Does distance determine who is in higher education?

Sørensen, Elise Stenholt and Høst, Anders Kamp

Kraks Fond Institute for Urban Economic Research

11 December 2015

Working Paper

Does distance determine who is in higher education?

Elise Stenholt Sørensen & Anders Kamp Høst

October 2016

KRAKS FOND
INSTITUTE FOR URBAN ECONOMIC RESEARCH

Kraks Fond Byforskning
Frederiksholms Kanal 30
DK- 1220 Copenhagen K

www.kraksfondbyforskning.dk

Does distance determine who is in higher education? [1]

Elise Stenholt Sørensen
Kraks Fond Institute for Urban Economic Research

Anders Kamp Høst
Ministry of Higher Education and Science

October 2016

Abstract

We assessed the effect of distance to higher education institutions on education enrolment. Furthermore, we analysed how parental education and geographic region affect the relationship between distance and enrolment. We employed Danish administrative data of high school students from 2006–2013 and found no relationship between distance and the decision to enrol in higher education, when controlling for individual and parental characteristics. However, the results did suggest a small negative association between distance and enrolment among students in cases where neither of their parents completed a higher education. However, this only applied in Central Jutland Region and Southern Denmark Region.

JEL: C25, I21, R10
Keywords: High school graduates, higher education, geographical accessibility, distance to school

[1] We would like to thank Lars Pico Geerdsen for insightful and constructive comments

1. Introduction

The purpose of this paper was to examine the following questions: Does distance to higher education institutions act as a deterrent to enrolment? Are high school students with parents who do not have a higher education, more affected by longer distances compared with high school students with parents who themselves have completed a higher education? And is the relationship between distance and enrolment more pronounced in some geographical regions of Denmark?

Distance to higher education institutions may act as a deterrent to attendance because of transaction costs. The greater the distance to an educational institution, the higher those costs. These may involve direct financial costs if the student has to leave home or to commute to their place of study. There may also be emotional costs associated with leaving home, and students may be reluctant to leave their home town because of their network of family and friends. Therefore, policy makers usually assume that the spatial distribution and thus the access to higher education is a relevant determinant of a country's educational results (Vestergaard, 2015). However, we know relatively little about whether and how distance affects educational decisions in general and in a Danish context in particular.

The empirical evidence from other western countries is inconclusive. Half of the studies we found, indicated a small negative effect (Frenette, 2009; Öckert, 2012; Spiess & Wrohlich, 2010), whereas the other half concluded that there was no effect(Cullinan, Flannery, Walsh, & McCoy, 2013; Steve Gibbons & Vignoles, 2009; Sá, Florax, & Rietveld, 2006). The application of international evidence in a Danish context is hampered by country variation with respect to the geographical dispersion of higher education institutions, general access to the educational system, and differences in the financial provision of the welfare systems. Hence, empirical evidence from Denmark is needed.

This study examines the effect of distance to higher education institutions on enrolment, using administrative data from Statistics Denmark in combination with detailed distance measures. This gives us rich information about the full population of Danish senior year high school students in 2006–2013, their demographic characteristics and educational abilities, as well as the family and parental background. As a first step, we mapped the geographical location of higher education institutions in Denmark, and measured the actual distance from each student's residential address to the nearest institution. Secondly, we estimated the relationship between distance to higher education and enrolment, specifying a discrete choice model with distance to the nearest college, at the time when the student graduated from high school[2], as an explanatory variable. We control for the individual characteristics and parental background of the high school students in order to separate the causal effect of distance from the correlation between parental residential sorting and distance.

The empirical analysis revealed that – controlling for socio-economic characteristics – there is no relationship between distance to higher education institutions and enrolment. However, our results suggest that there is a small but significant relationship between distance and enrolment for a subgroup of the sample; among students from families with no higher education, the negative association between distance and enrolment was statistically significant. However, this only applies to students from families living in the region of Central Jutland or Southern Denmark.

The rest of the paper is organised as follows. Section 2 reviews the existing empirical literature on the relationship between distance and educational attendance. Section 3 introduces the Danish educational system and maps the geographical location of educational institutions in Denmark. Section 4 outlines the theoretical framework describing why distance may influence educational decisions. Section 5 describes

[2] We use the term 'high school' equivalent to qualifying upper secondary education throughout the paper. Upper secondary educations consists of academic tracks such as mathematical, technical or linguistic tracks.

the empirical strategy, the model, and the data and variables used in the analysis. Section 6 presents the estimation results and section 7 presents relevant sensitivity analysis. Finally, section 8 discusses the results and presents the conclusions drawn.

2. Empirical literature

The literature on the effects of distance to school on educational participation is mainly rooted in two research questions: (1) Does geographical distance influence the decision to study at all, representing a deterrent effect, and (2) does proximity to different educational institutions have an effect on the choice between different education types. In addition, distance to school has been used in several studies as an instrument for college attendance, attempting to estimate the return to education e.g. (Card, 1993; Currie & Moretti, 2003; Kane & Rouse, 1995). This section provides a short literature review, with an emphasis on previous evidence of a deterrent effect, as the goal of this study is to examine whether distance acts as a barrier to studying at all.

2.1 Existing evidence of distance as a deterrent

The existing evidence on the relationship between distance to the following higher education institution, and the probability of enrolling is inconclusive. Half of the studies we found, indicated a small negative effect (Frenette, 2006; Öckert, 2012; Spiess & Wrohlich, 2010), whereas the other half indicated no effect (Cullinan et al., 2013; Steve Gibbons & Vignoles, 2009; Sá et al., 2006). Furthermore, some studies have found that the deterrent effects of distance are stronger or only apply for adolescents from low socio-economic status households (Denzler & Wolter C., 2011; Frenette, 2006).

Based on Swedish data, Öckert (2012) found evidence for a deterrent effect of distance on college enrolment among a full population sample of Swedish males (cohorts born in 1974–1976). Adolescents living more than 25 km from college have an average 4% point lower probability of enrolling in college than those who lived near a college[3]. Similar, Spiess & Wrohlich (2010) found a reduced university participation above a 12,5-km threshold, employing survey data encompassing 1.219 German high school leavers. Based on a Canadian household survey, Frenette's analysis from 2004 and 2006 demonstrated that students living beyond commuting distance (80 km) were less likely to attend university than students living within commuting distance. Further, he found that students from lower income families were particularly disadvantaged by distance (Frenette, 2004, 2006).

However, there is also empirical evidence suggesting that there is no deterrent effect of distance in data from The Netherlands, Ireland or the United Kingdom. Sá et al. (2006) employed survey data on 3.263 Dutch students finishing their senior high school year in 1998–2000. They found no deterrent effect on the general participation across the geographic access to community colleges and universities[4]. Similarly, based on survey data on 858 Irish high school leavers in 2007, Cullinan et al. (2013) found no deterrent effect of proximity to the nearest higher education institution, encompassing both community colleges and universities, on participation. However, they did find reduced participation among students without university-educated parents. Employing a full population sample of British adolescents at the age of 16, Gibbons & Vignoles (2009) found that distance to the nearest university had little or no deterrent effect on the probability of continuing to university.

[3] Conversely, Öckert (2012) found that adolescents living more than 25 km from college had a higher completion rate and earnings after graduation. This is probably due to the fact that among these adolescents, only those with good education grades found it worthwhile to continue to higher education.

[4] The distance measure employed by Sá et al. (2006) is calculated as the number of eligible colleges or universities weighted by the distance to each institution, giving institutions further away less weight. The included institutions are selected such that they can be attended by students with a given high school profile (Science and technology, Health, Culture and society, Economics and society). Hence, distance varies across space and individuals.

2.2. Related studies of distance and study choices

A handful of studies exploited college openings and expansions as a source of exogenous variations (Frenette, 2009; Holzer, 2007; Oppedisano, 2011) to overcome the possible problem of endogeneity due to household sorting. These studies examined the effect of the opening of a college within a given radius. Frenette (2009) compared students who lived within a radius of 80 km from a college, with students who lived further away. In the studies of Holzer (2007) and Oppedisano (2011), on the other hand, students were assumed to be affected by the opening, if it occurred within the same municipality or region[5]. All the studies found that an expansion of the local supply of higher education, increased participation in higher education. However, the studies lack variation, as college and university openings rarely occur. Therefore their results could should be applied with precaution.

A related empirical strand involves the effect of distance on the choice between alternative studies – a so-called diversion effect. Most of the studies find that distance influences the choice of study/institution among students – especially if the parents do not hold a university degree (Denzler & Wolter C., 2011; Frenette, 2009; Steve Gibbons & Vignoles, 2009; Griffith & Rothstein, 2009; Sá et al., 2006). Frenette (2009) estimated the effect of newly established community colleges and universities and concluded that the opening of a new university increases the share of young people enrolling at a university by 6,4 % points. However, this occured at the expense of a reduction in the probability of enrolling at a community college by 5 % points. Therefore, the expansion of higher education to a large extent lead to diversion effects. A similar result were found in a Danish study, showing that students resident more than 40 km from a university enrolled in college rather that university educations. However, this diversion effect were not present among students for whom at least one of the parents had a university degree (Høst & Sørensen, 2015).

A large part of the variation in the presence and size of the deterrent effects of distance may originate from country-specific differences. These differences encompass within-country variation in population density and location of higher education institutions, the general access to the educational system and the provisions of the welfare system. In addition, some of the differences may originate from differences in methodological approaches. This makes it difficult to apply the international evidence in a Danish context. Therefore, evidence based on Danish data is necessary to inform the policy debate about access to higher education in Denmark.

3. The Danish educational system and the geography of higher education institutions

In this section, we give a description of the Danish educational system and the requirements for admission to a higher education. We do this, to clarify the primary restriction to our sample; it only comprises students who have completed an upper secondary education. We also present the geographic location of colleges and universities from 2006 to 2013 and describe the general association between proximity and enrolment across municipalities.

3.1. The Danish educational system

Compulsory education in Denmark consists of primary and lower secondary elementary education, from grade 0 (age 5–6) to grade 9 (age 15–16). After 9th grade, further education is voluntary. Pupils can choose to leave the educational system, continue in 10th grade or enter upper secondary school. Students that do not directly enrol in upper secondary school may enter later, with no loss of rights or opportunities for enrolment.

[5] They did not have further quantification of the distance.

The upper secondary school comprises two main tracks of education: upper secondary education and vocational secondary education. Upper secondary education consists of academic tracks, such as mathematics, technical studies and linguistics. Vocational secondary education consists of branch-specific tracks such as carpentry, bricklaying, mechanics and hairdressing. Here students shift between school-based learning and practical apprentice training.

Tertiary education in Denmark is on three levels: lower tertiary education (LTE), intermediate tertiary education (ITE) and higher tertiary education (HTE). A formal requirement for admission to all tertiary education is an upper secondary education. However, some lower- and intermediate tertiary tracks allow students to use selected vocational secondary education to meet the admission requirement.

LTE, corresponding to 1–2 years of full time studies, includes a wide range of branch-specific education such as social and health assistants, agricultural or industrial diplomas and business education. ITE, comprising 3–4 years of education, encompasses a range of education aimed at the public welfare and health sectors, e.g. nurses and midwives, child care workers and elementary school teachers, as well as a range of technical training such as electrical and mechanical engineering. In the following, we refer to LTE and ITE institutions as colleges. HTE comprises all types of education at the universities (Bachelor and Master's programs) and range from 3 to 8 years of full time educational programs. It normally takes 5–6 years to complete this type of education.

3.2 The geography of higher education institutions and enrolment in Denmark

Figure 1 depicts the geographic location of colleges and universities in 2006–2013. Colleges in Denmark (small red dots) in general are geographically distributed widely across the country. This is in line with the Danish political strategy to ensure regional access to higher education[6]. Universities (dark red pentagons) are primarily located in the largest cities. It should be noted, that the following analysis does not distinguish between colleges and universities, but look at the total enrolment in higher educations.

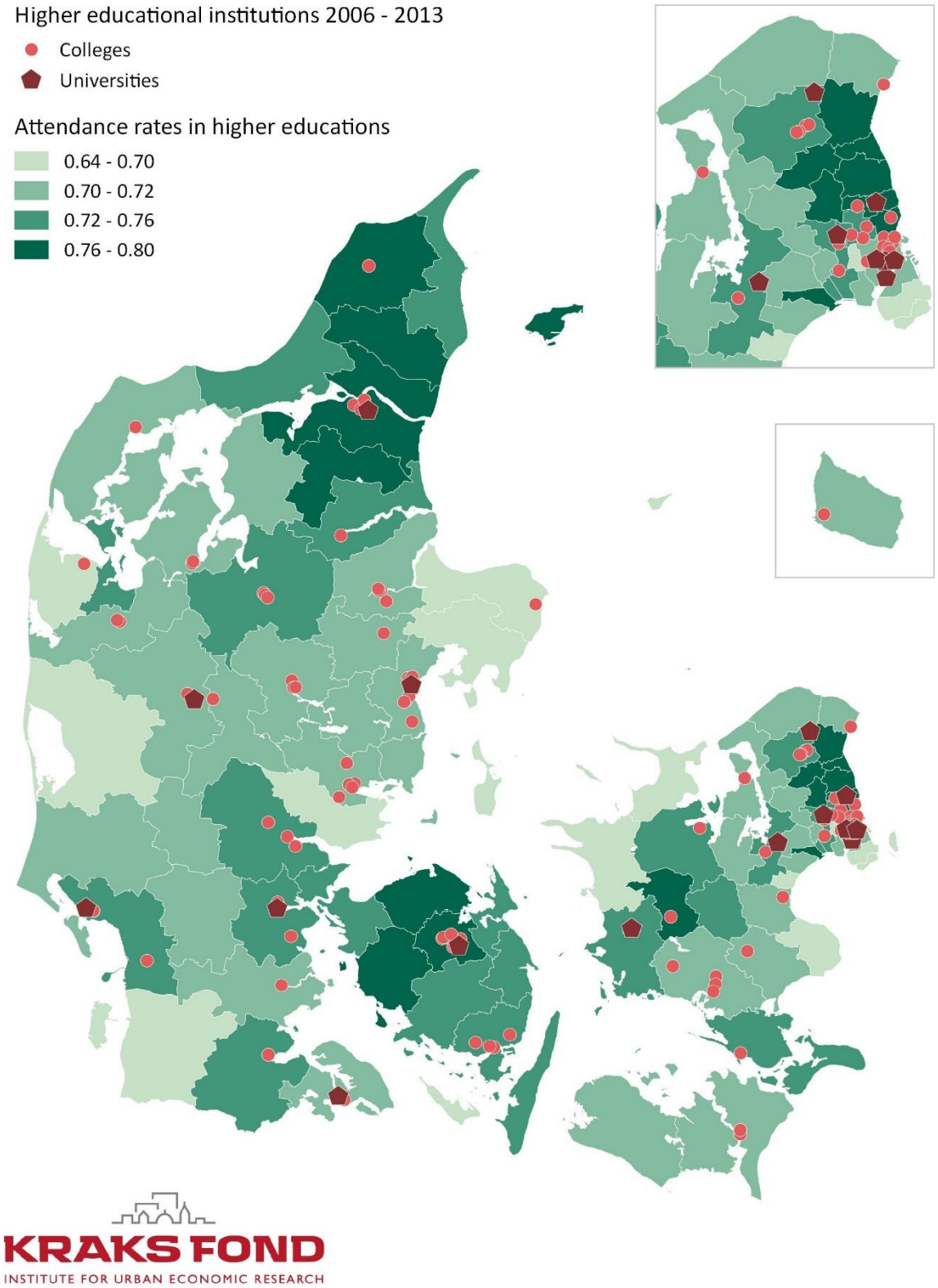
There are relatively large differences in higher educational enrolment rates across municipalities in Denmark. Figure 1 presents the percentages of all high school graduates within each municipality who enrolled in higher education within 2 years after graduation. The data comprise students who were between 18 and 21 years old at the time of graduation.

Further, Figure 1 shows that there is an overlap between the location of educational institutions and enrolment rates. The enrolment rates are highest in the municipalities with numerous higher educational institutions, i.e. urban municipalities such as the municipality of Odense and the municipality of Aalborg[7]. However, inconsistent with this pattern, the municipality of Århus and the municipality of Copenhagen have enrolment rates below the national average of 73%, 72% and 71%, respectively. And, at the same time, the municipalities in Northern Jutland and north of Copenhagen have above-average enrolment rates.

[6] The term 'colleges' covers both business academies and university colleges.

[7] In the municipality of Odense the enrolments rats were 76%, on average, and in municipality of Aalborg 79%, from 2006-2013.

Figure 1: Geographical differences in attendance rates in higher education



4. Theoretical framework

There is an extensive literature concerning the transition to higher education, and the reasons behind the decision to attend college or university. We present a short overview here, focusing on aspects that are relevant to educational decisions of high school graduates and the theoretical explanations for the effect of distance on study behaviour.

4.1 Choice behaviour of high school leavers

Why do young people choose to go into higher education? Conventional human capital theory looks at education as an investment good. The theory assumes that individuals are rational and forward-looking, making cost-benefit calculations on the basis of future returns of education e.g. (Becker, 1964). A student chooses to invest in education if the expected benefit is higher than the expected cost. The benefit is denoted by the expected increase in lifetime earnings. Costs include the expected forgone earnings, direct costs of tuition and transport, the effort required for graduation, and the risk of failure. Theories that consider education as a rational cost-benefit decision assume that direct and indirect costs would influence the decision to attend higher education.

The direct costs of obtaining a higher education, such as tuition fees, have received much attention in empirical research e.g. (Bishop, 1977; Fuller, Manski, & Wise, 1982). However, the direct financial costs are not relevant in the present study, as tuition fees for higher education are subsidised by the State of Denmark[8]. However, transactions costs could play a role and this will be addressed in detail in the following section. Another strand of the literature focuses on the relationship between participation in higher education and socio-economic background. Typical findings are that the education of the parents and household income increase the probability of enrolling in higher education e.g. (Acemoglu & Pischke, 2001; Shea, 2000).

Previous empirical studies have also found a series of other individual, family and school characteristics to be relevant. Gender plays a key role in participation; generally speaking, females are more likely to graduate from high school and to attend university (Frenette, 2006). Ethnicity is also an important determinant of differences in college enrolment (Black & Sufi, 2002) as well as the type of secondary school that the student has attended (Nguyen & Taylor, 2003). Finally, spatial aspects such as urbanisation and distance have been shown to play a role in choice behaviour, even though the geographic aspects have played a minor role in economic analyses of the transition to higher education (Stephen Gibbons & Silva, 2008).

4.2. Theoretical explanations for distance effects

One would expect a negative association between distance to higher education institutions and the probability of studying. In the literature the explanations for such an association are divided in four types:

1. Pecuniary costs: Distance is associated with cost, such as commuting costs, the cost of accommodation outside the parental home, and the cost of moving.
2. Emotional cost: Individuals are rooted in their local environment and may be unwilling to leave their social network.
3. Selection effects: Higher educated parents tend to live in the largest cities, where the job market matches their skills. Therefore, highly educated parents tend to live in areas a short distance from colleges and universities.
4. Peer effects: If people who are alike (e.g. academics, unskilled workers, migrants) tend to live in similar locations, with a similar distance to colleges and universities, they will have a mutual influence on each other concerning their study decisions.

[8] The few private institutions in our sample are excluded from the analysis, to ensure that the conditions of applications are kept equal.

Explanations 1 and 2 explain why distance to school can have a causal effect on study decisions, for either pecuniary or emotional reasons. The other two explanations, 3 and 4, indicate that study decisions may also be influenced by other factors, such as parental education and friends in high school. These factors are associated with the choice of residence and the distance to higher education institutions. Therefore, the empirical design should be able to separate the causal effect of distance from the non-causal effects from residential sorting. In other words, the attempt of the statistical models is to exclude selections effects via control variables.

5. Empirical framework

5.1. General model

The empirical analysis is based on the linear probability model in equation 1. The model estimates the probability of enrolling in higher education y_i given the geographical distance D, while controlling for individual-specific covariates $\boldsymbol{x}$, and family-specific covariates $\boldsymbol{z}$.

$$\mathrm{P}(y_i = 1|D_i, \mathbf{x}_i, \mathbf{z}_i) = \alpha + \delta D_i + \boldsymbol{\beta}\mathbf{x}_i + \boldsymbol{\gamma}\boldsymbol{z}_i + \varepsilon_i \quad (1)$$

where D_i is a continuous variable that measures distance to the nearest higher education institution. The associated parameter δ indicates the change in the enrolment rate by each kilometre increase in the distance to the nearest higher education institution. The vector $\mathbf{x}_i$ comprises individual characteristics such as gender, age, immigration status, year of graduation, type of secondary education, grade point average from elementary school, and geographical region. The vector $\boldsymbol{z}_i$ comprises socio-economic variables such as mother's and father's highest education, log of household income and family type. The term ε_i is an idiosyncratic error term, assumed to be independent and identically distributed. The explanatory variables $(D_i, \mathbf{x}_i, \mathbf{z}_i)$ are measured in the year of graduation, and the dependent variable (y_i) is measured up to 2 years after high school graduation, that is, if the student enrols at any higher education institution in Denmark within 2 years of graduating from high school.

5.2. Data and variables

We obtained data from administrative registers retained and maintained by Statistics Denmark. The population consists of the full population of high school graduates aged 18–21 years at the time of graduation during 2006–2011. Students living on an island without a bridge to the mainland are excluded from the analysis. Furthermore, we excluded students who enrolled in one of the few private institutions, where there are tuition fees, and students who enrolled in a military education, or an education Furthermore, students who enrolled in military, police, maritime and educations in creative subjects are omitted, mainly because we lack distance measures to these educations. In total, those students comprise about 5% of the population.

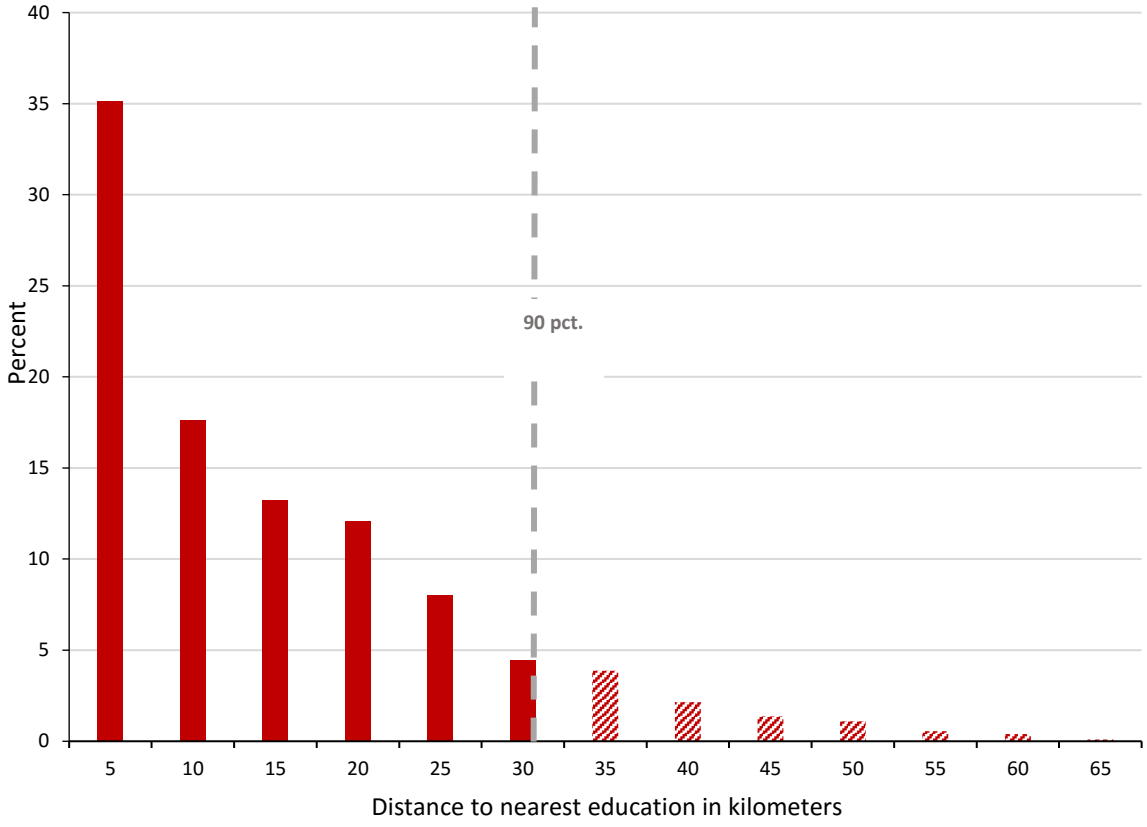
5.3. Distance measure

The explanatory variable of interest, i.e. distance to higher education institution, is measured as the distance via the road networks from the student's residential address to the nearest institution[9]. About 84% of the students in the sample were living with their parents when they graduated from high school. In these cases, we measured the distance from their parents' residence on 1 January of the year they graduated. For the residual group of students who left home before graduation, we measured the distance from the latest address where they were resident with one or two parents.

[9] To calculate the distances for the students, we use the exact geographic co-ordinates for each of the identified educational institutions and the co-ordinates for the bottom-left corner of the geographic defined quadrant 1 x 1 km in size in which the residence of the student was located.

The average distance to a higher education institution was 12 km for the observations in our sample, with a minimum distance is 0 km and a maximum distance of 67 km to the nearest college or university (see Table 1). Figure 2 shows a histogram of distance to the nearest higher education institution. The majority of the population have access to such an institution within a relatively short distance; over 90% are less than 30 km from the nearest education institution (illustrated by the solid red bars in Figure 2) and 35% are within 5 km (the hatched bars). For a more detailed illustration of the distance distribution in 1-km intervals, see Figure A1 in the Appendix.

Figure 2: Histogram of the distance to the nearest higher education institution



5.4 Control variables

Tables 1 and 2 present the descriptive statistics on the explanatory variables. About 57% of the sample consisted of females and almost half of the population graduated from high school when they were 19 years old (Table 1). In all, 92% of the sample consisted of students with Danish ethnicity, and the majority of the students (61%) graduated from a general high school ('Stx'). We also control for grade point average (GPA) in written maths and written Danish in 9th grade; both variables have a mean of around 7 and range from -3 to 12.

Table 1: Individual specific variables

Variable		Mean	Std. Dev.
Y: Enrolled in a higher education, within two years		0,734	
D: Distance to nearest higher education		12,146	11,620
Gender	Female	0,571	
	Male	0,429	
Age	Age 18	0,219	
	Age 19	0,496	
	Age 20 - 21	0,285	
Immigration status	Danish	0,925	
	First generation immigrant	0,033	
	Second generation immigrant	0,042	
Secondary education	Htx	0,077	
	Stx	0,610	
	HF	0,106	
	Hhx	0,204	
	IB	0,002	
	diplom	0,001	
GPA	Written Danish exam	7,197	2,324
	Written math exam	7,201	2,461
Year of graduation	2006	0,170	
	2007	0,151	
	2008	0,158	
	2009	0,171	
	2010	0,176	
	2011	0,175	
Region	Northern Jutland	0,114	
	Central Jutland	0,244	
	Suthern Denmark	0,228	
	Capital Region	0,269	
	Zealand Region	0,144	

N = 178.868

We have also included detailed information about student's socio-economic background, in the analysis. Table 2 shows the distribution of mother's and father's highest education, the family type and the log of the household income.

Table 2: Family-specific covariates

Variable		Mean	Std. Dev.
Family type	Nuclear family	0,609	
	Divorced with new partner	0,081	
	Single parent	0,157	
	Not living with parents	0,153	
	Mother missing	0,012	
	Father missing	0,049	
Mother's education	Elementary school	0,166	
	High School	0,053	
	Vocational	0,336	
	Short-cycle higher education	0,052	
	Medium cycle higher education	0,299	
	Long-cycle higher education	0,093	
Father's education	Elementary school	0,209	
	High School	0,051	
	Vocational	0,373	
	Short-cycle higher education	0,071	
	Medium cycle higher education	0,152	
	Long-cycle higher education	0,142	
Household income (Log)		13,069	0,399

N = 178.868

6. Estimation results

The descriptive analysis above have demonstrated that despite the fact that the majority of high school students in Denmark, live within commuting distance to a higher education, those living in the city regions are still more likely to attend a higher education. The following analysis will reveal, how much of this association, that are due to the fact, that higher educated parents tend to live in the largest cities, closer to higher education institutions (see Figure A2 in Appendix), and how much that are cause by the direct cost of longer distances. In short, we examine whether distance to higher education's refrain high school students from enrolling in a higher education.

We present the empirical analysis, in three parts. First, we present the results from the main analysis examining the average effects of distance to higher education. Secondly, we examine whether the effect of distance is dependent on the parents' formal education, by splitting the sample into four subsamples, one for each socio-economic. We then take a step further and examine whether the influence of parents' education on distance, differs between geographic regions in Denmark.

6.1. Main results

In this section, we present the results from the main analyses, examining the effect of distance to education on enrolment rates in higher education. We attempt to separate the causal effect of distance from the non-causal effects of residential sorting, using the linear probability model, describe in section 5.1.

Figure 3 illustrates the marginal effect of distance to nearest education institution, based on the parameter estimates from Table 3. Control variables were held to their average when the marginal effects were calculated. The dotted line in Figure 3 shows the relationship between distance and enrolment, without any controls. It confirms that individuals living far from higher education institutions, on average, are less likely

to enrol in higher education. However, controlling for individual-specific covariates (the red dashed line) the effect of distance vanishes. After controlling for both individual-specific and parental covariates, the relationship is statistically insignificant (the red solid line in Figure 3). In other words, the revealed negative relationship between distance and enrolment is explained solely by residential sorting of the parents (see Table 3 for parameter estimates). When we control for selection effects in terms of parents' socio-economic status, there is no relationship between distance to the nearest higher education institution and enrolment.

Figure 3: Participation in higher education by distance

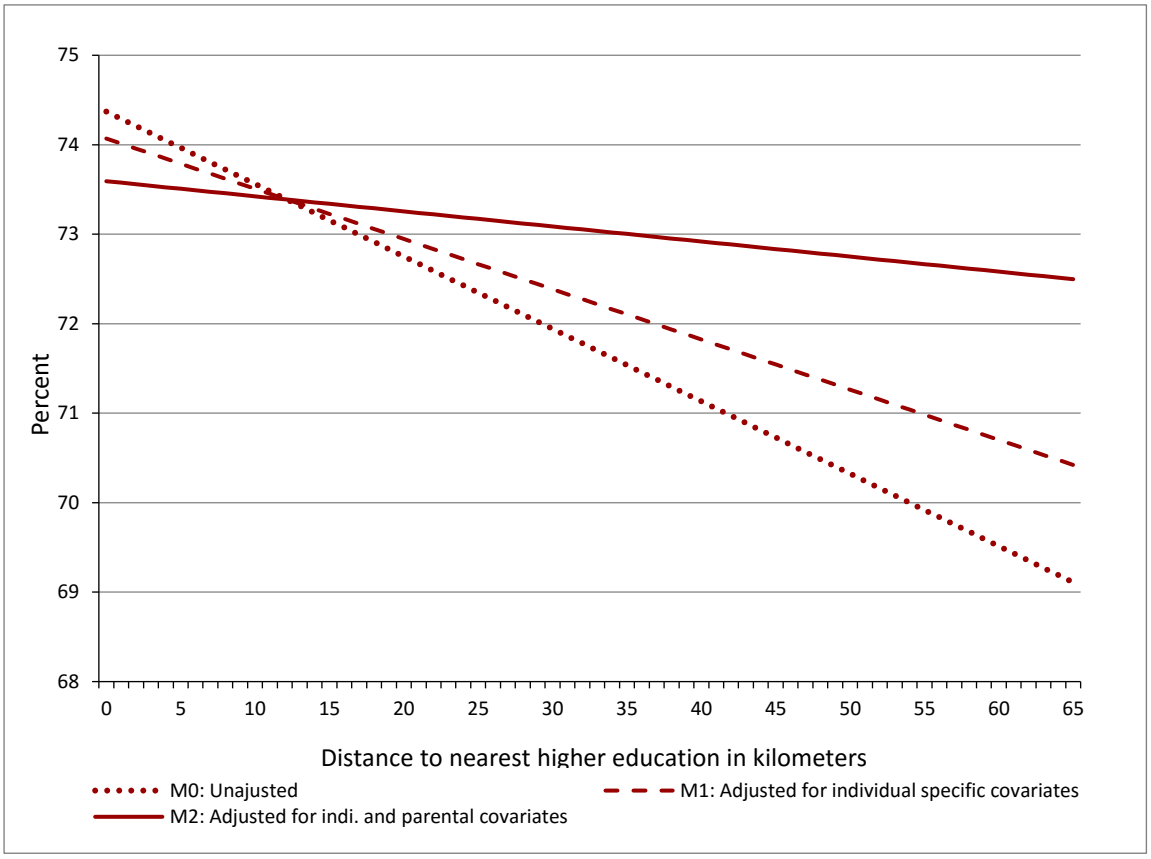


Table 3 shows the parameter estimates from the three different models, that we estimated. Column 1 gives the parameter estimates from Model 0, without controls, column 2, presents the distance effect, adjusted for individual specific covariates (Model 1) and in column 3, you see the distance estimates adjusted for both individual and parental covariates (Model 2).

Table 3: Linear regression model results: participation in higher education

	Model 0		Model 1		Model 2	
	B	se	B	se	B	se
Distance (km)	-0,0008***	0,0001	-0,0006***	0,0001	-0,0002	0,0001
Individual specific covariates:						
Gender (Ref.: Female)						
Male			-0,0723***	0,0022	-0,0823***	0,0022
Age (Ref.: Age 19)						
Age 18			-0,0002	0,0026	-0,0035	0,0026
Age 20 - 21			0,0056*	0,0024	0,0110***	0,0024
Immigration status (Ref.: danish)						
First generation immigrant			0,1412***	0,0057	0,1698***	0,0059
Second generation immigrant			0,1543***	0,0052	0,1910***	0,0054
Secondary education (Ref. Htx)						
Stx			-0,0044	0,0040	-0,0219***	0,0040
HF			-0,1196***	0,0050	-0,1163***	0,0050
Hhx			-0,1401***	0,0043	-0,1390***	0,0043
IB			-0,2322***	0,0217	-0,2517***	0,0216
diplom			0,1847***	0,0295	0,1930***	0,0294
Year of graduation (Ref.: 2006)						
2007			0,0313***	0,0036	0,0240***	0,0036
2008			0,0612***	0,0035	0,0532***	0,0035
2009			0,0746***	0,0035	0,0653***	0,0035
2010			0,0762***	0,0034	0,0641***	0,0035
2011			0,0704***	0,0035	0,0564***	0,0035
GPA Written Danish exam			0,0075***	0,0005	0,0063***	0,0005
Written math exam			0,0253***	0,0005	0,0225***	0,0005
Region (Ref.: Northern Jutland)						
Central Jutland			-0,0432***	0,0037	-0,0473***	0,0037
Southern Denmark			-0,0127***	0,0037	-0,0138***	0,0037
Capital Region			-0,0395***	0,0038	-0,0522***	0,0038
Zealand Region			-0,0242***	0,0040	-0,0269***	0,0040
Family specific covariates						
Family type (Ref.: Nuclear family)						
Divorced with new partner					-0,0288***	0,0039
Single parent					-0,0306***	0,0030
Not living with parents					-0,0457***	0,0031
Mother missing					0,0527***	0,0097
Father missing					0,0400***	0,0057
Mother's education (Ref.: Elementary school)						
High School					0,0153**	0,0051
Vocational					0,0058	0,0032
Short-cycle higher education					0,0260***	0,0052
Medium cycle higher education					0,0239***	0,0034
Long-cycle higher education					0,0505***	0,0047
Father's education (Ref.: Elementary school)						
High School					0,0134**	0,0052
Vocational					0,0028	0,0030
Short-cycle higher education					0,0241***	0,0046
Medium cycle higher education					0,0278***	0,0038
Long-cycle higher education					0,0570***	0,0042
Household income (Log)					0,0463***	0,0034
Intercept	0,7437***	0,0015	0,5422***	0,0071	-0,0354	0,0445
R2	0,000		0,068		0,077	
N	178.868		178.868		178.868	

*** p<0.001, ** p<0.01, * p<0.05

6.2. Parental education and distance effects

In line with previous studies, we further examined whether the effect of distance varies across parental education. We expected that, if there was a negative effect of distance in Denmark, we would see it among students from families in which neither of the parents holds a college or university degree. To test this hypothesis, we divided the sample into four subsamples according to the highest education of the parents. Group 1 includes students where neither mother or father has a qualifying educational degree[10]. Group 2 includes students for whom the highest education of both parents is a vocational education[11]. Group 3 includes students where the highest education of both parents is a LTE or ITE (2-4 years of higher education). Finally, group 4 includes students for whom at least one of the parents holds a university degree.

Estimation results are presented in Table 4. Each column in the table shows the parameter estimates for the full model for each educational group. The second column shows that students from group 2, i.e. families where the highest parental education is vocational, are less likely to enrol when distance is longer. However, the marginal distance effect is small – if distance increases by 10 km, the probability of enrolling decreases by 0,5% points, for students in this group. In other words, the probability of enrolling among those living furthest away (60 km) is on average 3% points lower than among students living next to a higher education institution (less than 1 km). The effect remains statistically insignificant different from zero for the other groups. In sum, when we split the sample with regard to highest parental education, we find little evidence suggesting that the effects of distance differ according to parents' education.

[10] That is, they have completed high school, they have completed primary school, or they have not completed any education. If the education information for the parents is missing, the student is also in this group. If one parent hold a vocational educational degree and the other parent has no education, the student is still in this group.

Table 4: Linear regression model results: participation in higher education by parental education

	1 : No vocational edu.		2: Vocational edu.		3: College edu.		4: University edu.	
	B	se	B	se	B	se	B	se
Distance (km)	-0,0005	0,0003	-0,0005***	0,0002	0,0001	0,0002	0,0002	0,0002
Individual specific covariates:								
Gender (Ref.: Female)								
Male	-0,0872***	0,0079	-0,0941***	0,0040	-0,0891***	0,0035	-0,0455***	0,0043
Age (Ref.: Age 19)								
Age 18	-0,0078	0,0094	-0,0085	0,0048	-0,0014	0,0043	0,0014	0,0048
Age 20 - 21	0,0137	0,0086	0,0136**	0,0042	0,0144***	0,0039	-0,0045	0,0054
Immigration status (Ref.: danish)								
First generation immigrant	0,1814***	0,0132	0,1665***	0,0111	0,1948***	0,0112	0,1167***	0,0126
Second generation immigrant	0,2213***	0,0112	0,2022***	0,0098	0,1622***	0,0107	0,0894***	0,0147
Secondary education (Ref. Htx)								
Stx	-0,0509**	0,0146	-0,0187*	0,0068	-0,0248***	0,0063	-0,0053	0,0097
HF	-0,1299***	0,0168	-0,1001***	0,0083	-0,1389***	0,0081	-0,1233***	0,0133
Hhx	-0,1853***	0,0153	-0,1640***	0,0070	-0,1123***	0,0070	-0,0629***	0,0117
IB	-0,3693***	0,0725	-0,2439***	0,0509	-0,2615***	0,0367	-0,1927***	0,0319
diplom	0,1798	0,0814	0,1957***	0,0463	0,2299***	0,0532	0,0674	0,0776
Year of graduation (Ref.: 2006)								
2007	0,0058	0,0121	0,0419***	0,0064	0,0139**	0,0058	0,0150*	0,0073
2008	0,0571***	0,0121	0,0785***	0,0063	0,0478***	0,0057	0,0108***	0,0073
2009	0,0976***	0,0120	0,0959***	0,0062	0,0487***	0,0056	0,0234***	0,0071
2010	0,0882***	0,0122	0,0949***	0,0062	0,0504***	0,0057	0,0172***	0,0071
2011	0,0819***	0,0126	0,0968***	0,0063	0,0379***	0,0057	-0,0001	0,0071
GPA Written Danish exam	0,0052**	0,0017	0,0076***	0,0009	0,0066***	0,0008	0,0058***	0,0010
Written math exam	0,0227***	0,0016	0,0219***	0,0008	0,0242***	0,0008	0,0220***	0,0010
Region (Ref.: Northern Jutland)								
Central Jutland	-0,0345*	0,0138	-0,0507***	0,0061	-0,0442***	0,0059	-0,0439***	0,0084
Southern Denmark	0,0175	0,0137	-0,0307***	0,0061	-0,0042	0,0059	-0,0098	0,0087
Capital Region	-0,0491**	0,0140	-0,0826***	0,0067	-0,0405***	0,0062	-0,0274**	0,0081
Zealand Region	-0,0344*	0,0148	-0,0394***	0,0067	-0,0170*	0,0064	-0,0153	0,0093
Family specific covariates								
Family type (Ref.: Nuclear family)								
Divorced with new partner	-0,0315*	0,0138	-0,0379***	0,0067	-0,0153	0,0063	-0,0335***	0,0081
Single parent	-0,0115	0,0109	-0,0289***	0,0054	-0,0341***	0,0048	-0,0330***	0,0058
Not living with parents	-0,0236***	0,0097	-0,0378***	0,0052	-0,0557***	0,0052	-0,0598***	0,0075
Mother missing	0,0268***	0,0207	0,0566***	0,0164	0,0336***	0,0194	0,0325***	0,0203
Father missing	0,0118***	0,0131	0,0320***	0,0102	0,0422***	0,0091	0,0313***	0,0133
Household income (Log)	0,0203	0,0110	0,0390***	0,0063	0,0665***	0,0059	0,0629***	0,0059
Intercept	0,3166**	0,1426	0,0613**	0,0818	-0,2751	0,0767	-0,1784	0,0795
R2	0,091		0,070		0,064		0,054	
N	14.998		63.001		67.530		33.339	

*** p<0.001, ** p<0.01, * p<0.05

6.3. Participation in higher education by distance, parents' education and geographic region

Finally, we examined whether there were regional differences in the distance effect among students from group 2. Thus, we restricted the sample to students from families where the highest parental education was a vocational education (*n* = 63.001). We split this sample into five different groups according to the geographic region of their residential address. If distance posed a barrier to attending a higher education institution, in terms of travel time, then we might expect that distance to have a smaller effect in the Capital Region, where the public transportation network is more extensive, compared with other regions.

To examine whether the distance effect was more significant in the peripheral regions of Denmark, we ran a linear regression model, equivalent to model (1), for each region. The results in Table 6 suggest that the distance effect was only present in the Region of Central Jutland and Southern Denmark. However, the effects were still relatively small, suggesting a marginal 0,9% point decrease in enrolment when the

distance was increased by 10 km. As expected, distance had no effect in the Capital Region. Likewise, we found no effect in the Northern Jutland Region and in Zealand Region. This is rather surprising considering the geographic distribution of higher education institutions (Figure 1) and the distance distribution for each region (as depicted in Figure A3 in the Appendix). Figure A3 in the Appendix shows that Northern Jutland Region had the most dispersed distance distribution of the regions in Denmark.

Table 5: Linear regression model results for students from families with a lower education: participation in higher education by geographic region.

	Northern Jutland		Central Jutland		Southern Denmark		Capital Region		Zealand Region	
	B	se	B	se	B	se	B	se	B	se
Distance (km)	0,0001	0,0004	-0,0009***	0,0003	-0,0008*	0,0003	0,0004	0,0007	-0,0006	0,0004
Individual specific covariates:										
Gender (Ref.: Female)										
Male	-0,0931***	0,0107	-0,0931***	0,0079	-0,1018***	0,0080	-0,0728***	0,0086	-0,1173***	0,0101
Age (Ref.: Age 19)										
Age 18	-0,0184	0,0131	-0,0245*	0,0104	-0,0160	0,0100	0,0026	0,0096	0,0095	0,0115
Age 20 - 21	0,0124	0,0111	0,0072	0,0081	0,0170*	0,0083	0,0136	0,0102	0,0268*	0,0111
Immigration status (Ref.: danish)										
First generation immigrant	0,1339***	0,0375	0,1805***	0,0240	0,1585***	0,0215	0,1813***	0,0214	0,1581***	0,0289
Second generation immigrant	0,1387**	0,0511	0,1657***	0,0233	0,1477***	0,0240	0,2493***	0,0145	0,1691***	0,0288
Secondary education (Ref. Htx)										
Stx	-0,0085	0,0179	0,0101	0,0133	-0,0081	0,0137	-0,0376*	0,0156	-0,0785***	0,0177
HF	-0,0599**	0,0219	-0,0793***	0,0162	-0,0910***	0,0169	-0,1433***	0,0188	-0,1348***	0,0211
Hhx	-0,1792***	0,0183	-0,1279***	0,0133	-0,1618***	0,0140	-0,1489***	0,0170	-0,2420***	0,0184
IB	-0,3854*	0,1797	-0,3261**	0,1220	-0,0690	0,0866	-0,3784***	0,0959	-0,1837	0,1509
diplom	0,1779*	0,0856	0,2722**	0,0854	0,2275*	0,0919	0,2571	0,2035	0,0584	0,1369
Year of graduation (Ref.: 2006)										
2007	0,0446**	0,0167	0,0704***	0,0126	0,0135	0,0127	0,0468**	0,0144	0,0276	0,0163
2008	0,0656***	0,0164	0,1001***	0,0124	0,0771***	0,0126	0,0732***	0,0140	0,0652***	0,0160
2009	0,1215***	0,0165	0,1033***	0,0123	0,0945***	0,0123	0,0719***	0,0140	0,0991***	0,0159
2010	0,0806***	0,0161	0,1086***	0,0124	0,0906***	0,0124	0,0897***	0,0139	0,0994***	0,0157
2011	0,0879***	0,0164	0,1128***	0,0126	0,0812***	0,0126	0,0863***	0,0141	0,1163***	0,0161
GPA Written Danish exam	0,0105***	0,0024	0,0058***	0,0018	0,0091***	0,0018	0,0073***	0,0019	0,0070**	0,0022
Written math exam	0,0157***	0,0023	0,0217***	0,0017	0,0204***	0,0017	0,0282***	0,0018	0,0212***	0,0021
Family specific covariates										
Family type (Ref.: Nuclear family)										
Divorced with new partner	-0,0255	0,0193	-0,0365*	0,0143	-0,0394**	0,0138	-0,0316*	0,0135	-0,0536***	0,0157
Single parent	0,0034	0,0153	-0,0468***	0,0116	-0,0154	0,0114	-0,0393***	0,0108	-0,0199	0,0135
Not living with parents	-0,0481***	0,0134	-0,0447***	0,0101	-0,0249*	0,0104	-0,0320*	0,0125	-0,0435***	0,0130
Mother missing	0,0374	0,0439	0,0390	0,0335	0,0518	0,0351	0,0918**	0,0353	0,0469	0,0384
Father missing	0,0117	0,0293	0,0530*	0,0211	-0,0063	0,0222	0,0529**	0,0202	0,0297	0,0248
Household income (Log)	0,0521**	0,0175	0,0509***	0,0129	0,0221	0,0131	0,0477***	0,0129	0,0189	0,0160
Intercept	-0,0998	0,2274	-0,1574	0,1670	0,2564	0,1696	-0,1771	0,1683	0,3535	0,2080
R2	0,071		0,067		0,071		0,079		0,077	
N	8.460		16.159		15.408		13.106		9.868	

*** p<0.001, ** p<0.01, * p<0.05

7. Robustness

This section presents a small selection of the extensive robustness analysis we conducted. The first section presents the main results split by graduation year and the second demonstrates the robustness check regarding the specification of the distance variable.

7.1. Participation in higher education by year of graduation

The share of high school graduates who continue in higher education has increased within the last decade. From 2006 to 2011 the share that enrolled in higher education within 2 years of high school graduation increased from 67% to about 77%. There may be several explanations for this tendency. One could be that the financial crisis in 2008 led to an increase in youth unemployment, which encouraged more students to

enrol in higher education (cf. Statistics Denmark)[12]. If this were the case, you would expect an increase in enrolment in the provinces, which were the most vulnerable to the financial crisis and where youth unemployment rates rose most significantly.

Another explanation could be that the admission requirements for higher education changed significantly in 2008. In addition, new legislation on a faster start to study came into effect in 2009. According to the law, students who applied for admission to a higher education institution no later than 2 years after graduation, could multiply their grade point average by 1,08, providing an incentive for students to pursue a higher education more rapidly after graduation. Based on these explanations, it is relevant to examine whether the result from Table 3 showing no relationship between distance and enrolment rates holds true for the whole period.

Table 6 shows the distance parameter estimate, by year of graduation. Each year, three models are estimated: Model 0, with no control variable, Model 1, adjusted for individual specific covariates and Model 2; adjusted for both individual and parental covariates. The parameter estimates from Model 2, demonstrate that the relationship between distance and higher education enrolment were more profound in the beginning of the period. In 2006 and 2007 the relationship between distance and enrolment remains statistically significant after controlling for individual and family specific covariates. This could be explained by the fact, that the enrolment rates have increased the most in municipalities in peripheral areas where the distance to above average.

Table 6: Parameter estimates: Participation in higher education by year of graduation

	Year 2006		Year 2007		Year 2008		Year 2009		Year 2010		Year 2011	
	B	se	B	se	B	se	B	se	B	se	B	se
Model 0: Unadjusted												
Distance (km)	-0,0016***	0,0002	-0,0013***	0,0002	-0,0014***	0,0002	-0,0004*	0,0002	-0,0007***	0,0002	-0,0004*	0,0002
Model 1: Adjusted for individual specific covariates												
Distance (km)	-0,0010***	0,0002	-0,0010***	0,0003	-0,0009***	0,0002	-0,0001	0,0002	-0,0004*	0,0002	-0,0002	0,0002
Model 2: Adjusted for indi. and parental covariates												
Distance (km)	-0,0005*	0,0002	-0,0005*	0,0003	-0,0006*	0,0002	0,0003	0,0002	-0,0001	0,0002	0,0001	0,0002
N	30.417		26.922		28.326		30.559		31.422		31.222	

*** p<0.001, ** p<0.01, * p<0.05

7.2. Different specifications of distance

To allow for non-linearities in the effect of distance to the nearest higher education institution, we have tried other specifications of the distance measure. Table 7 shows the parameter estimates of the distance measure in the final model (model 2) in four different specifications. Column 1 gives the continuous distance measure in kilometres, column 2 the natural log of the distance[13], column 3 a set of dummy variables, indicating the deciles of the distribution of the distance variable, and column 4 a set of dummies for each distance interval. The table presents the F-statistics and R^2-values for each model. The specification of the distance measure did not change the overall conclusion – that the observed relationship between distance and enrolment rate was insignificant when controlling for relevant covariates. Further, the F-statistic was highest for the continuous distance measures. In other words, a linear functional form of the distance measure is appropriate to capture the effect of distance.

[12] The unemployment rate among 15-24 years old increased from 7.5% to 14.2% from 2007 – 2011 cf. numbers from Statistics Denmark.

[13] Distance measure + 1.

Table 7: Parameter estimates: different specifications of distance in the final model

	Distance in km		Ln (distance)		Distance in deciles		Distance in categories	
	B	se	B	se	B	se	B	se
Continuous distance measure								
Distance in kilometers	-0,0002	0,0001						
Ln (distance + 1)			0,0003	0,0011				
Distance in deciles								
0-1 km (ref.)								
2 km					0,0068	0,0044		
3-4 km					0,0089*	0,0039		
5-6 km					0,0086	0,0045		
7-8 km					0,0101*	0,0047		
9-12 km					0,0141***	0,0042		
13-16 km					0,0046	0,0042		
17-21 km					0,0099*	0,0043		
22-29 km					0,0003	0,0045		
30-67 km					0,0002	0,0045		
Distance in categories								
0-5 km (ref.)								
6-10 km							0,0078*	0,0030
11-15 km							0,0032	0,0032
16-20 km							0,0018	0,0035
21-25 km							-0,0081*	0,0041
26-30 km							-0,0005	0,0054
31-40 km							-0,0048	0,0047
41-50 km							-0,0091	0,0071
51-60 km							-0,0186	0,0111
61-67 km							0,0110	0,0459
R2	0,0767		0,0767		0,0768		0,0768	
F	390,8996		390,8101		323,3317		323,3075	
N	178.868,000		178.868,000		178.868,000		178.868,000	

*** p<0.001, ** p<0.01, * p<0.05

8. Discussion and conclusions

This study assessed the role of distance to higher education institutions in Denmark, in the decision to enrol in higher education after high school. Policy makers often assume that the spatial distribution and access to higher education affects the rate of enrolment and graduation (Vestergaard, 2015). The international literature on the deterrence effects of distance on education decision is inconclusive. The reported effects range from small, negative effects (Frenette, 2006; Öckert, 2012; Spiess & Wrohlich, 2010) to no effect (Cullinan et al., 2013; Sá et al., 2006). Moreover, country-specific differences such as variation in the location of higher educations, the general access to the educational system, and the generous provisions of the welfare system may hinder the application of international evidence in a Danish context.

The descriptive results demonstrated that the majority of high school students in Denmark had access to higher education within commuting distance, when they graduated from high school: Over 90% are less than 30 km from the nearest educational institution and 35% of the students live within 5 km. However, those living closer to an education institution are still more likely to attend a higher education. We used a linear probability model to investigate if distance to higher education institutions has an effect on

enrolment in higher educations within 2 years after graduation. The econometric results showed that, on average, there is no effect of distance on enrolment in higher education. The revealed negative association between distance and enrolment can be explained by residential sorting of the parents.

We also examined whether the effect of distance varied across parental education and found small but significant relationship between distance and enrolment among student, for whom the highest education of both parents were a vocational education. The additional analyses revealed that this association were only present in Central Jutland Region and Southern Denmark Region. Therefore we conclude that students who do not have access to a local university or college are as likely to enrol in a higher education, as students who grow up near a university or college, when taking parental residential sorting into account.

The results have important implications for discussions surrounding access to higher education in outlying areas[14]. The current study has shown that there is no empirical evidence that distance to higher education institutions acts as a barrier to enrolment in Denmark. Instead, we see that parental education continues to play a key role when it comes to educational choice. The results suggest that increasing the capacity of higher education in outlying areas is not an efficient tool to increase enrolment rates in higher institutions among high school students. Future work could investigate the effect of distance on educational choices earlier in life, for example the effect of distance to upper secondary educations and vocational secondary educations.

[14] A potential shortcoming of the present study is, that is it based on observational data. Several studies point to the fact that, there is a risk of endogeneity issues as families who choose to live close to educational institutions may be quite different from those who live far away, on central variables that we cannot observe.

References

Acemoglu, D., & Pischke, J. S. (2001). Changes in the wage structure, familiy income and children's education. *European Economic Review*, *45*, 890–904.

Becker, G. S. (1964). *Human capital: a theoretical and empirical analysis, with special reference to education*. New York: National Bureau of Economic Research; distributed by Columbia University Press.

Bishop, J. (1977). The Effect of Public Policies on the Demand for Higher Education. *The Journal of Human Resources*, *12*(3), 285–307.

Black, S. E., & Sufi, A. (2002). *Who Goes to College? Differential Enrollment by Race and Family Background* (Working Paper No. 9310). National Bureau of Economic Research.

Card, D. (1993). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. *NBER Working Paper Series*, (4483).

Cullinan, J., Flannery, D., Walsh, S., & McCoy, S. (2013). Distance Effects, Social Class and the Decision to Participate in Higher Education in Ireland. *The Economic and Social Review*, *44*(1), 19–51.

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

Denzler, S., & Wolter C., S. (2011). Too Far to Go? Doese Distance Determine Study Choices? *The Institute for the Study of Labor (IZA)Discussion Paper Series*, (5712).

Frenette, M. (2004). Access to College and University: Does Distance to School Matter? *Canadian Public Policy*, *30*(4), 427–443.

Frenette, M. (2006). Too Far to Go On? Distance to School and University Participation. *Education Economics*, *14*(1), 31–58.

Frenette, M. (2009). Do universities benefit local youth? Evidence from the creation of new universities. *Economics of Education Review*, *28*, 318–328.

Fuller, W. C., Manski, C. F., & Wise, D. A. (1982). New Evidence on the Economic Determinants of Postsecondary Schooling Choices. *The Journal of Human Resources*, *17*(4), 477–498.

Gibbons, S., & Silva, O. (2008). Urban density and pupil attainment. *Journal of Urban Economics*, *63*, 631–650.

Gibbons, S., & Vignoles, A. (2009). *Access, choice and participation in higher education*. London School of Economics and Political Science.

Griffith, A. L., & Rothstein, D. S. (2009). Can't Get There from Here: The Decision to Apply to a Selective College. *Economics of Education Review*, *28*(5), 620–628.

Holzer, S. (2007). *The Expansion of Higher Education in Sweden and the Issue of Equality of Opportunity* (CAFO Working Paper No. 2007:5). Centre for Labour Market Policy Research (CAFO), School of Business and Economics, Linnaeus University.

Høst, A. K., & Sørensen, E. S. (2015). *Unges afstand til de videregående uddannelser. Hvor langt har de unge i Danmark til en videregående uddannelse? Og er der en sammenhæng mellem afstanden og deres uddannelsesvalg?* København: Kraks Fond Byforskning.

Kane, T. J., & Rouse, C. E. (1995). Labor-Market Returns to Two- and Four-Year College. *The American Economic Review*, *85*(3), 600–614.

Nguyen, A. N., & Taylor, J. (2003). Post-high school choices: New evidence from a multinomial logit model. *Journal of Population Economics*, *16*(2), 287–306.

Öckert, B. (2012). On the margin of succes? Effects of expanding higher education for marginal students. *Nordic Economic Policy Review*, (1), 111–155.

Oppedisano, V. (2011). The (adverse) effects of expanding higher education: Evidence from Italy. *Economics of Education Review*, *30*(5), 997–1008.

Sá, C., Florax, J. G. M., & Rietveld, P. (2006). Does Accessibility to Higher Education Matter? Choice Behaviour of High School Graduates in the Netherlands. *Spatial Economic Analysis*, *1*(2), 155–174.

Shea, J. (2000). Does parents' money matter? *Journal of Public Economics*, *77*(2), 155–184.

Spiess, K. C., & Wrohlich, K. (2010). Does distance determine who attends a university in Germany? *Economics of Education Review*, *29*, 470–479.

Vestergaard, M. (2015, August 2). Nye uddannelser kan være på vej i Udkantsdanmark. *Jyllands-Posten*. Retrieved from http://jyllands-posten.dk/politik/ECE7903292/Nye-uddannelser-kan-være-på-vej-i-Udkantsdanmark/

Appendix

Figure A1: Distance to nearest higher education institution in 1-km intervals

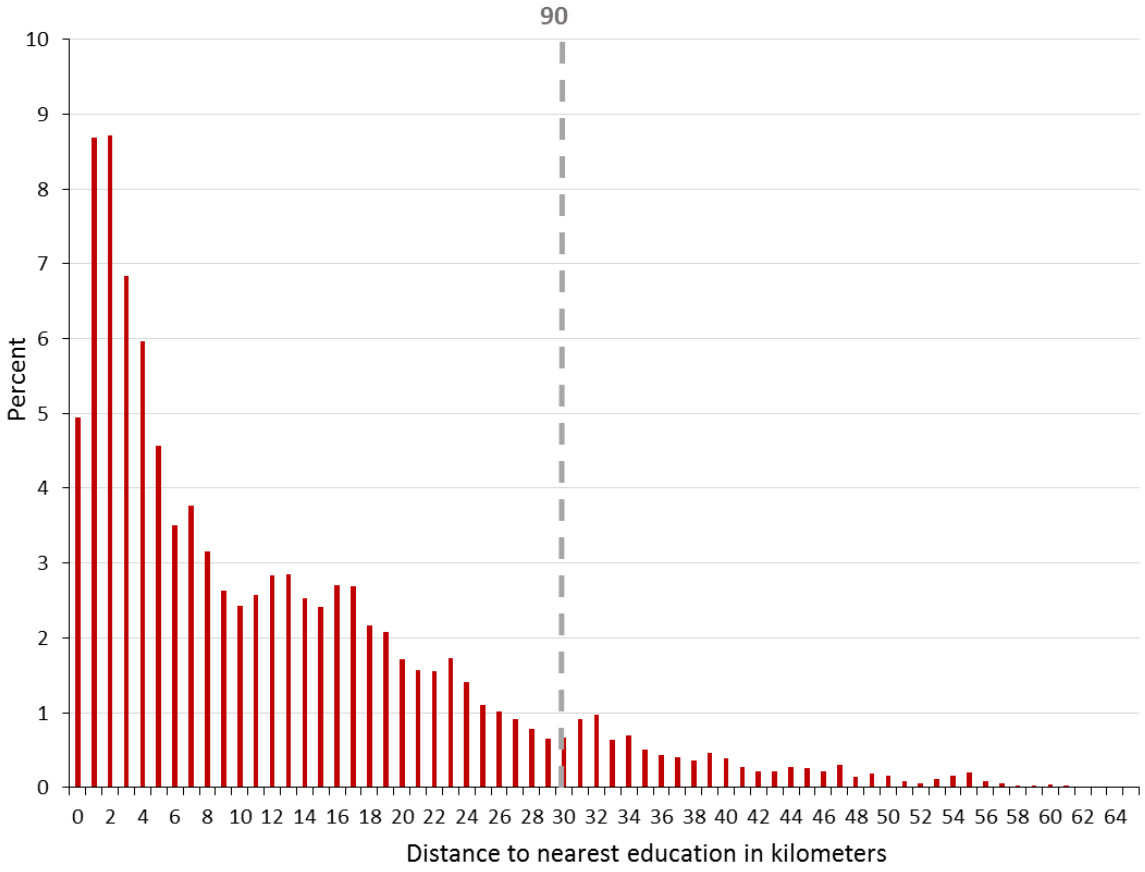


Figure A2: Distance to nearest higher education institution by parental education

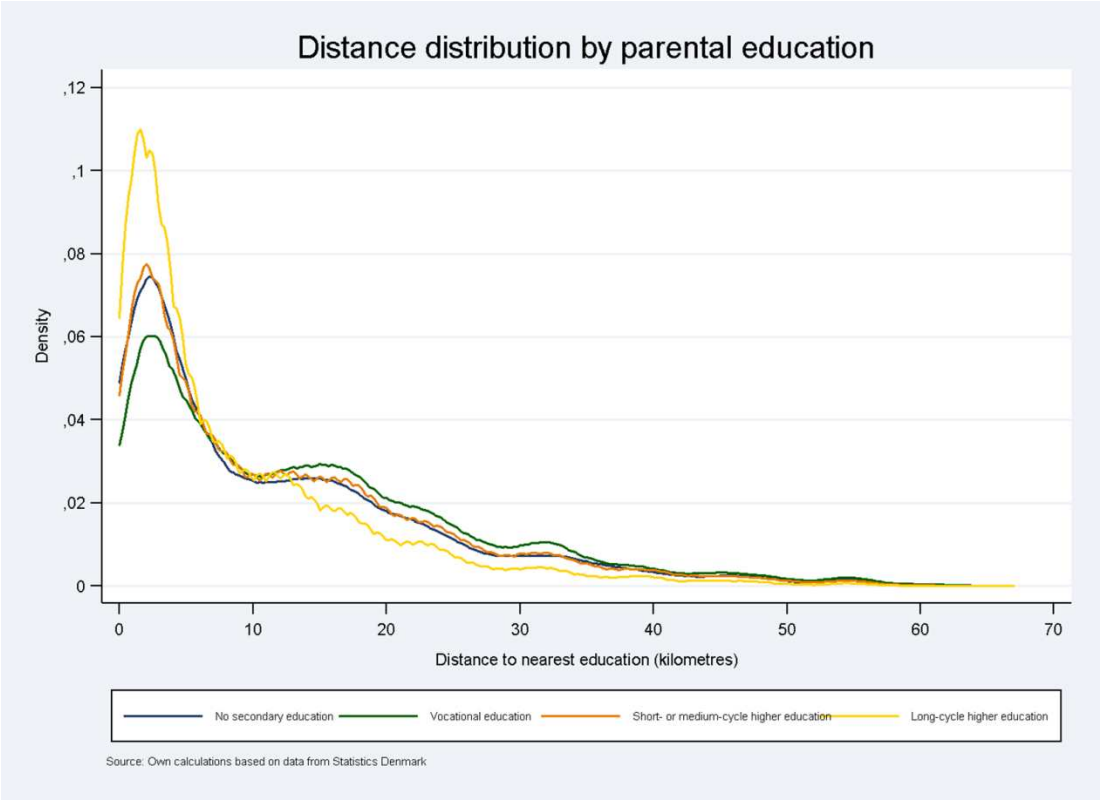


Figure A3: Distance to nearest higher education institution by geographical region