already mention that the tables become much larger now. Therefore, we recommend using magnifying glasses to consult the precise coefficients for the variables. The first model presented uses the approximation method and is called "Model 7". We will also show the results of a model without middle schools, namely "Model 8". The regression table for both models is provided in the appendix. We

[47] We tested for the most remarkable differences between the tracks (which is a non-exhaustive approach).

will start by discussing “Model 7” and its relevance for our analysis. Thereafter, we will point to the most remarkable differences in the determinants across tracks.

Does our approach work?

The approach of “Model 7” is compared with the situation where no middle school institutions are included. It yields relatively similar results: most statistically significant differences remained the same.

However, shortcomings exist. One possible shortcoming is that the analysis does not control for a “range” of parental preferences. It might be that parents are indifferent whether their student chooses ASO or TSO in the third year, but strongly dislike BSO. In that case, a certain institution could have been chosen because of preferences for many tracks. Now is assumed that each student prefers just one specific track and “dislikes” all the other tracks.

To interpret the results, the approach without the middle school institutions is examined not to potentially bias any conclusions. After this part, we show an example of how we interpreted the differences in coefficients with respect to each other. We will come back later to “Model 7”, the approximation method, when the ASO-track is compared with the whole sample.

Interpretation of results

Preferred track

First, the results indicate strong significantly negative correlations of TSO and BSO on school preferences. We noted that the coefficient for BSO is even more negative than for TSO. According to our calculations, ASO is offered approximately 38% of all times, TSO approximately 30% and BSO approximately 32%. However, most students follow general secondary education. A possible explanation lies in the infrastructure needed to offer the TSO/BSO-track. An average TSO/BSO-school institution might have less students than an average ASO-school institution, although specific, maybe rather expensive, infrastructure is needed to offer the track. The “average” infrastructure needed per student can be much higher for the TSO/BSO-track, giving a lower number of students per TSO/BSO-track.[48]

---

[48] The accuracy of this assumption depends of course on the correctness of our model.

Distance

Second, we clearly noted that distance has statistically significant different coefficients in the ASO-track compared to the TSO/BSO-track. Another potential way to look at the importance of distance across tracks is to summarize this variable for every track. We have strong indications that distance as a more important determinant for the ASO-track. We show an example of how to extract this from our model.

Suppose our student is not disadvantaged. We observe distance to have a coefficient of -0.54 in the ASO-track, “% school allowance” a coefficient of -0.987. If we look at the TSO-track these coefficients become -0.421 and -2.358 respectively, taking into consideration the interaction effects. If in the ASO-track the institution is approximately one kilometre further away, this can be compensated by having approximately 50% less students receiving school allowance. In the TSO-track, an increase in the travelling distance of one kilometre can be compensated by approximately 20% less students receiving school allowance. This means distance is valued more with respect to “% school allowance” in the ASO-track compared to the TSO-track.

One explanation for this result could lie in the role of supply and demand. As most pupils tend to follow general secondary education, this track is in higher supply than a TSO/BSO-track. Parents might know this and become more selective in their school preferences with respect to distance. However, we know this reasoning is rather speculative and should be confirmed by other research.

School composition

For school composition indications exist that, according to the specific SES-status of parents, preferences differ across tracks. For example, a student of which the mother has no degree of higher education seems to have a stronger preference to attend a school with a high percentage of similar students in the TSO/BSO-track compared to the ASO-track.

For the sake of simplicity, we summarized the average percentage of disadvantaged students for every track. The results indicate that this average percentage is considerably lower in general secondary education. This should be weighted for the number of students in each institution. We thus only have rather strong indications that a disadvantaged school composition is more negatively perceived in general secondary education.

This section concludes by comparing the ASO-track, which also includes a variable on school quality, with the whole sample, indirectly including all tracks. We look only at “matches” for students with the institution, so we don’t have to deal with missing observations from chosen tracks. This yields 37,560 observations instead of 31,765 or more than 5,500 additional observations.

6.2.3 The ASO-track compared to the whole sample

The tables are presented beneath each other as they both come from another dataset. We suggest the reader to put them next to each other for comparison.

final extension: ASO

	(1) Model 9
school choice	
positive advice	1.198***
	(0.0983)
distance	-0.523***
	(0.00578)
% mother did not finish higher education	-1.731***
	(0.0450)
% not speaking Dutch at home	0.0184
	(0.0701)
% school allowance	-1.450***
	(0.0577)
interaction quality & SES-status	0.117
	(0.152)
interaction distance & SES-status	-0.0289***
	(0.00477)
interaction 'education of mother'	1.410***
	(0.0654)
interaction 'language'	3.786***
	(0.149)
interaction 'school allowance'	1.837***
	(0.0987)
distance squared	0.00921***
	(0.000242)
Observations	894477
Pseudo R-squared	0.262

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

the final extension

	(1)
	Model 6
school choice	
positive advice	1.156***
	(0.0966)
distance	-0.497***
	(0.00533)
% mother did not finish higher education	-2.412***
	(0.0349)
% not speaking Dutch at home	-1.362***
	(0.0488)
% school allowance	-1.554***
	(0.0461)
interaction quality & SES-status	-0.341*
	(0.133)
interaction distance & SES-status	-0.0155***
	(0.00382)
interaction 'education of mother'	2.605***
	(0.0438)
interaction 'language'	2.935***
	(0.0710)
interaction 'school allowance'	1.805***
	(0.0668)
distance squared	0.00705***
	(0.000253)
Observations	2341607
Pseudo R-squared	0.284

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

The first table above ("Model 9") presents the final regression for the ASO-track, the second table above is summed over all tracks ("Model 6", ASO, TSO, BSO). As the importance of distance and school composition across tracks is discussed, the focus is on school quality and its interaction effect. The idea is to look whether the quality aspect is more important for the ASO-track than for the whole sample. This would mean therefore that quality has higher importance in general secondary education.

A statistical test shows that the coefficient on quality is significantly more pronounced in ASO. However, it is only significant at the 15% significance level. Further, no statistically significant effect of the interaction with a low SES-status for the ASO-track is found, while it is significantly negative for the whole sample. Parents of a lower SES-status seem to attach more importance to school quality in the ASO-track. Indeed, the interaction effect of school quality according to the SES-status of the parents is significant and more negative for the whole sample compared to the ASO-track. This is at a very low significance level, because the p-value is 11.22.

Indications thus appear that school quality is perceived as more important in ASO than on average, although still comparisons of "positive advice" with respect to other determinants are needed in both models. One can do this according to the example presented above.

As we had important limitations, we should not draw too much conclusions here at first sight. We however indicated that our variables on school composition and distance might be differently valued in the ASO-track, i.e., more negatively. This offers also additional indications that school quality should indeed be perceived as more important in general secondary education. Nevertheless, the results for quality are perhaps even more suggestive because of the data limitations.

6.2.4 Remarks on the generalization of the model

To conclude, it is important to repeat that the different approaches have limitations. We could not extract chosen tracks for all students in the table where we compared the different tracks. The assumptions we presented to approximate a real model without middle school institutions are furthermore at least partly inaccurate.

Therefore, certain findings can only be suspected to be more common in certain tracks. The results can try to indicate trends, but more research needs to complement our findings.

Section seven will now summarize the main results and introduce a discussion on these findings.

7 Discussion

For the hypothesis on school quality significant positive correlation is found. This is in favour of our first hypothesis (1a) of section 3. We noted that school quality is likely to be the most important positive determinant of preferences. We also tested whether parents with a lower SES-status attach relatively less importance to quality of education than parents with a higher SES-status, which was hypothesis 1b. No evidence was found for this. We formally tested this by including an interaction term of quality with the SES-status of the parents.[49] However, it needs to be mentioned that results could have occurred due to data limitations as well. It is plausible that some of these results are at least partly due to coincidence. We nevertheless believe that recommendations of schools by the inspectorate provide a relatively inclusive measure for school quality. This must be considered as an important value-added of this thesis.

Specific preferences for school composition according to the SES-status of the parents exist. Therefore, hypothesis 2a holds. This is true for all different determinants of school compositions. All coefficients of the school composition variables changed of sign by including the respective interaction effects. For the sake of ease, we didn't fully conclude on preferences for a certain track in our thesis according to the SES-status of the parents, which was hypothesis 2b. However, we found that the average percentage of disadvantaged students in a school institution is considerably lower in the ASO-track[50]. This should be weighted for the number of students in each institution of course[51]. Thus, only indications that this hypothesis is confirmed exist, although rather strong.

For distance, the predicted effects were found, both for hypothesis 3a and hypothesis 3b. Distance is a negative determinant of school preferences, but the marginal effect of distance is positive. This means that the negative effect becomes

[49] Remember that we had to look at the importance of quality for a disadvantaged student with respect to other determinants like distance.

[50] By summarizing this variable for each track.

[51] In the data section, we observed the number of students to drop by approximately 55 if more than 75% percent of all students are disadvantaged.

less pronounced if the school is further away. This thesis further remarks that the interaction between distance and SES-status exerted significant negative effects. Transport costs and practical factors might play a role here. Evidence was found that for a disadvantaged student, distance is higher valued with respect to “percentage of disadvantaged students” and “positive advice” than for a student who is not disadvantaged.

We also investigated determinants of secondary school preferences across different tracks. Unfortunately, a number of difficulties arose when looking at the importance of quality. Therefore, we looked at the determinants of school composition and distance across tracks.

Our results seem to indicate that the preferences for school composition indeed differ across tracks, as well as distance. For distance, a logical explanation could lie in the role of supply and demand. As most pupils tend to follow general secondary education, this track is in higher supply than a TSO/BSO-track. Parents might know this and become more selective in their school preferences with respect to distance. This is maybe also true with respect to their preferences for school composition. Especially for distance, we have found strong indications that it is perceived as a more important determinant of preferences in the ASO-track. Preferences for school composition seemed to be more complex to compare. However, we believe that in general a student with a lower SES-status will have a higher preference for a TSO- or BSO-track than for an ASO-track. This is in line with what we discussed earlier.

We also compared the results from the smaller sample consisting of the ASO track participants to the results from the whole sample. We suspect that school quality might be perceived as more important in the ASO-track than on average, depending on the significance level we adopt. Nevertheless, only the first steps for this analysis were provided. Because of substantial approximations, these results should be regarded as suggestive.

There are certain limitations of this study that one should keep in mind. We should note that Ghent is probably not a representative case for Flanders in some respect like school composition. It has a considerably higher percentage of disadvantaged students. It was also not possible to perform a clustering of standard errors in our regressions without substantial loss of information[52].

This thesis seems to support findings in favour of school quality but also in favour of school composition and distance. Like Tan (1998) Thys & Van Houtte (2016) and Wouters & Groenez (2014), we observed that the results are in favour of school composition as a slightly positive determinant of school preferences for disadvantaged students. This determinant turns strongly negative when the student is not disadvantaged. As we already noted earlier, we should be careful on assigning this result only to preferences. A bigger school (more students) might be more able to cope with problems like a non-mixed school composition. Another possibility is that students with a disadvantaged background indeed prefer smaller school institutions. This would then rather indicate preferences.[53] The rather extensively elaborated literature around our subject pointed also to the importance of school quality and distance as determinants. “School quality is highly positively valued, distance negatively”, appears as an often-cited result. A remarkable difference of our study is however that the perception of school quality does not depend on the SES-status of the parents. However, it is hard to say whether these results occur because of data limitations or because of real preferences Further, the higher importance of distance to schools for parents with a low SES-status must be confirmed by more literature in the future.

One of the more important reasons to look at determinants of secondary school preferences was to combat segregation. We found that in Ghent there is still a differentiated landscape of schools according to the SES-status of the parents[54]. Wouters & Groenez (2015) already pointed to the importance of a mixed school composition as we observed in the introduction. The existing measures might be insufficient to deal with this situation (e.g. geographic criteria, local consultation platforms, middle school institutions). The stigma of TSO and BSO might also be improved by the government, for example by showing employment opportunities.

[52] Due to data limitations.

[53] For example, smaller schools may use a more individual approach in relation with their students.

[54] Note that we do not explicitly state that this is because of preferences. It is however a finding we presented in our data section.

The ASO-track is by far the most popular track, although this choice must be made in accordance with the ability of the student, as we noted earlier. Hastings & Weinstein (2008) furthermore suggested for public good markets the provision and framing of information, thereby increasing efficiency in the school choice process. It is possible that parents are not fully informed about the quality of education in schools (and the possible advantages this has on educational outcomes). An important remark is of course that segregation has indeed some advantages, but we consider the disadvantages to be more pronounced. A possible solution for schools to combat segregation could lie in the recruitment policy of personnel (teachers) as well. Good teachers might prefer to work in institutions where the population is not too disadvantaged. If so, the government can use its power to randomly assign teachers over different school institutions.

More econometric research on determinants of school preferences in Flanders should be conducted. Effects of determinants might differ over regions, which is to be included in future studies, e.g. countryside versus city. Cities are considered to have more demand and supply in education than the countryside. Furthermore, on determinants of primary education more literature could be added for Flanders. This is especially true for econometric research.

8 Bibliography

Ball, S. J., Bowe, R., & Gewirtz, S. (2006). School choice, social class and distinction: the realization of social advantage in education. *Journal of Education Policy*. https://doi.org/10.1080/0268093960110105

Black, S. E. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *The Quarterly Journal of Economics*, *114*(2), 577–599. https://doi.org/10.1162/003355399556070

Burgess, S., Greaves, E., Vignoles, A., & Wilson, D. (2011). Parental choice of primary school in England: what types of school do different types of family really have available to them? *Policy Studies*. Retrieved from http://www.tandfonline.com.kuleuven.ezproxy.kuleuven.be/doi/abs/10.1080/01442872.2011.601215

Burgess, S., Greaves, E., Vignoles, A., & Wilson, D. (2015). What Parents Want: School Preferences and School Choice. *The Economic Journal*, *125*(587), 1262–1289. https://doi.org/10.1111/ecoj.12153

cijfermateriaal leerlingenkenmerken. (n.d.). Retrieved July 17, 2017, from /cijfermateriaal-leerlingenkenmerken

De Cremer, J. (2010). Een analyse van de Vlaamse maatschappelijke ruimte in het licht van de geografie van het secundair onderwijs.

De Morgen. (n.d.). 4 op 10 leerlingen in middelbaar onderwijs zijn kansarm | Binnenland | De Morgen. Retrieved July 13, 2017, from https://www.demorgen.be/binnenland/4-op-10-leerlingen-in-middelbaar-onderwijs-zijn-kansarm-b5986b68/

Deming, D. J., Hastings, J. S., Kane, T. J., & Staiger, D. O. (2014). School Choice, School Quality, and Postsecondary Attainment. *American Economic Review*, *104*(3), 991–1013. https://doi.org/10.1257/aer.104.3.991

Dhaene, G. (2015). Econometrie.

Doorlichten van instellingen. (n.d.). Retrieved July 31, 2017, from /doorlichten-van-instellingen

Groenez, S., Van den Brande, I., & Nicaise, I. (2004). Sociale ongelijkheid in het Vlaamse onderwijs becijferd. *Over. Werk. Tijdschrift van Het Steunpunt WAV*, *14*, 112–116.

Hastings, J. S., & Weinstein, J. M. (2008). Information, School Choice, and Academic Achievement: Evidence from Two Experiments. *The Quarterly Journal of Economics*, *123*(4), 1373–1414. https://doi.org/10.1162/qjec.2008.123.4.1373

Holfve-Sabel, M.-A. (2014). Learning, Interaction and Relationships as Components of Student Well-being: Differences Between Classes from Student and Teacher Perspective. *Social Indicators Research*, *119*(3), 1535–1555. https://doi.org/10.1007/s11205-013-0557-7

Kesternich, I. (2016). Micro-econometric Models.

Kezdi, G. (2003). Robust standard error estimation in fixed-effects panel models. Retrieved from https://papers.ssrn.com/soL3/papers.cfm?abstract_id=596988

Lankford, H., & Wyckoff, J. (1992). Primary and secondary school choice among public and religious alternatives. *Economics of Education Review*, *11*(4), 317–337. https://doi.org/10.1016/0272-7757(92)90040-A

Lavy, V. (2015). *The Long-Term Consequences of Free School Choice* (Working Paper No. 20843). National Bureau of Economic Research. https://doi.org/10.3386/w20843

lokale overlegplatforms in Vlaanderen. (n.d.). Retrieved July 31, 2017, from /lop

McFadden, D., & McFadden, D. (1977). *Modelling the Choice of Residential Location* (text No. 477). Cowles Foundation for Research in Economics, Yale University. Retrieved from http://EconPapers.repec.org/RePEc:cwl:cwldpp:477

Murray-Harvey, R., & Slee, P. T. (2007). Supportive and Stressful Relationships With Teachers, Peers and Family and Their Influence on Students' Social/Emotional and Academic Experience of School. *Journal of Psychologists and Counsellors in Schools*, *17*(2), 126–147. https://doi.org/10.1375/ajgc.17.2.126

Musset, P. (2012). School choice and equity: current policies in OECD countries and a literature review. Retrieved from http://repositorio.minedu.gob.pe/handle/123456789/2510

Nelson, R., & Ehren, M. C. M. (2014). Review and synthesis of evidence on the (mechanisms of) impact of school inspections. Retrieved from https://research.utwente.nl/en/publications/review-and-synthesis-of-evidence-on-the-mechanisms-of-impact-of-s

Nichols, A., & Schaffer, M. (2007). Clustered errors in Stata. In *United Kingdom Stata Users' Group Meeting*. Retrieved from http://fmwww.bc.edu/repec/usug2007/Stata%20UGM%202007/crse.pdf

Officieel en vrij onderwijs, onderwijsnetten en - koepels. (n.d.). Retrieved July 17, 2017, from /nl/officieel-en-vrij-onderwijs-onderwijsnetten-en-koepels

Opdenakker, M.-C., & Van Damme, J. (2007). Do School Context, Student Composition and School Leadership Affect School Practice and Outcomes in Secondary Education? *British Educational Research Journal*, *33*(2), 179–206.

Penninckx, M. (2017). Effects and side effects of school inspections: A general framework. *Studies in Educational Evaluation*, *52*, 1–11. https://doi.org/10.1016/j.stueduc.2016.06.006

Petegem, K. V., Aelterman, A., Keer, H. V., & Rosseel, Y. (2008). The influence of student characteristics and interpersonal teacher behaviour in the classroom on student's wellbeing. *Social Indicators Research*, *85*(2), 279–291. https://doi.org/10.1007/s11205-007-9093-7

Pinxten, M., Fraine, B. D., Noortgate, W. V. D., Damme, J. V., & Anumendem, D. (2012). Educational choice in secondary school in Flanders: the relative impact of occupational interests on option choice. *Educational Research and Evaluation*. Retrieved from http://www.tandfonline.com.kuleuven.ezproxy.kuleuven.be/doi/abs/10.1080/13803611.2012.702991

Rothstein, J. M. (2006). Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions. *American Economic Review*, *96*(4), 1333–1350. https://doi.org/10.1257/aer.96.4.1333

Schneider, M., & Buckley, J. (2002). What Do Parents Want from Schools? Evidence from the Internet. *Educational Evaluation and Policy Analysis*, *24*(2), 133–144.

Sierens, S. (2006). Immigratiesamenleving, onderwijs en overheid in vlaanderen: een gespannen driehoeksverhouding. In *Onderwijs onderweg in de immigratiesamenleving* (pp. 9–32). Academia Press. Retrieved from http://hdl.handle.net/1854/LU-4305256

Tan, B. (1998). *Blijvende sociale ongelijkheden in het Vlaamse onderwijs*. UFSIA. Retrieved from http://www.centrumvoorsociaalbeleid.be/sites/default/files/D%201998%206104%2002.pdf

Thys, S., & Van Houtte, M. (2016). Ethnic composition of the primary school and educational choice: Does the culture of teacher expectations matter? *Teaching and Teacher Education*, *59*, 383–391. https://doi.org/10.1016/j.tate.2016.06.011

Tooley, J. (1997). On School Choice and Social Class: A Response to Ball, Bowe and Gewirtz. *British Journal of Sociology of Education*, *18*(2), 217–230.

van Kempen, R., & şule Özüekren, A. (1998). Ethnic Segregation in Cities: New Forms and Explanations in a Dynamic World. *Urban Studies*, *35*(10), 1631–1656. https://doi.org/10.1080/0042098984088

Vanderhoeven, J. L. (2004). *Positief omgaan met verschillen in de leeromgeving. een visie op differentiatie en gelijke kansen in authentieke middenscholen*. Garant.

Wondratschek, V., Edmark, K., & Frölich, M. (2013). The Short- and Long-Term Effects of School Choice on Student Outcomes - Evidence from a School Choice Reform in Sweden. https://doi.org/10.2139/ssrn.2360549

Wouters, T., & Groenez, S. (2014). De evolutie van schoolse segregatie in het Nederlandstalige onderwijs in België. *Tijdschrift Voor Sociologie*, *35*(3), 185–211.

Wouters, Thomas, & Groenez, Steven. (n.d.). *Overheidsbeleid en schoolse segregatie* (No. SSL/2015.8/2.3.1) (p. 79). Leuven: Steunpunt SSL.

9 Appendix

9.1 The datasets

This part of the appendix will briefly summarize other (important) variables we did not cover in our thesis. It's also a structured presentation of all datasets we used to perform our analysis.

Dataset of first-year enrollers: contains students choosing for the first year of secondary education and living in the district of Ghent (NIS-code between 44000 and 45000). This dataset only contains those students who attended primary education last year and it looks at first-year choices from 2002 until 2014. The most important variables are

> "location number": location number of a certain school institution. One school can have different locations.
>
> "location number2": based on student flows, certain school institutions are merged together in this variable. Most of the time it deals with school institutions of different schools taken together.
>
> "disadvantaged": binary variable, it has value 1 if the pupil is rather disadvantaged, 0 in other cases. The is considered disadvantaged if she/he fulfils at least one of the following criteria (also binary variables):
>
> > "education": mother has no degree of higher education
> >
> > "school allowance": receives school allowance
> >
> > "language": does not speak Dutch at home
>
> If for one of those variables a value is missing, the value on "disadvantaged" becomes "missing" as well.

"students followed over years": this dataset contains all students living in the district of Ghent and following education there as well. Here one can follow a student over all the years of secondary education (via an identification number). Important variables are:

Latitude and longitude coordinates on school institutions

The sector codes to which pupils belong (useful for calculating distances)

School where the student has followed its primary education (useful for clustering standard errors).

Number of administrative level

Acquired certificates for all students (see Word-document)

District of the pupil and the school institution

Dataset containing the centroid of all statistical sectors in Belgium (with latitude and longitude coordinates as most important variables)

Dataset containing names of choice options (specific options within a track like ASO)

Dataset of all secondary school institutions, containing information on every school year (from 2002 until 2014). Important variables are:

Number of students

Number of first-year enrollers

Percentages of students disadvantaged in some respect (“mother has no degree of higher education”, “student does not speak Dutch at home”, “student receives school allowance”). These variables provide an indication of school composition.

Dataset containing all secondary school institutions (2): same dataset as above, but different school institutions are not merged now according to student flows (which gives a higher number of observations).

Dataset “Inspection”: contains recommendations on schools based on an analysis by the inspectorate.

Identification is based on the school (i.e., all different institutions belonging to the same school).

Dataset about study records and certificates: Word-document we consulted to interpret the values on the variable “acquired certificate”

9.2 List of tables and figures

Table A1: sensitivity analysis on distance

sensitivity analysis on distance

	(1) Model 1	(2) Model A
school choice		
positive advice	0.831***	0.740***
	(0.0663)	(0.0663)
distance	-0.400***	
	(0.00209)	
percentage of disadvantaged students	-1.629***	-1.594***
	(0.0207)	(0.0207)
distance 2		-0.399***
		(0.00212)
Observations	2341607	2341607
Pseudo R-squared	0.253	0.251

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

We slightly changed our coordinates on the centroids of statistical sectors to which our students belong. We called the new variable “distance 2”, the model “Model A”. We added 0.01 to the latitude and the longitude coordinate, increasing the average travelling distance for a student now with 117.41 metres compared to the calculations used in our regressions. The coefficients on quality of education and school composition have slightly changed. This is due to the new numbers calculated for distance. The coefficient on distance is almost the same as in “Model 1”, the first simple conditional logit regression. The pseudo R-squared has only changed by 0.2 percentage points.

If we change the coordinates for example by distracting 0.01 from all latitude coordinates and adding 0.01 to all longitude coordinates in the above approach, the average travelling distance for a pupil decreases by 46.43 metres. This also does not significantly change “Model 1”. We present our calculations beneath. The new model here is called “Model B”, and its variable for distance “distance 3”.

Table A2: sensitivity analysis on distance

sensitivity analysis on distance

	(1) Model 1	(2) Model A	(3) Model B
school choice			
positive advice	0.831***	0.740***	0.817***
	(0.0663)	(0.0663)	(0.0663)
distance	-0.400***		
	(0.00209)		
percentage of disadvantaged students	-1.629***	-1.594***	-1.594***
	(0.0207)	(0.0207)	(0.0207)
distance 2		-0.399***	
		(0.00212)	
distance 3			-0.404***
			(0.00209)
Observations	2341607	2341607	2341607
Pseudo R-squared	0.253	0.251	0.248

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Table B: clustering of standard errors (in the final regression, “Model C”)

clustering the standard errors

	(1) Model 6	(2) Model C
school choice		
positive advice	1.156***	1.545***
	(0.0966)	(0.252)
distance	-0.497***	-0.500***
	(0.00533)	(0.0272)
% mother did not finish higher education	-2.412***	-2.523***
	(0.0349)	(0.195)
% not speaking Dutch at home	-1.362***	-1.193***
	(0.0488)	(0.293)
% school allowance	-1.554***	-1.478***
	(0.0461)	(0.227)
interaction quality & SES-status	-0.341*	-0.516**
	(0.133)	(0.189)
interaction distance & SES-status	-0.0155***	-0.0274**
	(0.00382)	(0.00900)
interaction 'education of mother'	2.605***	2.473***
	(0.0438)	(0.164)
interaction 'language'	2.935***	2.789***
	(0.0710)	(0.274)
interaction 'school allowance'	1.805***	1.822***
	(0.0668)	(0.153)
distance squared	0.00705***	0.00713***
	(0.000253)	(0.00119)
Observations	2341607	1928851
Pseudo R-squared	0.284	0.304

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

The Model on the right (“Model C”) shows the situation where we used clustering on the standard errors. The standard errors increased which is not a desired outcome. Note that the coefficients on the variables have also slightly changed as well as the pseudo R-squared. This is due to the lower amount of observations.

Table C: The final extension across tracks ("Model 7", "Model 8")

As noted earlier, "Model 7" presents the approximation method we proposed. "Model 8" shows the same regression where all middle schools are wiped out (no approximations). One can observe the coefficients in both models to be very similar, although sometimes more pronounced differences exist as well.

final extension across tracks

	(1) Model 7	(2) Model 8
school choice		
distance	−0.500***	−0.540***
	(0.00555)	(0.00737)
% mother did not finish higher education	−1.642***	−1.772***
	(0.0435)	(0.0726)
% not speaking Dutch at home	0.0261	−0.140
	(0.0683)	(0.0996)
% school allowance	−1.363***	−0.987***
	(0.0553)	(0.0911)
interaction distance & SES-status	−0.0215***	−0.0209**
	(0.00429)	(0.00642)
interaction 'education of mother'	1.334***	1.707***
	(0.0611)	(0.0967)
interaction 'language'	3.553***	4.946***
	(0.137)	(0.169)
interaction 'school allowance'	1.463***	1.827***
	(0.0915)	(0.135)
distance squared	0.00832***	0.00952***
	(0.000242)	(0.000273)
TSO	−1.957***	−1.744***
	(0.0877)	(0.107)
BSO	−2.606***	−1.967***
	(0.121)	(0.151)
distance * TSO	0.0790***	0.119***
	(0.0157)	(0.0180)
distance * BSO	0.0135	0.0492*
	(0.0216)	(0.0237)
distance squared * TSO	−0.00665***	−0.00697***
	(0.000937)	(0.00103)
distance squared * BSO	−0.00492***	−0.00439***
	(0.000959)	(0.000972)
interaction distance & SES-status * TSO	0.0336**	0.0270*
	(0.0103)	(0.0117)
interaction distance & SES-status * BSO	0.0695***	0.0565**
	(0.0180)	(0.0198)
% mother did not finish higher education * TSO	0.454***	0.445**
	(0.131)	(0.159)
% mother did not finish higher education * BSO	−0.519**	−0.982***
	(0.185)	(0.227)
% not speaking Dutch at home * TSO	−2.927***	−3.012***
	(0.153)	(0.203)
% not speaking Dutch at home * BSO	−1.476***	−1.769***
	(0.120)	(0.174)
% school allowance * TSO	−1.200***	−1.371***
	(0.140)	(0.173)
% school allowance * BSO	0.0179	−0.404**
	(0.120)	(0.156)
interaction 'language' * TSO	−1.003***	−1.677***
	(0.209)	(0.265)
interaction 'language' * BSO	−0.651***	−1.200***
	(0.160)	(0.212)
interaction 'school allowance' * TSO	0.564***	−0.101
	(0.128)	(0.160)
interaction 'school allowance' * BSO	0.642***	0.0891
	(0.111)	(0.145)
interaction 'education' * TSO	1.424***	1.096***
	(0.0798)	(0.0991)
interaction 'education' * BSO	2.549***	2.252***
	(0.115)	(0.139)
Observations	3312513	1166413
Pseudo R-squared	0.312	0.338

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

FACULTY OF BUSINESS AND ECONOMICS
Naamsestraat 69 bus 3500
3000 LEUVEN, België
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be