Human Capital, Migration and Regional Inequality

How the location of education institutions influences our society

Lange, Elise Stenholt

UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

Human Capital, Migration and Regional Inequality

HOW THE LOCATION OF EDUCATIONAL INSTITUTIONS INFLUENCES OUR SOCIETY

ELISE STENHOLT SØRENSEN
Department of Geosciences and Natural Resource Management

Human Capital, Migration and Regional Inequality

How the Location of Educational Institutions Influences our Society

ELISE STENHOLT SØRENSEN

Human Capital, Migration and Regional Inequality

PhD thesis
Elise Stenholt Sørensen
April 2021

Principal academic supervisor:
Professor Lars Winther, Department of Geosciences and Natural Resource Management, Faculty of Science, University of Copenhagen

University of Copenhagen

Professor Alessandra Faggian, Gran Sasso Science Institute – Italy
Associate Professor Mika Haapanen, University of Jyväskylä – Finland
Associate Professor Høgni Kalsø Hansen (chair), IGN

Table of Contents

List of Papers

This PhD thesis consists of the following four papers:

Paper 1. Youth Participation in Education: The Impact of Distance to School and Parental Education

Paper 2. Moving out of Rural Areas: Estimating the Causal Effect of General Upper Secondary Education on Youth Migration

Paper 3. Gaps in Higher Education Choice and Participation: Understanding the Role of Distance to Higher Education

Paper 4. Regional Graduate Retention Trajectories in Denmark: Education, Location and City Size

Acknowledgements

The writing of my PhD thesis has been an educational journey, a long process of leaning, reflecting, doubting and widening my view of the field of human capital and inter-regional migration. I am grateful for having had the opportunity to work and focus on specific topics in which I have been interested for many years, questions about the formation of cities, what shapes human choices over the life-cycle and the role of education institutions in our society. Both personally and professionally, these last five years have been enormously formative and educational. The biggest lesson has been that no one creates anything on his or her own: everything that we create is made with the help, support and interaction of other people. Although we tend to value individual performance and accomplishment in society today, we tend to forget all those who have stood behind the accomplishment. Although this thesis is the result of many hours of focused work spent in solitude, it is also the outcome of inspiring conversations, comments and interactions with many different people, to whom I owe my greatest thanks.

First, I would like to thank Lars Pico Geerdsen for encouraging me to pursue a PhD and giving me the opportunity to do so. I am grateful for the many formative discussions we had during our weekly running-club at the Kraks Fond. Lars Pico Geerdsen has been a great source of inspiration that showed me how academic scientific research and studies that are relevant and useful for society can go hand in hand. This thesis was financed by the Kraks Fond, to which I am grateful for making the funding available and investing in important knowledge within the field of urban economics. In this connection, I owe my greatest thanks to the former director of Kraks Fond, Stig Christensen, who had the vision and courage to establish a research centre in this field and make this PhD thesis possible.

I would also like to thank my supervisor, Professor Lars Winther, for kind and competent guidance and many inspirational discussions. I also thank the research group Transformation of Cities and Landscapes - Geoinformatics at the Section of Geography, in the Department of Geosciences and Natural Resource Management, University of Copenhagen, for valuable comments, and for widening my perspectives on economic geography. In addition, I owe thanks to my colleagues in the research groups Transformation of Housing and Places, BUILD, Aalborg University, for helpful comments and personal

support. In particular, Hans Thor Andersen read early drafts and was generous with insightful comments and constructive criticism.

I would also like to thank my former colleagues at the Kraks Fond, for their helpful comments and encouragement and for creating an open, creative and fun working environment. I owe a special thanks to Cecilie Weatherall for helping me plan and organise the timeline of my PhD and to Bence Boje-Kovacs for reading several of my papers and providing me with constructive comments. I would also like to thank my co-author Anders Holm for collaboration and inspiring conversations in conducting the research for Paper 2 of the thesis.

The statistical analysis in the thesis is based on a large amount of composite data sources maintained by Statistics Denmark. I therefore owe a big thank you to our helpful contact person at Statistics Denmark, Thomas Lauterbach, for providing us with access to data sources. I would also like to thank my manuscript editors Robert Parkin and Mark Harvey Simpson for their careful editing and useful suggestions. Their editorial help has lifted the reading of the papers substantially.

On the personal level, I am grateful to physiotherapist Lisbeth Ebbesen for giving me the hope and faith that I could return to work on my PhD thesis after a bicycle-accident left me with a brain concussion. I would like to thank my friend and wonderful neighbour Ida Lawrence for creating the front cover of the thesis and some of the illustrations in the thesis. In addition, she has given me countless support in providing dinner and childcare when it was needed the most. I would also like to thank my friend Louise Vang Jensen for helping me structure and highlight the main conclusion in the introduction to the thesis. I am also thankful to my friend Nanna Nielsen for editing and proof-reading the Danish summary of the thesis. In addition, my friend Anne Marie Boesen provided unreasonable optimism and inspiring professional discussions about how research could be applied to society.

I owe debts to my parents, Jannie and Niels, and my brother Gustav for their endless encouragement and understanding. Their support was priceless during the last couple of months. I am also grateful to my son, Sebastian, for constantly reminding me of the importance of everyday life besides my PhD project. Above all, I am grateful to my husband Peter for his love, patience and endless sparring and counselling. He took on the role of my "domestiques rider" in this long race and helped me cross the finishing line.

Dansk sammenfatning

Selv i et lille land som Danmark er der store regionale forskelle hvad angår økonomisk vækst og udvikling. Økonomiske forskelle mellem land og by og mellem regioner har altid eksisteret. Men i løbet af det seneste årti er de regionale forskelle i økonomisk udvikling blevet større. Denne tendens til øget regional ulighed er ikke kun et dansk fænomen, men et fænomen som har optaget forskere, økonomer, regionale planlæggere og politikere over hele den vestlige verden i de seneste år.

Blandt toneangivende forskere inden for byøkonomi og regional udvikling er der enighed om, at en af de helt centrale forklaringer på den øgede regionale ulighed skal findes i vores samfunds overgang fra industrisamfund til videnssamfund. I takt med at den globale konkurrence er øget, har flere virksomheder fået større fokus på innovation og højt specialiserede produkter. Det betyder, at arbejdsstyrkens kompetencer og uddannelsesniveau er blevet en endnu vigtigere faktor for skabelsen af produktivitet og økonomisk vækst. Befolkningens uddannelsesniveau, den såkaldte humane kapital, er altså helt afgørende for både den nationale og den regionale udvikling i dag.

I forskningslitteraturen har det ansporet til en øget interesse for at undersøge sammenhængen mellem høj uddannelse og regional udvikling. En række centrale studier har vist, at der både er positive individuelle og samfundsmæssige effekter forbundet med et højt uddannelsesniveau i befolkningen. En stor andel af personer med høj uddannelse i en by eller i en virksomhed har fx den positive effekt, at de også gør de andre personer i virksomheden mere produktive og innovative via vidensdeling osv. Desuden er høj uddannelse ofte forbundet med lavere kriminalitet, højere sundhed, højere politisk engagement osv. på et samfundsmæssigt plan. Der er derfor mange gode grunde til, at regioner i stigende grad ønsker at øge befolkningens uddannelsesniveau eller tiltrække højtuddannede fra andre regioner.

Dansk kontekst

I en dansk politisk kontekst er der også stor fokus på at løfte befolkningens uddannelsesniveau og tiltrække højtuddannet arbejdskraft til alle egne af landet som et led i at skabe en mere lige regional udvikling. Den regionale placering af offentlige arbejdspladser, især placeringen af uddannelsesinstitutioner, er blevet et varmt politisk emne i Danmark i de seneste år. Med formålet at skabe "et land i balance" har den nuværende regering foreslået, at der skal flere

uddannelsesmuligheder uden for de større byer, og at de nuværende videregående uddannelser skal decentraliseres.

Der er flere argumenter for at have fokus på lokalisering af ungdomsuddannelser og videregående uddannelsessteder. Et centralt argument er argumentet om, at alle i befolkningen skal have lige adgang til uddannelsesmuligheder, uanset hvor i landet man vokser op. Hvis nogen afskæres muligheden for uddannelse, fordi afstanden er for lang, er det både et socialt og samfundsmæssigt problem, at de menneskelige ressourcer ikke bliver udnyttet til fulde. Et andet centralt argument handler om, at lokalisering af uddannelsesinstitutioner kan påvirke bosætning og regionale flyttemønstre. Mange unge flytter i forbindelse med uddannelsesstart, og derfor kan uddannelsesstedernes placering måske have betydning for, hvor de flytter hen.

Der mangler et solidt vidensgrundlag og empirisk evidens for at kunne vurdere om decentralisering af uddannelsesinstitutioner i Danmark vil have den ønskede effekt. Det er uklart, hvorvidt oprettelsen af flere uddannelsestilbud rundt om i landet vil få flere unge i gang med en ungdomsuddannelse? Det er også uklart, hvorvidt de nyuddannede unge vil blive boende efter endt uddannelse eller i højere grad vælger at flytte til en af de større byer? Hvis det sidste er tilfældet, så vil det øgede uddannelsesniveau ikke komme den lokale uddannelsesby til gode.

Bidrag og indhold

Indtil nu har forskningslitteraturen primært fokuseret på de højtuddannede personer. En lang række studier har undersøgt, hvorvidt afstand til universiteter spiller en rolle for, om unge søger ind på universitet. Desuden har en stor gren af litteraturen beskæftiget sig med højtuddannedes regionale flyttemønstre og hvilke regionale faktorer, der er afgørende for at tiltrække og fastholde højtuddannet arbejdskraft. Studier har blandt andet vist, at højtuddannede er langt mere geografisk mobile end andre befolkningsgrupper. Længere uddannelse er med til at øge ens muligheder og incitament for at flytte over længere geografisk afstand. Det er dog uklart, i hvilken grad dette også gælder for unges første uddannelsesvalg – nemlig deres valg af ungdomsuddannelse. Denne ph.d.-afhandling bidrager til den internationale forskningslitteratur ved at undersøge, om unges valg af ungdomsuddannelser har en betydning for deres senere første flyt hjemmefra.

Disse spørgsmål har motiveret mig til at skrive denne ph.d.-afhandling. Formålet med afhandlingen er at skabe et bedre vidensgrundlag for at kunne

vurdere, hvorvidt decentralisering af ungdomsuddannelser og videregående uddannelser er et godt værktøj til at skabe bedre regional udvikling i hele landet? Og hvorvidt skabelsen af flere uddannelsestilbud kan have utilsigtede effekter på den regionale ulighed.

Derfor er jeg kommet frem til følgende forskningsspørgsmål: Hvilken betydning har den geografiske placering af vores uddannelsesinstitutioner for unges uddannelsesvalg? Er lange afstande med til at afholde nogle unge fra at tage en ungdomsuddannelse? Hvilken betydning har uddannelse for bosætningsmønstre og urbanisering i Danmark? Og hvilken sammenhæng er der mellem placeringen af de videregående uddannelser og andelen af nyuddannede, der bliver boende i uddannelsesbyen efter endt uddannelse? Disse spørgsmål er centrale for at forstå, hvordan den geografiske placering af uddannelsesinstitutioner påvirker vores samfund, og de vil blive grundigt behandlet i afhandlingens fire artikler.

Undersøgelserne er baseret på registerdata fra Danmarks Statistik samt omfattende afstandsberegninger fra bopælsadresser til samtlige ungdomsuddannelser og videregående uddannelser i Danmark. Alle analyser anvender data med detaljerede oplysninger på individniveau. Ved hjælp af avancerede statistiske metoder og kausale analysedesign undersøger studierne betydningen af nærhed til uddannelse for uddannelsesvalg, og uddannelsesvalgets betydning for regional mobilitet, samtidig med at modellerne tager højde for en række centrale faktorer som fx individuelle karakteristika, forældrebaggrund og geografi.

Formål

Formålet med denne ph.d.-afhandling er at undersøge, hvilken betydning den geografiske placering af uddannelsesinstitutioner har for befolkningens uddannelsesvalg, deres bosætningsmønstre og den regionale udvikling i Danmark. Afhandlingen er delt op i to overordnede dele. Den første del, artikel 1 og 2, undersøger spørgsmål, der omhandler geografisk placering af ungdomsuddannelser. Den anden del, artikel 3 og 4, kigger nærmere på betydningen af de videregående uddannelsers regionale placering.

Artikel 1 undersøger om afstand fra bopæl til det nærmeste gymnasium har en betydning for unges uddannelsesvalg. Indledningsvist kortlægges den geografiske placering af gymnasier i dag, og hvor langt de unge har til en gymnasieuddannelse. Resultaterne viser, at danske gymnasier har en god

geografisk fordeling, og de færreste unge oplever at have langt til et gymnasieudbud. Analyserne viser også, at afstanden fra bopæl til det nærmeste gymnasium har en betydning for optag på en gymnasieuddannelse blandt nogle unge. Betydningen af afstand varierer efter familiebaggrund, således at unge fra lavt uddannede familier er mest påvirkede af afstand, når de skal vælge, om de vil fortsætte i en ungdomsuddannelse efter folkeskolen. Overordnet set er effekterne af afstand for uddannelsesoptag dog relativt små i størrelse sammenlignet med eksempelvis betydningen af karakterer eller forældrenes uddannelsesniveau.

Artikel 2 fokuserer på årsagerne til migration fra land til by blandt unge mellem 18 – 24 år. Analysen viser, at unges valg af ungdomsuddannelse spiller en afgørende rolle for deres flyttemønstre. En gymnasial ungdomsuddannelse øger tilbøjeligheden til at flytte til en større by markant. På et samfundsmæssigt plan betyder det, at når flere unge vælger at gå i gymnasiet, så har det betydning for den øgede urbanisering, som vi ser i dag.

Artikel 3 studerer unges afstand til de videregående uddannelser samt betydningen af afstand til uddannelse for deres uddannelsesvalg. Analysen omfatter unge, der har fuldført en gymnasial ungdomsuddannelse. Resultaterne peger på, at afstanden til de videregående uddannelser ikke spiller en afgørende rolle for, hvorvidt unge påbegynder en videregående uddannelse efter gymnasiet. Til gengæld peger analysen på, at afstand til det nærmeste universitet kan være afgørende for, hvorvidt unge vælger en universitetsuddannelse eller en kortere videregående uddannelse. Hvis afstand til universitetet bliver lang, er unge fra lavtuddannede hjem mere tilbøjelige til at vælge en kortere videregående uddannelse.

Artikel 4 analyserer, i hvilken grad uddannelsesbyer med videregående uddannelser formår at fastholde deres højtuddannede dimittender i byen og regionen efter endt uddannelse. Artiklen fokuserer på nyuddannede dimittender fra alle videregående uddannelser i Danmark og analyserer deres flyttemønstre, fra de bliver færdiguddannede og op til 10 år efter. Artiklen undersøger hvilke faktorer, der har betydning for fastholdelse af dimittender og konkluderer, at uddannelsesbyens størrelse, regionale placering og det lokale arbejdsmarked er afgørende for, hvorvidt dimittenderne bliver boende i uddannelsesbyen.

Konklusion

For at opsummere så kan ph.d.-afhandlingens resultater samles i tre overordnede konklusioner. For det første viser analyserne af afstands betydning for uddannelsesvalg, at afstanden fra bopæl til ungdomsuddannelse har en betydning blandt de 15-årige, hvorimod afstand til de videregående uddannelser har en begrænset rolle for, hvorvidt unge vælger at læse videre efter gymnasiet eller ej. Analyserne peger derfor på, at afstand til uddannelse har en betydning for unges uddannelsesvalg, når de forlader folkeskolen, hvorimod unge, der har fuldført gymnasiet, er villige til at pendle eller flytte for at tage en videregående uddannelse.

For det andet viser resultaterne, at størstedelen af de nyuddannede unge, der har fuldført en gymnasial ungdomsuddannelse, flytter væk fra deres hjemby og til en af Danmarks største byer. Det samme gælder ikke for de unge, der har fuldført en erhvervsfaglig ungdomsuddannelse, eller de unge der ikke opnår en ungdomsuddannelse. Dermed bekræfter analysen, hvad internationale studier har vist: At uddannelse åbner for nye muligheder og øger individers geografiske mobilitet. Unge forlader altså ikke deres hjemby på grund af *manglende* uddannelsesmuligheder lokalt - de fraflytter netop *på grund af* de muligheder, som de lokale uddannelser har givet dem.

For det tredje viser analysen af de højtuddannedes flyttemønstre, at uddannelsesinstitutionens geografiske placering er afgørende. I de mindre uddannelsesbyer er det et fåtal af de nyuddannede, der bliver boende efter endt uddannelse. Det tyder på, at en eventuel decentralisering af de videregående uddannelsessteder ikke vil øge det lokale uddannelsesniveau i arbejdsstyrken betragteligt, da langt de flest dimittender vil flytte væk igen efter endt uddannelse.

Abstract

As regional inequalities in economic prosperity have increased in Denmark in recent years, the political tools for slowing down this development are taking up space in public debates. The geographical location of public institutions, in particular educational institutions, has become a political hot topic in recent years in Denmark. In order to address the regional imbalance, the current Danish government has suggested that geographical access to higher education should be improved by creating more educational opportunities across the country.

There are several arguments for focusing on the decentralisation of public educational institutions. One central argument concerns the right to equal opportunities for all citizens, including the opportunity to harness human resources. If distance to education acts as a barrier to participation in education, it becomes both a social problem and a failure to utilize a source of human capital. Another central argument is that the location of educational institutions might affect residential choice and inter-regional migration.

However, Denmark lacks the empirical evidence to inform these arguments. Hence, it is uncertain how the creation of more educational opportunities will influence enrolment in education programmes and inter-regional migration. The aim of this thesis is to obtain insights into the impact of geographical access to educational institutions. By focussing on two central educational choices for young people, choice of youth education and choice of higher education, this thesis provides new evidence for the associations between geographical proximity to education, educational choices and migration during the life-cycle of the individual.

The main contribution of this thesis is therefore that it provides new knowledge about how the initial choice of youth education influences young people's decision-making with respect to migration. Research has primarily focused on highly educated individuals and the determinants and consequences of the spatial mobility of recent university graduates. However, there is a gap in the literature concerning what shapes those highly educated individuals' initial choices on their educational trajectories. Evidence about the determinants of educational choices and migration by young people is needed to acquire a fuller picture of the sorting and divergence of regional human capital.

In this thesis I use quantitative, statistical methods to examine the research questions. The analysis in the papers is based on longitudinal data from the administrative population register of Statistics Denmark, combined with detailed distance measures between place of residence and educational institutions in Denmark. All analyses are conducted using micro-level data in which individuals are the unit of analysis. The thesis is structured into five introductory chapters and four academic papers. The four papers cover different perspectives of the educational paths of individuals. Papers 1 and 2 centre on the role of upper secondary education, while papers 3 and 4 examine higher education institutions.

Paper 1 studies to what extent distance from home to general upper secondary education institutions influences participation. In addition, the paper examines to what extend young individuals from disadvantaged backgrounds are more affected by geographical constraints. Using data on the total population of 15-year-olds in Denmark, from 1982 – 2018, the study provides causal evidence of the effect of distance to school. The results suggests that the influence of this distance is most prevalent among young people from lower-educated families, and the influence diminishes if one of the parents holds a university degree.

Paper 2 examines the effect of completion of a general upper secondary education programme on young individuals' migration patterns. The paper focuses on migration out of rural areas among young people between 18 and 24 years old. The results suggest that completion of a general upper secondary education significantly increases the probability that a person will move out of a rural area.

Paper 3 considers the role of distance to higher education in young people's decision-making about their educational paths. The study focuses on young people who have graduated from a general upper secondary education programme. Overall, the results suggests that, in the current educational landscape in Denmark, distance to higher education plays no significant role in whether young people decide to enrol in higher education or not.

Paper 4 provides descriptive evidence of important factors for the retention of graduates in the region where their higher education institution is located. The paper focuses on newly educated graduates and analyses their migration decisions up until 10 years after graduation. The findings point to the importance of regional location and the local labour market for the retention of highly educated graduates.

In this thesis, I paint a broader picture of the impact of geographical access to educational institutions than has previously been available for Denmark. Overall, this thesis contributes to the debate on the role of local educational institutions in regional development.

Chapters

CHAPTER 1

Introduction

This PhD project is motivated by the increasing regional divide in economic performance that has been witnessed in Denmark during the past decade (Sørensen, 2021)[1]. Economic disparity across space is not solely a Danish phenomenon, it is also a global phenomenon that has concerned politicians, policy makers, economists and urban scientists across the Western hemisphere in recent years. In the US, a study of the historical development of long-term unemployment shows that regional disparities in joblessness are large and have become "semi-permanent" in the past decade (Austin, Glaeser, & Summers, 2018). A similar development is found in an analysis of regional development in the European Union (EU), which shows that inter-regional inequality within the EU has grown and finds that there are several different modes of regional economic performance that respond to different development challenges and opportunities in Europe's regions (Iammarino, Rodriguez-Pose, & Storper, 2019).

Regional differences in economic prosperity have existed in the past. During most of the 20th century, economists predicted that regional economic disparities would converge, as the poorer regions were considered to have high growth rates and were therefore expected to catch up with the wealthiest regions (Austin et al., 2018). However, the developments in regional inequality in the past decade have flipped this picture. Economic development between regions has stopped converging and has begun to diverge (Eliasson, Haapanen, & Westerlund, 2020).

This development has spurred a growing interest in the determinants of the regional divide and also a concern about its potential consequences for society. In the US, Austin et al. have voiced their concern that the spatial economic divisions seen in the US "loom as the backdrop to our political divisions" (Austin et al., 2018, p. 152). In the EU context, Iammarino et al. write that "regional economic divergence has become a threat to economic progress, social cohesion and political stability in Europe." (Iammarino et al., 2019, p. 1).

[1] Chapter 4 provides an overview of the development in regional growth in Denmark over the past 20 years, and a description of the institutional and political context of the regional divide.

In Denmark, regional divergence is also a public and political concern. The incumbent Danish Prime Minister Mette Frederiksen expressed her concerns about social cohesion across the country in her New Year's speech to the nation on 1 January 2021. She argued that "opportunities have become more unevenly distributed"[2] across Denmark and that the government would work to reduce regional inequalities (Frederiksen, 2021).

1.1 The Danish context

In the Danish political context, we have seen a change in political focus from socially stratified policies to place-based policies. Especially two central events have sparked this change in policy focus. Firstly, the global financial crisis in 2007 impacted large manufacturing companies and agriculture in Denmark in particular. This meant areas located outside the largest urban regions lost many workplaces as employment in these areas was more reliant on large manufacturing companies and agriculture (Larsen et al., 2014). Secondly, a range of large-scale reforms in the organisation of local and regional government were put into effect on 1 January 2007, which resulted in the loss of many public workplaces in rural and peripheral areas (Andersen & Etzerodt, 2018). Overall, researchers agree that these extensive reforms amplified the economic recession in rural and peripheral areas that had come in the wake of the financial crises (Andersen & Etzerodt, 2018).

A strong public and political discourse has unfolded about public investment being (unfairly) concentrated in the largest urban areas and especially in the Danish capital Copenhagen, with rural and peripheral areas being left behind. Since 2015, many political initiatives have been taken with the aim of creating a 'Denmark in Balance'. This political focus on peripheral and rural areas has continued under the present Social Democratic government. The current Danish Minister of the Interior and Housing Kaare Dybvad Bek has stated that "We need a country in balance – because Denmark is too small for large differences[3]" (Bek, 2021). In order to address the current regional imbalance, he has suggested that local accessibility to higher education should be improved,

[2] My own translation from Danish "Alligevel er mulighederne blevet mere skævt fordelt."
[3] My own translation from Danish: "Vi skal have et land i balance – for ja, Danmark er for lille til store forskelle."

stating that Denmark needs "more educational opportunities across the country"[4] (Bek, 2021).

The underlying assumption behind the political goal of creating more educational opportunities (in all corners of the country) is that the education level of a region's population is a key determinant of regional growth and development. However, in a Danish context, there is a lack of empirical evidence about the extent to which the presence of educational institutions in an area affects levels of participation in education among the population of that area. Hence, it is uncertain how the creation of more educational opportunities will influence enrolment in education programmes. Likewise, there is uncertainty about whether areas designated as locations for new educational institutions will benefit from them, as graduates might move to other areas after graduation. This thesis will address the aforementioned uncertainties. I will draw on the key international research within this field, and also provide new empirical evidence in the case of Denmark.

1.2 Human capital and regional development

The idea that a highly educated workforce is essential for regional growth and development is uncontroversial and has been demonstrated in a long range of research studies (Corcoran & Faggian, 2017; Faggian & McCann, 2009). Romer (1986, 1990) and Lucas (1988) were the first to suggest that "embodied skills and competences" of the labour force are a key factor for economic productivity. Hence, they were the progenitors of the idea that persistent regional differences in economic growth between regions and nations can be explained by spatial differences in human capital. The notion "human capital" is usually used in economic theory to conceptualise the economic value of a person's skills, training and education level[5]. With the emergence of "knowledge intensive" jobs in post-industrial societies, many researchers agree that human capital has become increasingly important for regional development (Faggian & McCann, 2009; Florida, 2002; Glaeser, 2011). The recent decades' globalization and technological progress have made human capital even more

[4] My own translation from Danish: "Derfor har jeg fra første dag i ministerstolen fremhævet flere uddannelsesmuligheder rundt om i landet som et godt sted at starte."

[5] As Faggian and McCann (2009) observe, the notion of human capital has been extended and redefined in a variety of ways since its original use. See Chapter 2 for a more detailed description of the definition of human capital and the widening of the concept in the literature.

important for innovation and economic growth (Faggian & McCann, 2009; Florida, 2002).

Therefore, regions can benefit from their local human capital. The regional stocks of human capital can be augmented in two different ways: either by increasing educational participation among local youths, or by attracting human capital from other regions (Faggian & McCann, 2009). Consequently, establishing a local educational institution in a region would seem to be a reasonable strategy to achieve the goal of increasing the human capital of that region. However, establishing such a new educational institution will not necessarily enhance the education level of the local population nor do the job of contributing to the socio-economic success of the region.

Firstly, it is not evident that geographical distance to educational institutions is an important factor for young people's choice of education in a small country like Denmark. From international studies, we know that distance to higher education institutions has little or no association with enrolment when the distance is below 25 km. We lack knowledge about the present geographical dispersion of education institutions in Denmark and about how this dispersion affects the extent to which young people enrol in further education.

Secondly, even though the establishment of an educational institution in a region enhances the human capital of the region by producing graduates, some of this human capital can be lost as soon as any graduate moves to another region. A large literature has shown that education in itself makes people more migratory (Corcoran & Faggian, 2017). Hence, there is a chance that the highly educated graduates may leak to other regions after graduating. At present, empirical evidence for Denmark about which factors are important for retaining newly educated graduates within the borders of a region is sparse.

1.3 Research questions

Based on the above considerations, the aim of this thesis is to clarify to what extent the creation and re-location of new educational institutions will decrease regional disparities in human capital.

This has resulted in two overall research questions:

1. Does proximity to educational institutions increase the creation of human capital locally?
2. To what extent do levels of human capital influence inter-regional migration and settlement in urban areas?

In order to answer the first research question, I examine the role of distance to education at two different stages in the educational life course. The first stage investigated is when young people, at age 15, have left compulsory education and need to decide whether to continue in upper secondary education. The second stage is when young people, at age 18, graduate from upper secondary education, and need to decide whether to continue in higher education or not. Likewise, I examine, the influence of education on migration decisions (research question 2) at two different stages in the educational life course, namely after graduation from upper secondary education and after graduation from higher education.

By answering these two research questions this thesis provides new knowledge about how the initial choice of youth education influences young people's decision making on their migratory path. While many studies have shown that a person's level of education increases their geographical mobility, little is known about how a person's initial choice of education affects their subsequent migration patterns (Faggian & Franklin, 2014). Research has primarily focused on highly educated individuals and the determinants and consequences of the spatial mobility of recent university graduates (Corcoran & Faggian, 2017). However, there is a gap in the literature about what shapes those highly educated individuals' initial choices on their educational path. Evidence about the determinants of youth education choices and migration is needed to acquire a fuller picture of the sorting and divergence of regional human capital.

1.4 Empirical strategy

In order to understand which factors that influence the regional stock of human capital and migration flows at the regional level, we need to investigate the determinants of young individuals' decisions to enrol in education and of their decisions to migrate. I believe that knowledge about individual behaviour at the micro-level is essential if regional policies should be properly targeted. I have therefore formulated four empirical research questions, one question for each paper in the thesis: The first two questions consider the impact of local youth educations and the next two research questions consider the role of higher education.

Empirical research questions:

1. Does distance from home to general upper secondary school influence participation in general upper secondary education?
2. What is the effect of completing a general upper secondary education on an individual's decision to move out of a rural area?
3. Does geographical distance between home and higher education institutions play an important role for young people's decision making on their educational paths?
4. How does education level, prior migration and regional location influence highly educated individuals' decision to migrate after graduation?

I use quantitative, statistical methods to examine the research questions. The analysis in the four research papers is based on panel data from the administrative population register of Statistics Denmark, in combination with detailed distance measures between place of residence and educational institutions in Denmark. All analysis is conducted with micro-level data where individuals are the unit of analysis. Using the rich data source of individual longitudinal data retrieved from Statistics Denmark, I have had the unique possibility of examining education choice and migration patterns across time periods, across space, and across socio-economic backgrounds. Every single individual who lives in Denmark has a unique personal identification number in the administrative registers. This feature allows us to track individual decision-making over time and across the life cycle. The research questions address an inter-disciplinary

field, including geography, sociology, demography and economics. I will refer to the literature within regional economics and use an econometric methodology to examine the questions. By virtue of my interdisciplinary background in sociology and geography, I will also draw on theoretical considerations and methodological aspects from sociology and geography in the thesis.

1.5 Outline of PhD thesis

The thesis is structured in five introductory chapters, including this introduction, and four academic papers. The five chapters set the theoretical, methodological and institutional background of the four papers. Chapter 2 gives an introduction to key concepts and theories within the research field. Chapter 3 describes the methodological framework and provides my considerations about the benefits and weaknesses of the quantitative methods I have used. Chapter 4 gives a survey of the Danish historical and institutional context. Finally, chapter 5 summarises the main findings of the four academic papers and provides concluding reflections about the policy implications of the findings. The four papers in this thesis cover one research question each, covering different perspectives across the educational path of an individual. Paper 1 and 2, centre on the role of upper secondary education and paper 3 and 4 examine higher education institutions, as illustrated in Figure 1.

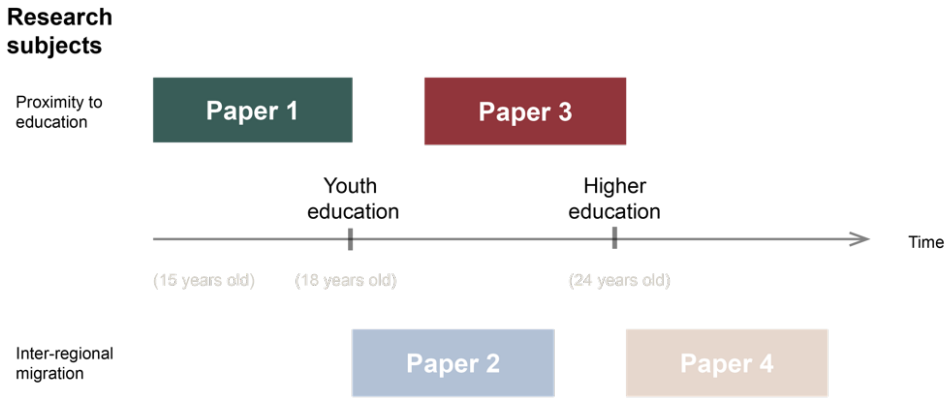

Figure 1: Papers included in the thesis

Paper 1 studies the impact of distance from home to general upper secondary education institutions on participation in general upper secondary education. The results suggest that distance has a small overall impact on participation in general upper secondary education. The effect of distance is moderated by parental background, and the effect is largest for young people from less-educated families. The small effect size is an unsurprising result, as general upper secondary education institutions are relatively evenly spread throughout Denmark. Hence, on average, young people live within a relatively short distance to their nearest general upper secondary education institution.

Paper 2 examines the effect of completion of a general upper secondary education programme on young individuals' migration patterns. The paper focuses on migration out of rural areas among young people between 18 and 24 years old. First, the paper provides evidence that youth migration from rural to urban areas has increased significantly between 1989 – 2017, while out-migration in all other age groups has not changed. The paper examines the determinants of migration among young people, and thereby adds nuances to our understanding of overall net-migration out of rural areas. The results suggest that completion of a general upper secondary education increases a person's probability of moving out of a rural area significantly. In addition, the analysis shows that the effect of general upper secondary education on out-migration varies across parental education, in the sense that young adults from low socio-economic backgrounds are most affected by their choice of youth education programme.

Paper 3 considers the role of distance to higher education in young people's decision making on their educational paths. The study focuses on young people who have graduated from a general upper secondary education programme. For this group, the geographical distance from home to higher education could potentially deter some individuals from enrolment in higher education, if the distance is too large. In addition, distance to different types of higher education programmes could be a deciding factor when choosing a study programme. Overall, the results suggests that (in the present educational landscape in Denmark) distance to higher education plays no significant role for young people's decision to enrol in higher education or not. However, when students who have decided to enrol in higher education are choosing between specific types of higher education study programmes, the distance to different study

programmes has a significant influence on the choice of study programme, in that the closest education programmes are more likely to be chosen.

Paper 4 provides descriptive evidence of important factors for retention of graduates in the region where their higher education institution is located. The paper focuses on newly educated graduates and analyses their migration decisions up until 10 years after graduation. The findings point to the importance of regional location and the local labour market for the retainment of highly educated graduates. In addition, the results show that previous student migration plays a significant role, as students who moved into a region to study are twice as likely to move again after graduation, compared to students who studied in their home region. Finally, the results suggests that university graduate are more migratory than college graduates.

In this thesis, I paint a broader picture of the impact of geographical access to educational institutions than has previously been available for Denmark. By focussing on two central educational choices for young people, choice of further education and choice of higher education, this thesis provides evidence on the association between geographical proximity to education, education choice and migration during the life cycle of the individual. Overall, the papers contribute to the debate about the role of local education institutions in regional development.

1.6 References

Andersen, J. G., & Etzerodt, S. F. (2018). Placering af offentlige institutioner (Location of public institutions). *Økonomi & Politik*, *91*(4), 34–49.

Austin, B., Glaeser, E., & Summers, L. (2018). Jobs for the Heartland: Place-Based Policies in 21st-Century America. *Brookings Papers on Economic Activity*, *49*(1 (Spring)), 151–255.

Bek, K. D. (2021, March 26). Kaare Dybvad: Storbyerne skal ikke være rigmandsenklaver, som lærere besøger i arbejdstiden. *Altinget*.

Corcoran, J., & Faggian, A. (2017). 1. Graduate migration and regional development: An international perspective. In *Graduate Migration and Regional Development*. Edward Elgar Publishing.

Eliasson, K., Haapanen, M., & Westerlund, O. (2020). Regional concentration of university graduates: The role of high school grades and parental background. *European Urban and Regional Studies*, *27*(4), 398–414.

Faggian, A., & Franklin, R. S. (2014). Human Capital Redistribution in the USA: The Migration of the College-bound. *Spatial Economic Analysis*, *9*(4), 376–395.

Faggian, A., & McCann, P. (2009). Human capital and regional development. In *Handbook of Regional Growth Development Theories*. Edward Elgar.

Florida, R. (2002). *The Rise of the Creative Class: And How It's Transforming Work, Leisure, Community and Everyday Life*. Basic Books.

Frederiksen, M. (2021, January 1). Mette Frederiksens nytårstale 1. Januar 2021 ("Mette Frederiksens New Year's Speech January 1st. 2021"). In *New Year's Speech 2021*. The Prime Minister's Office.

Glaeser, E. (2011). *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier.* New York: Penguin Press.

Iammarino, S., Rodriguez-Pose, A., & Storper, M. (2019). Regional inequality in Europe: Evidence, theory and policy implications. *Journal of Economic Geography*, *19*(2), 273–298.

Larsen, J. N., Andersen, H. T., Haldrup, K., Hansen, A. R., Jacobsen, M. H., & Jensen, J. O. (2014). *Boligmarkedet uden for de store byer* (SBI rapport No. 2014:05). Aalborg Universitet.

Sørensen, E. S. (2021). *Human capital, migration and regional inequality: How the location of educational institutions influences our society* (PhD thesis). University of Copenhagen, Copenhagen.

CHAPTER 2

Literature and Central Concepts

2.1 Introduction

The analysis of the creation of human capital and migration flows has come to occupy a more central role in urban and regional research over the last three decades. Originally, research in this field was primarily focused on the location decisions of firms and industries (Faggian & McCann, 2009; McCann, 2012; Mellander & Florida, 2014). This focus stems initially from the work of Alfred Marshall, who argued that firms cluster in order to achieve the advantages of shared inputs, shared labour markets and knowledge spillover (Marshall, 1890). In the older industrial economy, firms' location decisions were strongly influenced by distance to natural resources and market transportation costs (Weber, 1909). As most jobs in industrial firms were standardized, there was little interest in examining the role of human capital in a firm's growth. However, with the emergence of more knowledge-intensive jobs in the post-industrial economy, human capital, skills and workers' creativity became crucial for innovation and economic growth (Faggian & McCann, 2009; Florida, 2002). Thus, the emergence of a "knowledge-based economy" (Bell, 1973; Drucker, 1993) aroused the research interest in human capital from a regional perspective.

Most advanced societies have shifted from industrial to post-industrial economies. With the shift in production and services, individual skill formation and workers' educational competences became more important for a firm's productivity (Scott, 2008). Many countries experienced an increasing divergence in economic growth rates, both with other countries and internally, between regions. This divergence was hard to explain by means of traditional neo-classical macroeconomic models (Abreu, 2014), which predicted economic convergence over time. This led to new theoretical contributions by Romer (1986, 1990) and Lucas (1988) suggesting that the embodied skills and competences of labour (human capital) is a key factor in explaining differences in growth across space. These theoretical contributions gave rise to an increasing research interest in the role of human capital from a regional perspective.

The present chapter summarises the central concepts and empirical research that is occupied with the link between human capital, inter-regional migration and regional development. The aim of the overview is to constitute a backdrop for research into the present PhD thesis. I begin by defining the notion of "human capital" in Section 2.2. Then, in Section 2.3., I provide a brief summary of the central theoretical developments that have associated human capital with economic growth in a non-spatial framework. In Section 2.4, I describe the various theoretical links between high levels of human capital and regional economic development. Section 2.5 describes the determinants of inter-regional migration and the reason why highly educated individuals are more geographically mobile than others. Section 2.6 presents the role of educational institutions in local regions, while Section 2.7 summarises and gives some conclusions. Finally, in Section 2.8 I give an overview the contributions of this thesis to the literature.

2.2 Definition of Human Capital

The notion of human capital places a monetary value on the individual's education, skills and training. The sum of these factors can be used in labour to produce economic value in a productive activity. The centrality of the notion of human capital to economic theory was in large part due to Gary Becker, whose 1964 book *Human Capital* provided the first unifying framework on the subject. In this framework, the two most identifiable ways of investing in human capital are years spent in education and on-the-job training (Becker, 1964). There was a general consensus in the literature about this definition of human capital until the 1990s.

However, Becker's original notion of human capital has been extended and redefined in a variety of ways since the 1990s (Faggian & McCann, 2009), many of which are much broader than its original usage by Becker. With the emergence of new growth theories in which knowledge spill-overs and human capital play a central role, the simple notion of human capital was found to have fallen short in its explanatory power regarding the new models of growth. This led to a widening of the concept as an increasing source of diverse data became available for measuring "embodied knowledge and skills" in empirical analyses (Faggian & McCann, 2009).

First, the notion of human capital was extended to include any natural or physical health and ability which improves an individual's acquisition of knowledge and skills. Second, the term was extended by Robert Putnam

(1993) to include the value of the individual's social network (Putnam, Leonardi, & Nonetti, 1993). With inspiration from sociology (Bourdieu, 1986) and political science, Putnam introduced the concept of "social capital", which included any social or institutional features that foster an individual's learnings and skills. Putnam's central point is that social institutions i.e., the interaction and trust between humans, is a resource with an impact on economic growth. The third extension comes from the concept of "creative capital", which was popularised by Richard Florida (Florida, 2002, 2005). This notion emphasises the knowledge and skills obtained under employment and measures human capital in terms of skills via occupations (Mellander & Florida, 2014).

Florida (2002) used occupational analysis to divide the workforce into three main occupational classes: the creative class, the working class, and the service class. High levels of human capital overlap with the "creative class", but according to Florida they are not the same thing (Mellander & Florida, 2014). Florida's notion of creative capital focuses on the role of social norms and values and the networks based on them, which has much in common with Putnam's idea by Putnam (Faggian & McCann, 2009). According to Faggian and McCann (2009), these recent extensions imply that the notion of human capital has become both blurred and so loosely defined that is it difficult to measure empirically. They argue in favour of leaving the original Becker (1964) definition of human capital intact and discussing other notions of capital in terms of "informal institutions". In this thesis, I deploy the notion of human capital in connection with educational levels. For the particular group that I study, namely young adults and recent graduates, educational levels are an appropriate proxy for skills, since other skills – from jobs and training, for instance – have not yet been accumulated.

2.3 Theories of Human Capital and Economic Growth

Growing up in the 1980s and 1990s, I tend to take it for granted that higher education leads to innovation and economic growth. The influence of higher education on economic growth, innovation, welfare and health has been taken as a matter of course in my generation and has never been controversial. Hence, I have never questioned the reasoning behind the increasing public investment in raising the population's levels of education. However, it was not until the 1980s that the role of human capital was formulated as a theoretical economic growth model in macroeconomics. Paul Romer (1986) and Robert Lucas (1988) were the first to introduce human capital as an "endogenous" source of growth

related to the classic Cobb-Douglas production function, a contribution for which they received the Nobel Prize in Economics in 2018 ("The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2018,"). Their work gave the individual worker's skills and education a central role in economic growth models.

In the traditional neoclassical growth models (Solow, 1956; Swan, 1956), total output (production) is explained by the product of two traditional inputs: capital input (K) (machinery, equipment and buildings) and labour input (L) (personal hours worked). Within a firm, the capital input (L) and the labour input (L) have diminishing returns to scale. This means that increasing the input does not create an equivalent output. Economic growth is therefore explained by a third "generic" input, often referred to as "total factor productivity" (TFP) or simply "the residual" (often denoted A). The TFP is a measure of (unexplained) economic efficiency, and neoclassical economists assumed that the TFP was a product of technological change and innovation, which were *external* to the firm and available to everybody.

Romer (1986, 1990, 1994) suggested that this technological progress was in fact something *internal* to the firm, which could be endogenized in the classic Coob-Douglas production function, hence the name of the new theory: "endogenous growth theory". According to endogenous growth theory, growth is primarily the result of endogenous and not external forces. Romer argued that knowledge is not subject to diminishing returns to scale, like capital or real estate. Thus, Romer's new growth theory reached the opposite conclusion than the neoclassical models, which predicted a catching up process in which the poorest nations would catch up with the richest, leading to economic convergence. Quite the contrary, Romer's model was able to explain the increasing *divergence* between countries according to their levels of human capital.

Lucas (1988) proposed human capital as "real engine of growth" and identified human capital as a possible explanation for endogenous growth. Thus, investment in human capital, innovation, and knowledge are significant contributors to economic growth. Lucas (1988) proposed this version of the aggregate production function, incorporating human capital into the equation:

$$Y = AK^{\beta}(uhL)^{1-\beta}h_a^{\gamma} \qquad (2.1)$$

where K is physical capital, L is the number of workers, u is the number of working hours per worker and h is human capital (Faggian & McCann, 2009). Human capital thus appears twice in the equation. The first time, it represents the effect of individual human capital on the productivity of each worker. The second time, h_a^{γ} represents the average level of human capital. If positive externalities (knowledge spill-overs) exist, then $\gamma > 0$, and the term represents the positive externalities that human capital generates for the whole economy.

To summarize, Lucas (1988) distinguishes between the individual benefits and the external benefits of human capital. Even though the theory is a-spatial in essence, Lucas suggests that production will accumulate in locations with higher general levels of human capital because the productivity of a given worker will increase when other highly skilled workers are nearby and thus create positive knowledge spillover (Moretti, 2004b). Therefore, divergences in growth rates between nations and regions can be explained with reference to the share of human capital (Lucas, 1988; Romer, 1986). However, application of the endogenous growth model in a spatial context is not straightforward, assuming that the migration patterns of individuals with high levels of human capital are biased towards particular regions. In this case, national growth may be associated with both regional growth and regional decline. Therefore, the link between human capital and regional development will be addressed in more detail in the preceding section.

2.4 Human Capital and Regional Development

The theories of Romer (1986) and Lucas (1988) have spurred a large body of literature examining the role of human capital in regional economic growth. Regional increases in human capital can influence regional growth in two ways. The first concerns private returns to education: individuals with higher levels of education will, on average, obtain higher wages (Moretti, 2004b). In addition, higher education is also related to better health outcomes (Adler & Newman, 2002), the lower probability of unemployment and lower crime rates (Moretti, 2004b). In sum, highly educated people will pay higher local taxes and be less costly for local public authorities.

The second type of influence concerns social returns to education: spatial concentrations of human capital will create productivity gains and positive externalities. There is evidence that regions and cities become more productive when the share of the population with a college degree increases, including among workers who did not pursue a college education (Moretti, 2004a, 2004c;

Rauch, 1993). Moretti argued that high levels of human capital also create positive non-economic externalities, as highly educated individuals engage more in society (Moretti, 2004b).

From a regional policy perspective, local educational levels can be raised in two ways: encourage local residents to undertake higher education by making it more accessible, or attract the highly educated from other regions. These two strategies apply in both national and regional settings[6]. However, the net migration flows of highly educated individuals play a more central role in regional development settings because regions are "open systems" with a high degree of in- and outflows of human capital. Labour markets in different regions are much more connected, and the barriers to migration in terms of language, norms, culture and distance from friends and family are lower between regions than between nations (Faggian & McCann, 2009).

In the following two sections, I consider the two different strategies for raising the local stock of human capital in more detail. First, I briefly review the determinants of inter-regional migration flows with a focus on high-skilled individuals. Then I review the literature on increasing the local stock of migration by educating a larger share of the local population. I will focus on the role of proximity to local higher education institutions.

2.5 Determinants of migration

Highly educated individuals are more likely to migrate between regions. The association between educational level and geographical mobility is uncontroversial and has been documented in a large range of empirical studies across countries (see, among others, Ahlin et al., 2018; DaVanzo, 1978; Faggian et al., 2006, 2007a, 2007b; Plane, 1993). The literature suggests that increasing levels of individual education will increase the likelihood of individual geographical mobility.

There are several explanations for individuals with high levels of human capital being more spatially mobile. A central explanation is that highly educated individuals have higher returns to migration and lower costs of migrating. This implies that overall they are more likely to benefit from migrating. This explanation stems from "human capital migration theory",

[6] There is a large literature on migration of the highly educated and the brain-drain from a global perspective, which I do not describe in this summary.

proposed by Sjaastad in 1962. Sjaastad began the tradition of threating migration as an investment in future benefits (Cadwallader, 1992). The idea is that through migration individuals can obtain access to opportunities beyond their present labour market. Like any other investment, migration is associated with costs and returns, and an individual will decide to migrate when the (perceived) returns are larger than the costs of migrating (Sjaastad, 1962).

There are many reasons why highly educated people could have higher net benefits of migrating compared to the less educated. (Faggian, Corcocan, & Partridge, 2015; Faggian & McCann, 2009) provide exhaustive overviews of the empirical literature. I will briefly describe the main arguments. First, highly educated individuals have higher information-gathering skills. This gives them better access to information about opportunities in alternative locations, as underlines by (Levy & Wadycki, 1974). Higher education is associated with the ability to obtain and analyse information efficiently. Another explanation is related to the fact that previous migration lowers the barriers to future migration decisions (DaVanzo, 1983). A large proportion of the highly educated group have found it necessary to move, as HEIs are often located in central locations in the largest cities. Therefore, as the majority of the highly educated have tried to move across regions before, it is easier for them to move again in the future. It has also been suggested that the more highly educated face lower psychological costs associated with moving. Finally, the more highly educated have lower risks of migrating to a new labour market because they face lower risks of becoming unemployed.

Given the fact that inter-regional migration is highly selective (Greenwood, 1975), the consequences for the economy in the regions of origin and destination can be substantial (Faggian, Rajbhandari, & Dotzel, 2017). Therefore, a large literature has emerged concerned to determine which factors serve to attract and retain individuals with high human capital. The idea of "push and pull" factors in the economic literature originates with the pioneering work of Ravenstein (1889). The literature has focused on two main types of regional attractors (push and pull factors): economic factors and amenities (Faggian et al., 2015; Ravenstein, 1889; Storper & Scott, 2009; V. Venhorst, Dijk, & Wissen, 2011). The economic drivers of migration include unemployment rates, wage levels, and housing prices. Local amenities include climate, natural amenities (access to nature), cultural amenities (presence of theatres, museums), local public goods, and also negative amenities such as crime rates and pollution (Faggian et al., 2015).

(Millington, 2000) emphasises the importance of conducting migration studies disaggregated by age groups. He shows that labour market characteristics (economic determinants) are most important early in life and that this factor declines with age. Conversely, he finds that housing and local amenities (such as climate) become more important through the life-cycle (Faggian et al., 2015). Faggian and Franklin (2014) emphasise that young people are less sensitive to climate than older people when choosing destination region. In addition, they argue that climate characteristics act more as pull than push factors for migration. However, Plane (1993) and Whisler et. al. (2008) find that the main determinant of migration differs across different stages of the individual's life-cycle.

2.6 The Role of Local Educational Institutions

To recap, from a regional perspective there are two main ways to increase the local level of human capital: either to attract high human capital from other regions, or to educate more people locally. However, the high mobility of high-skilled individuals produces a complication because, even when produced locally, human capital can easily leak out of the area and therefore not generate the predicted multiplier effects. Therefore, a local strategy to educate more people only works if a region simultaneously manages to retain a large proportion of its highly educated graduates. The following section will summarize the present research concerning 1) the role of local educational institutions on local levels of education, and 2) which factors play a role in in a local area retraining graduates after they have obtained their degrees.

Distance to education

There is a vast literature examining the association between distance from home to higher education institutions and participation in higher education (Cullinan, Flannery, Walsh, & McCoy, 2013; Frenette, 2006; Gibbons & Vignoles, 2012; Griffith & Rothstein, 2009; White & Lee, 2020). Overall, the literature finds no association or only a relatively modest association between these factors, suggesting that young adults with a relatively long distance to school (above 25 kilometres) have a lower probability of enrolling in education. The results are sensitive to the countries where they were conducted and to the models that were used. The central problem is that the majority of such studies fail to pay close attention to issues of causality. Because of data limitations, most empirical studies are not able to disentangle the influence of distance from the influence

of residential sorting. There is an obvious challenge in these studies in determining whether the negative association between distance and educational choice is due to distance or to the fact that highly educated families tend to locate in urban areas where distance to schools is shorter on average. Thus, the association that is found could be solely due to the fact that parental education is associated with both distance to school and educational choice. I seek to solve this causality issue in the literature in Paper 1.

To summarize, the literature shows that, even ignoring the causality issues in most studies, there are moderate negative effects of distance to schools. Thus, the overall conclusion is that family background and grades in compulsory education have much more explanatory power than distance. From a regional policy perspective, a focus on attracting highly educated families and increasing learning through compulsory schooling might be a more sufficient investment (Gibbons & Vignoles, 2012). Other barriers to education are more important than geographical distance, such a tuition fees, affordable student housing etc.

Retention of highly educated graduates

From a regional perspective, it is important to retain the human capital that has been "produced" in the region after graduation in order to obtain the desired positive externalities of human capital. Thus, several empirical studies have examined the migration patterns of highly educated graduates (Faggian et al., 2007a, 2007b, 2017; Haapanen & Tervo, 2012; V. A. Venhorst, 2013; V. Venhorst et al., 2011), the overall conclusion being that there are large regional differences in the ability to retain them. Haapanen and Tervo (2012) find that in Finland 70 percent of students in the Helsinki region (the capital region) stay there for 10 years after graduation. However, there are also large regional differences. In respect of universities located in smaller university cities in the periphery, only 37 percent of graduates were still living in the region after 10 years. Venhorst (2013) shows that the majority of graduate migration between periphery and centre in the Netherlands goes towards the four largest metropoles. Most graduates come from regions close to their region of education, and they move to a neighbouring region after graduation.

2.7 Summary

The above summary of the literature has shown that the link between human capital and regional development is a complex question. It has shown that national growth in human capital does not necessarily imply equal regional growth.

A national strategy that increases the overall share of human capital can lead to regional growth in some areas and at the same time speed up the brain-drain in other regions. Thus, several studies point out that increases in human capital increase regional divergence (Austin, Glaeser, & Summers, 2018; Eliasson, Haapanen, & Westerlund, 2020; Iammarino, Rodriguez-Pose, & Storper, 2019). The general rise in educational levels that have been seen all over the world is also a driver of unequal regional growth. In many places this is a vicious circle. Thus, national growth may be associated with both regional growth and regional decline.

2.8 Contribution

Overall, the empirical literature has mainly focused on highly skilled individuals (Faggian & Franklin, 2014): evidence for the initial choices regarding education and inter-regional migration in the life-cycle is sparse. The aim of the present PhD is to fill out this gap in the literature. Figure 2 shows where the evidence for educational choices and migration are "thick" (that is, well documented in the literature) and where it is "thin" in the sense of being limited.

Figure 2 is in the form of a decision tree which presents a scheme of the individual choice set that applies after leaving elementary school. After leaving elementary school at age 15, the individual is confronted with three choices: to enrol in a vocational education and training (VET) course, to proceed to general upper secondary education, or to leave the education system altogether and enter the labour market. The literature on the potential geographical obstacles for first educational choice is limited. I am aware of two empirical studies that have examined the association between distance from home to school and enrolment in upper secondary education (Dickerson & McIntosh, 2013; Falch, Lujala, & Strøm, 2013). After graduation from upper secondary education (VET or general upper secondary education), the individual can decide to move or stay in the region. Empirical evidence for such initial migration decisions is also sparse. Faggian and Franklin (2014) examined the migration of the college-bound, that is, of the decision to migrate from high school to college. However, the other two groups (those with a VET education and those who entered the labour market after elementary school) are not included in the study. Hence, the evidence for youth migration between the ages of 18 and 24 is also sparse.

Those group young people who completed a general upper secondary education, which in makes up about 65 percent of the typical birth cohort, have the option of continuing their educational path and enrolling in higher

education. For this group, the geographical distance from home to higher education could potentially deter some individuals from enrolment in higher education if the distance is too long. In addition, the distance to different types of higher education programmes could be a deciding factor when choosing study programme. The association between distance from home and enrolment in higher education has been extensively examined in the international literature. The present PhD contributes the literature by providing empirical evidence for the situation in Denmark specifically.

Finally, after completing higher education, the individual faces the choice of staying in the region of studies or moving to another region. This question has received a lot of attention in the literature, as mentioned beforehand (Faggian et al., 2015, 2017). Thus, the two "branches" in Figure 2 are depicted as thicker and are highlighted. Although my own contribution to this large body literature may be limited, I provide new evidence from Denmark about the role of city size in determining graduate migration.

Looking overall at the decision tree below, it is evident that the lowermost branches of the "evidence tree" are the thinnest. This suggests that the geographical disparities in initial decisions about education and inter-regional migration are still underexplored. As there is a large degree of path-dependency in educational and residential choices, knowledge about these early decisions is essential if the regional disparities in human capital and migration are to be fully understood. The main contribution of this thesis is that it provides new knowledge about what influences these initial migration decisions.

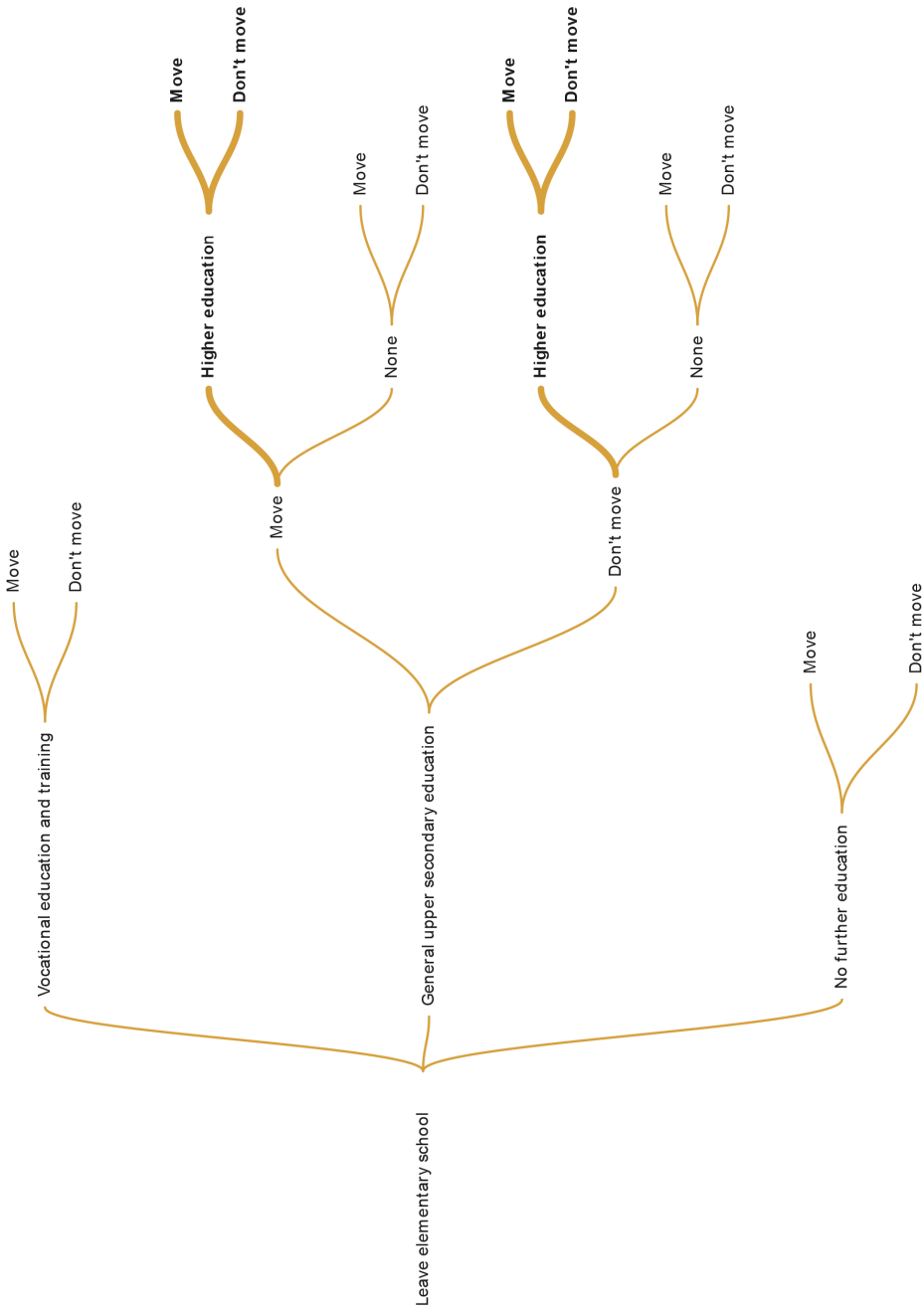

Figure 2: Individual decision tree

2.9 References

Abreu, M. (2014). Neoclassical Regional Growth Models. In M. M. Fischer & P. Nijkamp (Eds.), *Handbook of Regional Science* (pp. 169–191). Berlin, Heidelberg: Springer.

Adler, N. E., & Newman, K. (2002). Socioeconomic disparities in health: Pathways and policies. *Health Affairs (Project Hope)*, *21*(2), 60–76.

Ahlin, L., Andersson, M., & Thulin, P. (2018). Human capital sorting: The "when" and "who" of the sorting of educated workers to urban regions. *Journal of Regional Science*, *58*(3), 581–610.

Austin, B., Glaeser, E., & Summers, L. (2018). Jobs for the Heartland: Place-Based Policies in 21st-Century America. *Brookings Papers on Economic Activity*, *49*(1 (Spring)), 151–255.

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.

Bell, D. (1973). *The coming of post-industrial society: A venture in social forecasting.* New York: Basic Books.

Bourdieu, P. (1986). The Forms of Capital. In J. G. Richardson (Ed.), *Handbook of Theory and Research for the Sociology of Education* (pp. 241–258). Greenwood Press.

Cadwallader, M. (1992). *Migration and Residential Mobility: Macro and micro approaches*. Wisconsin: The University of Wisconsin Press.

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.

DaVanzo, J. (1978). Does Unemployment Affect Migration?- Evidence from Micro Data. *Review of Economics & Statistics*, *60*(4), 504.

DaVanzo, J. (1983). Repeat Migration in the United States: Who Moves Back and Who Moves On? *Review of Economics & Statistics*, *65*(4), 552. doi: 10.2307/1935923

Dickerson, A., & McIntosh, S. (2013). The Impact of Distance to Nearest Education Institution on the Post-compulsory Education Participation Decision. *Urban Studies*, *50*(4), 742–758.

Drucker, P. F. (1993). *Post-capitalist society.* (Reprint.). Oxford: Butterworth Heinemann.

Eliasson, K., Haapanen, M., & Westerlund, O. (2020). Regional concentration of university graduates: The role of high school grades and parental background. *European Urban and Regional Studies*, *27*(4), 398–414.

Faggian, A., Corcocan, J., & Partridge, M. (2015). 22 Interregional migration analysis. In C. Karlsson, M. Andersson, & T. Norman (Eds.),

Handbook of Research Methods and Applications in Economic Geography. Cheltenham, UK: Edward Elgar Publishing Limited.

Faggian, A., & Franklin, R. S. (2014). Human Capital Redistribution in the USA: The Migration of the College-bound. *Spatial Economic Analysis*, *9*(4), 376–395.

Faggian, A., & McCann, P. (2009). Human capital and regional development. In *Handbook of Regional Growth Development Theories*. Edward Elgar.

Faggian, A., McCann, P., & Sheppard, S. (2006). An analysis of ethnic differences in UK graduate migration behaviour. *The Annals of Regional Science*, *40*(2), 461–471.

Faggian, A., McCann, P., & Sheppard, S. (2007a). Human Capital, Higher Education and Graduate Migration: An Analysis of Scottish and Welsh Students. *Urban Studies*, *44*(13), 2511–2528.

Faggian, A., McCann, P., & Sheppard, S. (2007b). Some Evidence That Women Are More Mobile Than Men: Gender Differences in U.k. Graduate Migration Behavior. *Journal of Regional Science*, *47*(3), 517–539.

Faggian, A., Rajbhandari, I., & Dotzel, K. R. (2017). The interregional migration of human capital and its regional consequences: A review. *Regional Studies*, *51*(1), 128–143.

Falch, T., Lujala, P., & Strøm, B. (2013). Geographical constraints and educational attainment. *Regional Science and Urban Economics*, *43*(1), 164–176.

Florida, R. (2002). *The Rise of the Creative Class: And How It's Transforming Work, Leisure, Community and Everyday Life*. Basic Books.

Florida, R. (2005). *The Flight of the Creative Class: The New Global Competition for Talent*. New York: Harper Business.

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

Gibbons, S., & Vignoles, A. (2012). Geography, choice and participation in higher education in England. *Regional Science and Urban Economics*, *42*(1–2), 98–113.

Greenwood, M. J. (1975). Research on Internal Migration in the United States: A Survey. *Journal of Economic Literature*, *13*(2), 397–433.

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.

Haapanen, M., & Tervo, H. (2012). Migration of the Highly Educated: Evidence from Residence Spells of University Graduates. *Journal of Regional Science*, *52*(4), 587–605.

Iammarino, S., Rodriguez-Pose, A., & Storper, M. (2019). Regional inequality in Europe: Evidence, theory and policy implications. *Journal of Economic Geography*, *19*(2), 273–298.

Levy, M. B., & Wadycki, W. J. (1974). What Is the Opportunity Cost of Moving? Reconsideration of the Effects of Distance on Migration. *Economic Development and Cultural Change*, *22*(2), 198–214. doi: 10.1086/450705

Lucas, R. E. (1988). On the Mechanics of Economic Development. *Journal of Monetary Economics*, (22), 3–42.

Marshall, A. (1890). *Principles of economics*. Macmillan London.

McCann, P. (2012). *Urban and regional economics*. London: Routledge.

Mellander, C., & Florida, R. (2014). The Rise of Skills: Human Capital, the Creative Class, and Regional Development. In M. M. Fischer & P. Nijkamp (Eds.), *Handbook of Regional Science* (pp. 317–329). Berlin, Heidelberg: Springer Berlin Heidelberg.

Millington, J. (2000). Migration and Age: The Effect of Age on Sensitivity to Migration Stimuli. *Regional Studies*, *34*(6), 521–533.

Moretti, E. (2004a). Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, *121*, 175–212.

Moretti, E. (2004b). Human Capital Externalities in Cities. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics. Volume 4. Cities and Geography* (First edition). Amsterdam: Elsevier B.V.

Moretti, E. (2004c). Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions. *The American Economic Review*, *94*(3), 656–690.

Plane, D. (1993). Demographic Influences on Migration. *Regional Studies: The Journal of the Regional Studies Association*, *27*(4), 375–383.

Putnam, R. D., Leonardi, R., & Nonetti, R. Y. (1993). Social Capital and Institutional Success. In *Civic Traditions in Modern Italy*. *Making Democracy Work* (pp. 163–186). Princeton University Press.

Rauch, J. E. (1993). Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities. *Journal of Urban Economics*, *34*(3), 380–400.

Ravenstein, E. G. (1889). The laws of migration. *Journal of the Royal Statistical Society*, *Vol. 52*(No. 2), 241–305.

Romer, P. M. (1986). Increasing Returns and Long-Run Growth. *The Journal of Political Economy*, *94*(5), 1002–1037.

Scott, A. J. (2008). *Social Economy of the Metropolis: Cognitive-Cultural Capitalism and the Global Resurgence of Cities*. Oxford ; New York: Oxford University Press.

Sjaastad, L. A. (1962). The Costs and Returns of Human Migration. *Journal of Political Economy*, *70*(5), 80–93.

Solow, R. M. (1956). A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics*, *70*(1), 65–94.

Storper, M., & Scott, A. J. (2009). Rethinking human capital, creativity and urban growth. *Journal of Economic Geography*, *9*(2), 147–167.

Swan, T. W. (1956). Economic growth and capital accumulation. *Economic Record*, *32*(2), 334–361.

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2018. (n.d.). Retrieved April 27, 2021, from NobelPrize.org website: https://www.nobelprize.org/prizes/economic-sciences/2018/summary/

Venhorst, V. A. (2013). Graduate migration and regional familiarity. *Tijdschrift Voor Economische En Sociale Geografie*, *104*(1), 109–119.

Venhorst, V., Dijk, J. V., & Wissen, L. V. (2011). An Analysis of Trends in Spatial Mobility of Dutch Graduates. *Spatial Economic Analysis*, *6*(1), 57–82.

Weber, A. (1909). *Theory of the Location of Industries*. Chicago: University of Chicargo Press.

White, P. M., & Lee, D. M. (2020). Geographic Inequalities and Access to Higher Education: Is the Proximity to Higher Education Institution Associated with the Probability of Attendance in England? *Research in Higher Education*, *61*(7), 825–848.

CHAPTER 3

Research Design and Methods

When conducting scientific research, it is essential to be aware of one's assumptions. All research and knowledge are based on more or less implicit assumptions of how the world functions. In this chapter, I will describe my scientific theoretical approach, present the methods and data used in the research papers, and reflect on the consequences for my research findings. First, in Section 3.1, I clarify the methodological approach applied in the thesis. Then, in Section 3.2 I define the notion of causality and in Section 3.3 present a simple model of the research design. Section 3.4 presents the data sources, and Section 3.5 discusses the limitations and advantages of the applied method and data sources.

3.1 Methodological individualism

The aim of the thesis is to understand the causes of regional differences in human capital. I believe that in order to understand what has caused a social phenomenon like regional differences in human capital, one needs to study and explain the individual behaviour that causes the social phenomenon – here, for example, individuals' decisions to migrate or participate in education. In other words, in conducting my research I have adopted a form of methodological individualism.

The methodological approach entailed in the notion of "methodological individualism" dates back to the famous sociologist Max Weber and the later work of the economist Joseph Schumpeter (Gilje & Grimen, 2002). Weber argued that any social phenomenon in society can be traced back to roots in individual behaviour. Hence, in order to understand any social phenomenon, the researcher must interpret the individual actions that caused it. In other words, this entails a research design in which macro-level phenomena in society are explained by examining individual decision-making at the micro-level.

Methodological individualism can be criticised for its inability to explain how structures in society also influence individuals' decisions (Gilje & Grimen, 2002, Chapter 8). James Coleman, a central figure in the rational choice movement in sociology, suggested a version of methodological individualism that incorporated social structural constraints (Hedström & Ylikoski, 2010).

Coleman underlines the need to link the micro- and macro-levels in the social sciences. He argued that social science theories should be committed to showing how individual actions in sum function in a way that creates a certain association at the macro-level between social phenomena, and how, conversely, individuals are influenced by social phenomena (Gilje & Grimen, 2002, Chapter 8).

In the papers, I use statistical models and micro-level individual data to examine overall phenomena in society, making Coleman's scientific approach implicit in the way I conduct my research. In Figure 3 below I illustrate the methodological approach and show how inferences can be drawn from the individual behaviour I examine in this thesis to bear on the association between the overall social phenomena. I have borrowed this famous figure, which is often called "Coleman's boat" because of its shape, from Coleman.

According to Coleman, a full explanation of the association between two macro-level factors (arrow 4) should be explained as macro-to-micro and micro-to-macro transitions. First, arrow 1) shows how the location of institutions at the macro level places constraints on the individual (e.g. distance to education). Arrow 2) indicates the individual behaviour (e.g. how educational choices impact on migration). Lastly, arrow 3) shows how the sum of individual behaviour accumulates to produce macro-level phenomena (e.g. regional differences in human capital).

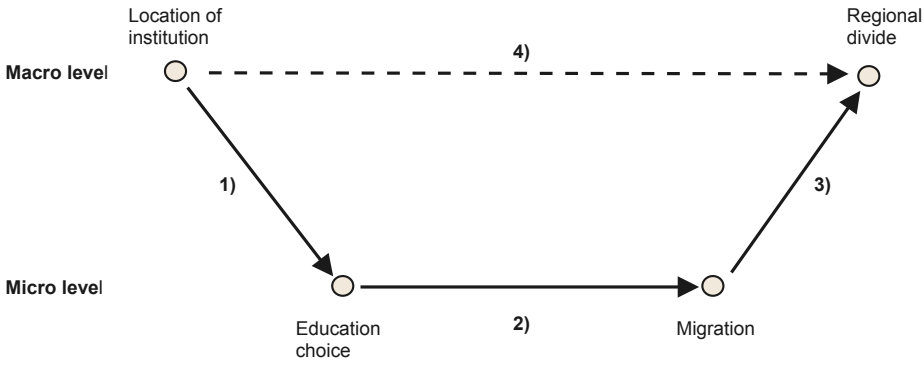

Figure 3. Coleman's boat

3.2 On causality

Research designs that can identify a causal effect are seen as the acme of econometric methods. But what does it mean to say that an association is causal? And why is it difficult to identify? A central characteristic of the concept of *causality* is the notion of producing (Cadwallader, 1992, p. 76). That is, if X is a cause of Y, then a change in X should always produce a change in Y. The association is not causal simply because a change in X is always associated with a change in Y (Cadwallader, 1992, p. 76). Unfortunately, we cannot observe this phenomenon in data directly, that is, in real time, hence, we have only indirect evidence of the covariation between the characteristics and the phenomena. There are at least three conditions which must hold in order to claim that X causes Y (Cadwallader, 1992). First, an effect cannot precede its cause. Thus, a causal effect implies a relationship that takes place over time. Second, a causal effect entails the presence of a relationship between two variables. Thus, if you know the values of one variable, you should have a better chance of forecasting the value of the other variable. Hence two variables should not be independent of each other. Third, the relationship between the two variables should not be spurious. A spurious relationship occurs when a variable Z causes both X and Y so that the relationship between X and Y disappears when Z is controlled for. However, this is harder to detect empirically. This variant could also indicate that Z is an intervening variable, so that X might cause Z and Z might cause Y.

3.3 Research design

In this section I present an illustration of the two overall hypotheses I test in the four papers. The particular hypotheses are derived from theoretical explanations, which I try to test against observable data. Figure 4 below illustrates the two different causal hypotheses I test in this thesis. First, I test whether there is a causal relationship between distance to school (D) and educational choice (X) (arrow 1). I add a range of control variables, most importantly socioeconomic status (SES), which is likely to influence both D (through residential sorting) and X (through social inheritance in the educational system). Hence, if controlling for SES makes the relationship between distance (D) and enrolment in education (X) disappear, then the relationship is due to spurious causation. In Paper 1, I also apply a research design which should make it probable that the relationship between distance to upper secondary schools and enrolment is causal. That is, I use changes in

distance that are not correlated with SES and other individual characteristics, and evaluate the effect of a change in distance.

Second, I test whether educational choice (X) causes the decision to move (Y) (arrow 2). Again, the main control variables are SES, geographical variables and individual characteristics, as shown in Figure 4. In Paper 2, I apply an additional IV research design to eliminate the possibility of selection on variables that I do not observe in the data.

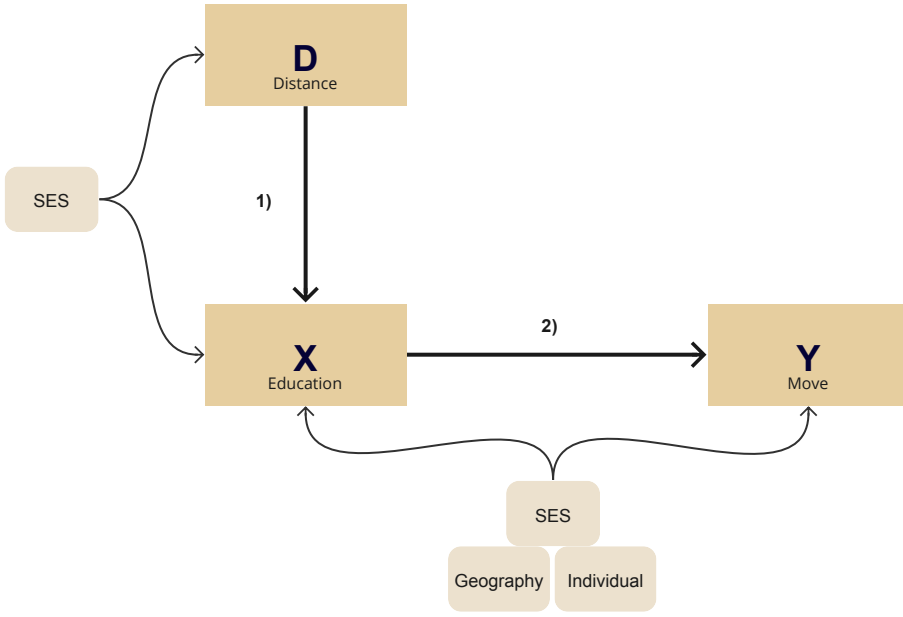

Figure 4. Illustration of empirical models

3.4 Methods and data

In this thesis I examine my research questions through a combination of descriptive statistics, statistical methods and causal research designs. I use rich administrative datasets to examine participation in education and inter-regional migration among young people in Denmark. The empirical analysis is based on a composite dataset, which I have linked together from a number of administrative data sources. The administrative data sources are retrieved from Statistics Denmark. The administrative registers contain annual information of the total Danish population from 1980 to 2020. In addition, I have sought to verify the location information of all youth- and higher education institutions in

Denmark back in time. The address list for the institutions was provided by the Danish Ministry of Children and Education. However, the list of institutions and their exact locations can be misleading in cases where institutions have been merged at the administrative level, but where the location of the place of study remain unchanged. Hence, I have manually surveyed the list of institutional addresses and information about their opening and closure to arrive at the number and location of institutions in the time period used in each respective analysis. These periods are 1982 to 2014 for upper secondary education and 2006 to 2011 for higher education.

The exact geographical location of the place of study is essential in order to calculate the distance from home to upper secondary education and to higher education respectively. The distances calculated are those along the road network. The distance measures are used in Papers 1, 2, and 3. In Papers 1 and 3, distance to school is the key explanatory variable, while in Paper 2, distance to nearest school is used as an instrumental variable.

3.5 Advantages and limitations

The main limitation of the data source is that I can only study factors that can be operationalised thorough the variables that are present in the administrative data sources. A wide range of variables are registered, but of course there are limitations. The administrative data sources provide only information about individual facts, such as gender and age, and individual actions, such as educational enrolment or moves between residential addresses. I have no information about individual considerations and motives behind individual actions. Valuable information about parents, siblings, partners and children can be sourced, making it possible to take social background and relatives into account in the analysis. However, the administrative data lack information about friends, close neighbours and other social networks that might be important for both the decision to move and the decision to participate in education.[7]

The advantages of the applied quantitative methods and the data sources are that the sources contain information about the whole Danish population, eliminating selection bias regarding the research sample. As the administrative

[7] As long as social networks and ties are random and not correlated with distance to school, this will not bias the results.

data are collected and registered by the relevant public authority, personal post-rationalization is also eliminated. The statistical methods make it possible to detect patterns and associations in the large data source and assess whether the association is substantial or trivial in the overall picture. In addition, by using statistical methods, one can examine the influence of one factor while taking other factors into account. However, a theory is needed to render a statistical association probable as a causal relationship.

Finally, quantitative methods are objective and have a large degree of external validity. I will argue that the method is ideal for examining policy interventions or general incidents that happen to the vast majority. In order to decide the best use of society's resources, knowledge and empirical evidence about the majority, the average citizen, is essential. I consider it important that politicians base their choices on objective knowledge and that their interventions are in favour of the majority of the population group they wish to target.

3.6 References

Cadwallader, M. (1992). *Migration and Residential Mobility: Macro and micro approaches*. Wisconsin: The University of Wisconsin Press.

Gilje, N., & Grimen, H. (2002). *Samfundsvidenskabernes forudsætninger: Indføring i samfundsvidenskabernes videnskabsfilosofi* (1. udgave.). Kbh: Hans Reitzel.

Glaeser, E. L. (1999). Learning in Cities. *Journal of Urban Economics*, *46*(2), 254–277.

Hedström, P., & Ylikoski, P. (2010). Causal Mechanisms in the Social Sciences. *Annual Review of Sociology*, *36*(1), 49–67.

CHAPTER 4

Institutional Setting and Historical Context

In this chapter, I will provide some stylized facts about the historical development of educational attainment in Denmark and the geographies of its educational institutions and human capital, as well as the country's economic growth. I will draw comparisons with educational attainment in other OECD countries in order to place the Danish case in its international setting. The descriptive statistics provide a backdrop to the following statistical analysis in the four papers in the present PhD dissertation. First, the next section provides a brief introduction to the Danish education system.

4.1. The Danish Education System

Figure 5 below illustrates the mainstream education system in Denmark. I will focus on youth- and higher education programmes in the following description, as the transitions from compulsory education into youth education and higher education are the focus of this PhD dissertation. For a presentation of adult education and training, see the booklet *The Danish Education System* (Ministry of Higher Education and Science, 2016). The education system in Denmark is universal, with no tuition fees for compulsory, upper secondary- and higher education. In addition, the Danish state provides financial support (SU) to all students over the age of eighteen who are enrolled in a youth- or higher education programme (Ministry of Higher Education and Science, 2016). Students receive a monthly student grant of approximately Euros 850 (at 2021 rates)[8] during their studies. Thus, the amount of government support to students is not trivial, the aim being that it should be possible for any individual, regardless of family background, to pursue an education.

Compulsory education in Denmark consists of 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 education, continue on to 10th grade, or enter upper secondary education. In Denmark, upper secondary education

[8] This is equivalent to DKK 6,321 before income taxes in 2021. If the student is still living at home, the SU rate depends on parental income and is between approximately Euros 124 and 367 (equivalent to DKK 920 to DKK 2,725).

programmes have two main tracks of education: general upper secondary education, and vocational education and training (VET). General upper secondary education programmes are primarily aimed at preparing the student for further studies, whereas the VET programmes are targeted at an occupation in a specific trade or industry (Ministry of Higher Education and Science, 2016).

General upper secondary education is divided into three types of education programme: a general academic programme (HF and STX), a commercial programme (HHX) and a technical programme (HTX). The general upper secondary education programmes are typically three-year programmes[9] that prepare students for further studies. The formal requirement for admission to the majority of higher education programmes is usually a general upper secondary education degree.

The VET programmes provide more than a hundred occupation-specific tracks and offer many different types of education. Vocational education is categorized into four main categories: 1) care, health and pedagogy, 2) administration, commerce and business service, 3) food, agriculture and hospitality, and 4) technology, construction and transportation. The vocational system is a dual educational system, such that students shift between school-based learning and practical apprenticeship training. The duration of programmes varies from 2 to 5½ years, depending on the profession, but is most typically 3½ to 4 years. All vocational education programmes qualify the student for labour market entry as a skilled worker, and the vast majority of students enter the labour market right after graduation (Ministry of Higher Education and Science, 2016) .

[9] However, some programs (HF) are only 2 years.

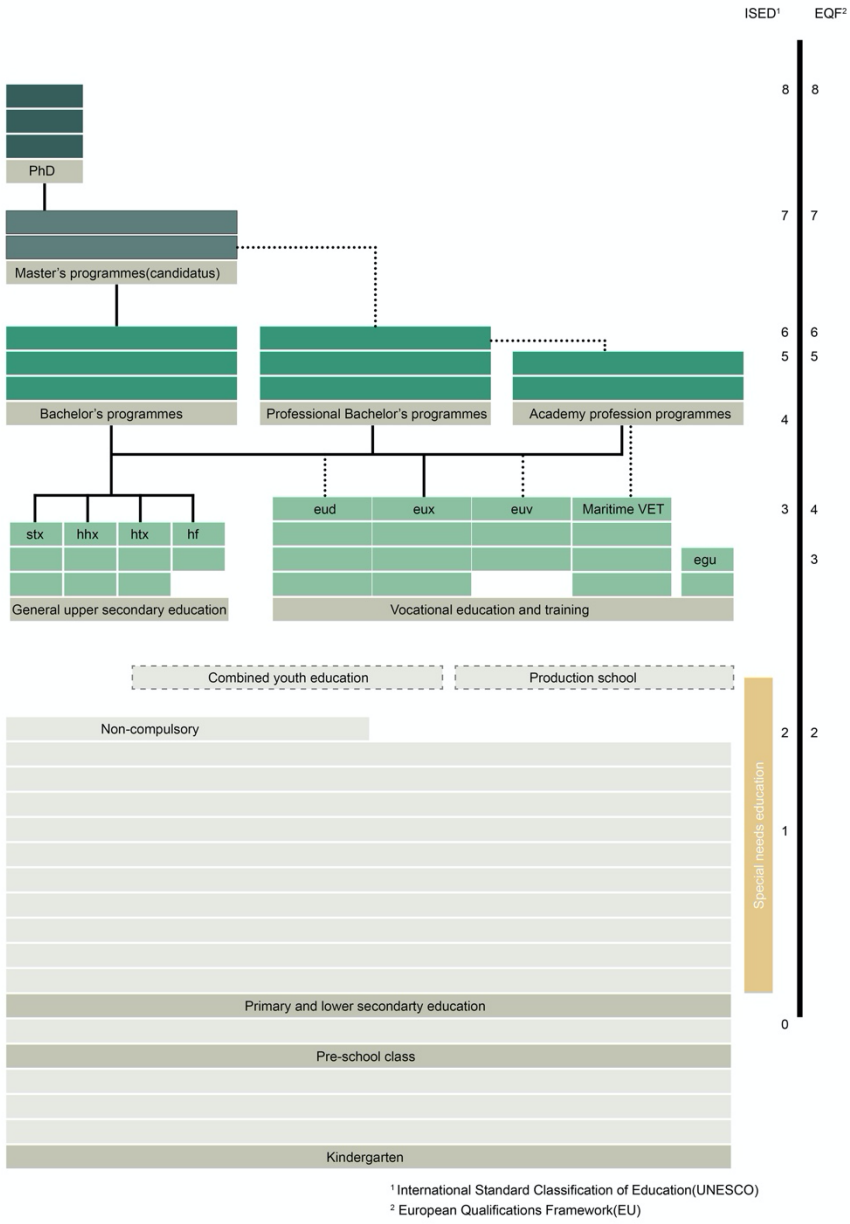

Figure 5. The Danish Education System.
Note: this figure is derived from the booklet *The Danish Education System* (Ministry of Higher Education and Science, 2016).

Higher education in Denmark can be divided into three main types of programme: academy profession programmes, professional bachelor's programmes, and university-level programmes. The academy profession programmes offer short-cycle education oriented towards a specific profession. The duration of these programmes varies from 1½ years to 2½ years and combines theoretical studies with a practice-oriented approach. The professional bachelor's programmes offer professionally oriented medium-cycle educations lasting typically 3½ years. The education programmes are primarily aimed at the public welfare and health sector, the largest such programmes being for childcare workers, elementary schoolteachers, nurses and midwives, and mechanical engineers.

The university-level programmes are provided at the universities, which offer research-based study programmes. These programmes can be divided into three categories: the bachelor's programme, the master's programme and the PhD. The bachelor's degree is awarded after three years and qualifies for further studies or a professional career. The vast majority of students choose to continue on to a two-year master's programme. A master's degree qualifies students for a professional career and for further academic work in the three-year PhD programme.

4.2. Historical development of education in Denmark

In this section, I will provide an overview of the historical development of educational attainment in Denmark in two separate sections, one on the historical development of educational choices and attainment by young people, and secondly choice and attainment in higher education.

Youth education

Figure 6 shows the historical development of youth education attainment in Denmark between 1982 and 2019. The figure reveals two key stylized facts of the development in Danish youth education. First, the share of young people without a post-compulsory education fell by half between 1982 and 2019. Secondly, the most popular choice of upper secondary education has shifted from VET to general upper secondary education in the past four decades.

In Denmark today, about 20 percent of elementary school leavers do not complete an upper secondary education before the age of 25. Instead they typically enter the labour market as unskilled workers with no further education.

Compared to other OECD countries, the share of young adults 25 to 34 years old with education below upper secondary education in Denmark is relatively high. The OECD average was 15 percent in 2019, and the average for European countries 13 percent (OECD, 2020). However, the high share is equal to the share in Denmark's neighbouring Nordic countries, Norway and Sweden.

The high share of younger adults without any qualifying education is a political concern. For more than a decade, there has been a politically stated objective that more than 95 percent of each birth cohort should complete a youth education (Regeringen, 2011). Despite the ongoing political focus on raising the share of elementary school leavers who pursue further education, the share of those with no post-compulsory education has been relatively constant during the past twenty years, as can be seen in Figure 6.

However, the choice between general upper secondary education programmes and VET programmes has changed dramatically among young people in Denmark. Figure 6 shows that about 57 percent of a birth cohort completed a general upper secondary education in 2019. Historically, in 1982 only about 23 percent of elementary school leavers completed a general upper secondary education. Conversely, the share of young people who completed a VET programme fell from 32 percent to 16 percent, between 1982 and 2019.

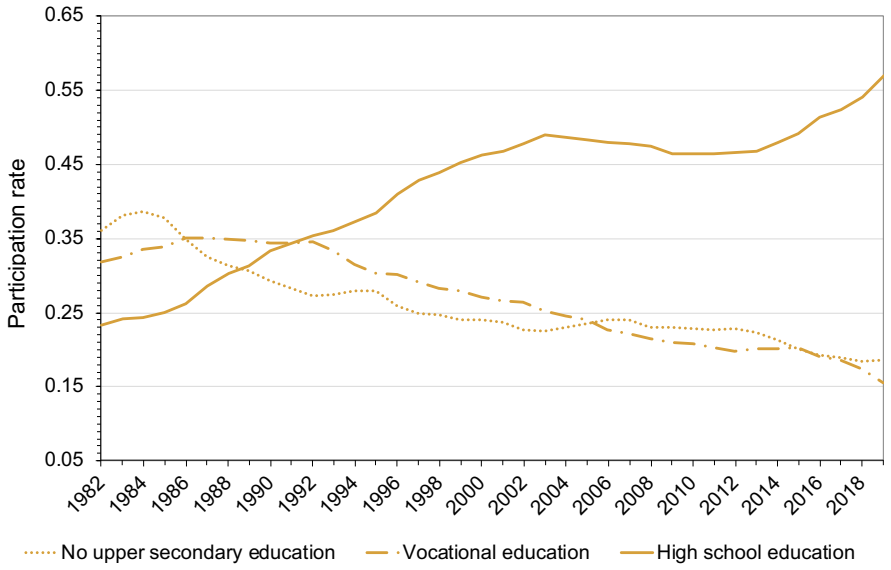

Figure 6. Completion of youth education in Denmark, 1982 – 2019.

Note: the figure shows first completed youth education at age 24. Calculations are based on a full population of 24-year-olds for each year. In cases where a person completes two different youth education programmes before turning 24, first education is used. Individuals for whom there is no information on their education are not shown on the figure. Therefore, the numbers do not add up to 100 percent.

There are several possible explanations why general upper secondary education has become the most popular choice of upper secondary education. General upper secondary education provides access to a wide range of (academic) higher education programmes and disciplines. VET programmes, conversely, are occupation-specific and are directed to a single profession. This implies that choosing a VET programme forces young people to make a choice about their future profession at the time they enrol into post-compulsory education. If, on the other hand, elementary school leavers choose a general upper secondary education, they are able to postpone their choice of profession.

General upper secondary education also has the reputation of offering students a good social life, as the gender distribution in general upper secondary education is almost equal, whereas VET programmes are typically dominated either by women (such as the hairdresser programme) or men (such as the carpentry programme). These factors contribute to general upper secondary being the first choice for the majority of young adults after secondary school, even among those individuals who have no intention of pursuing a higher education. About 73 percent of students who complete an upper secondary education continue on to higher education within two years of graduation (Sørensen, 2021)

Higher education

Figure 7 shows the development of higher education attainment between 1982 and 2019 among the full population of 35-year-old individuals resident in Denmark. The figure shows that the share of those with a university-level education has risen significantly during the last four decades, from about 5 percent in 1982 to 22 percent in 2019. The remaining higher education programmes have also increased, though at more modest rates. In total, the share of those with higher education among the 35-year-olds increased from 22 percent in 1982 to 50 percent in 2019.

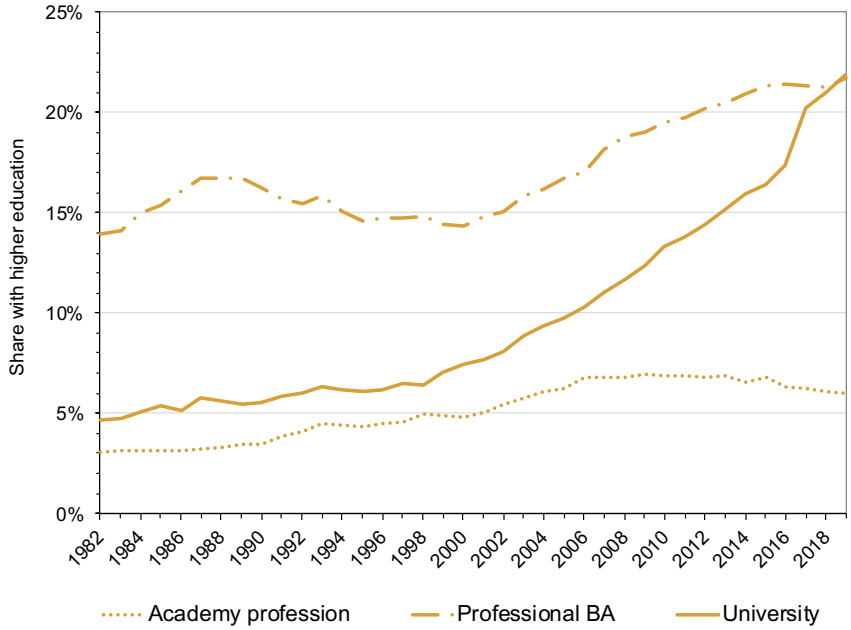

Figure 7. Completion of higher education in Denmark, 1982 – 2019.
Note: the figure shows the highest completed tertiary education among the 35-year-olds. Based on own calculations of Statistics Denmark's administrative register of education (UDDA).

According to the OECD, the share of younger adults with a higher education increased by roughly ten percentage points between 2009 and 2019 across all OECD countries (OECD, 2020, p. 39). On average in OECD countries, 45 percent of the 25–34-year-olds obtained a higher education qualification in 2019. The share in Denmark is approximately five percentage points higher than the OECD country average and at the same level as Norway and Sweden.

Total public expenditure on educational institutions is substantial in all OECD countries. In 2017, Denmark spent approximately 3.7 percent of its GDP on primary, secondary and upper secondary education, which was about 0.6 percentage points above the OECD average (OECD, 2020, p. 287). Taking into account public spending on higher education, Denmark spent approximately 1.6 percent of its GDP. In comparison, the OECD average was 1.0 percent of total GDP, ranging from 0.5 percent of GDP in Japan and Luxemburg to 1.8 percent in Norway.

Total public expenditure on education in Denmark is therefore substantial and above the OECD average. Education in Denmark is provided by the state

and requires no payment from the student. This universal access to education is a cornerstone of the Danish welfare state and one of the key arguments ensuring public support of the high tax levels. In this regard, it is essential that all young adults who wish to pursue education have the possibility to attend. Hence, it is also essential to map out the regional dispersal of education in Denmark and evaluate the degree of regional accessibility for all individuals. The next section will therefore describe the geography of education institutions in Denmark.

4.3. The geography of education institutions

In this section, I will briefly explain how new educational institutions can emerge in new locations and close down in old ones. Furthermore, I will clarify who has the authority to decide where education institutions are located and whether they are relevant to Danish society or not. Finally, I will provide a brief overview of the current location of upper secondary and higher education in Denmark and their geographical accessibility for young students. First, I will describe the geography of upper secondary education in Denmark, followed by the geography of higher education institutions.

Youth education

Youth education in Denmark, or general upper secondary schools and VET schools, is provided by independent and self-governing institutions. The vast majority of the institutions are state institutions that are fully funded by the state. Only about 10 percent of the upper secondary schools are self-governing private institutions. However, the fees for students in private schools are relatively small[10] because these private institutions are heavily subsidised by the government.

The opening and closure of institutions

Youth educational institutions have the right to merge with another institution or to divide. Any decision to merge or divide a state institution is made by the institution's board of directors, subject to approval by the minister (Ministry of Children and Education, 2019). The minister can decide to establish a new

[10] The annual fee at a private general upper secondary educational institution in central Copenhagen (Det Frie Gymnasium) was DKK 18,000,- in 2019, equivalent to approximately Euros 2,400 annually.

institution if it satisfies an expected need for a certain education institution. Likewise, the minister can decide to close an upper secondary school if the ministry considers there is no need for it anymore.

The creation of a new general upper secondary school usually springs from a wish on the part of the present institution. If that institution wishes to split into several new study places, it needs to apply the ministry for permission to create a new school. The procedure for authorization if such a move should involve the Regional Council, which gives its own recommendation about the geographical location of the new youth education institution.

Location of upper secondary schools

Figures 8 and 9 show the geographical location of upper secondary schools in Denmark. There are large numbers of both general upper secondary schools and VET schools located all over the country, which are widely geographically dispersed. In general, general upper secondary schools are more geographically dispersed than the VET schools. Thus, the average distance from a young 15-year-old's home to school is about a kilometre less to the nearest general upper secondary school than to the nearest VET school. On average, the distance from home to the nearest upper secondary school was approximately 6.5 km in 2014 among the total population of elementary school leavers. Hence geographical access to upper secondary schools in Denmark is generally good.

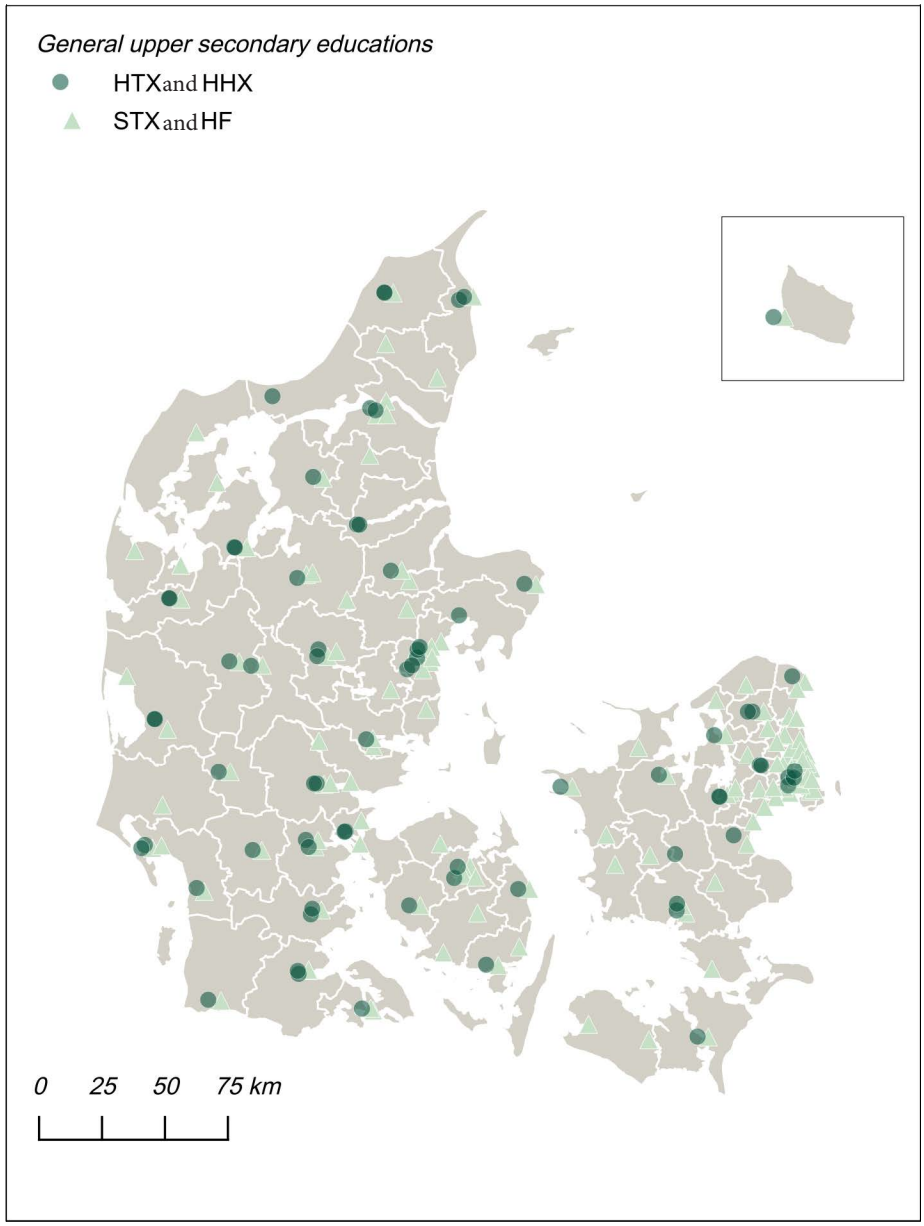

Figure 8. Location of general upper secondary schools in 2020

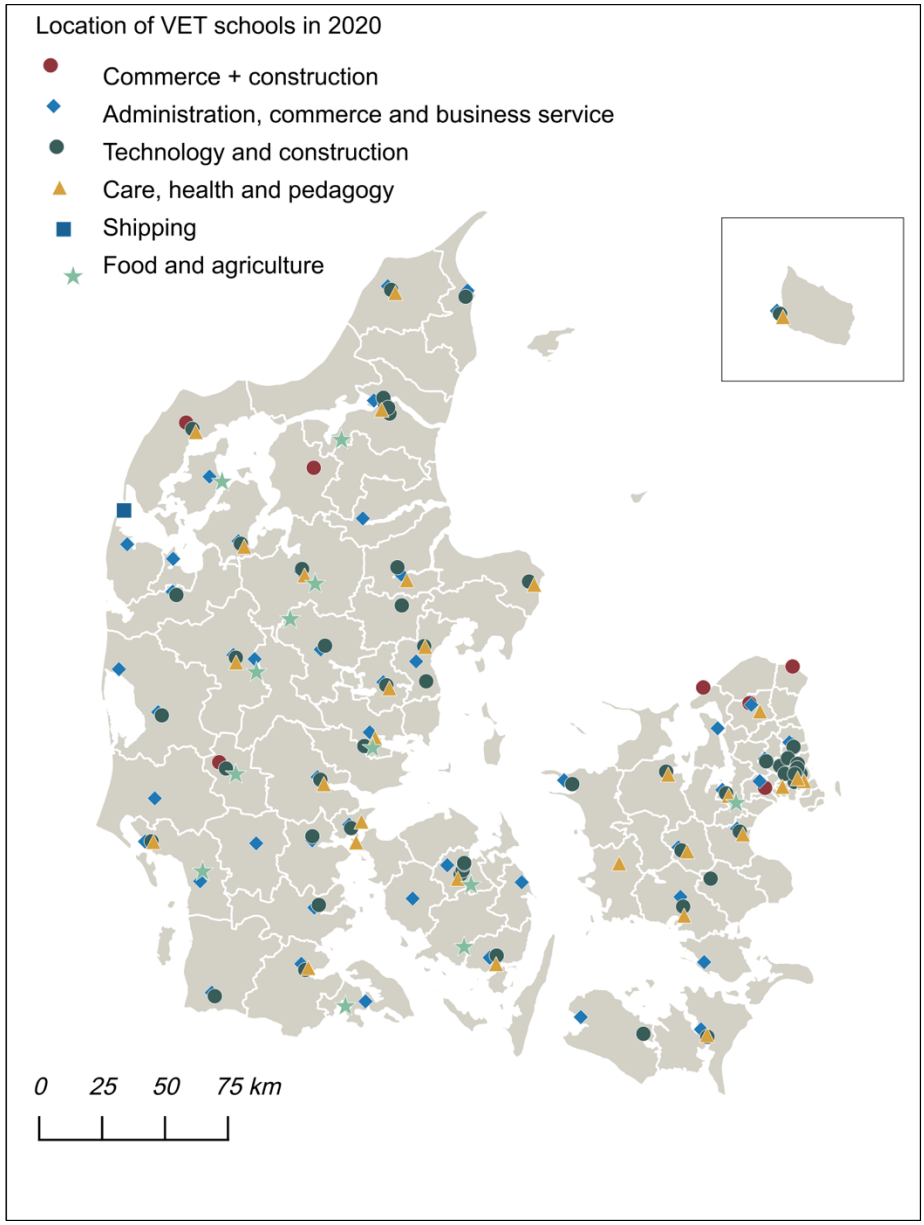

Figure 9. Location of VET schools in 2020

Higher education

In Denmark, the main HEIs, namely the universities, university colleges and academies of professional higher education, are independent and self-governing institutions. They are regulated by the Ministry of Higher Education and Science and are fully state-funded.[11]

Opening and closure of HEIs

The establishment of a new HEI should be approved by the Minister (Ministry of Higher Education and Science, 2013). Usually, the initiative to create a new HEI comes from the established HEIs. New HEIs can emerge in two ways: either the HEI creates a new place of study in a new geographical location, or it creates a place of study in one of its existing institutions in the same geographical location. The second option is obviously the most common, as it is expensive and time-consuming to establish buildings and facilities in a new location.

Whichever form of establishment is chosen, the HEIs need to submit an application to the Danish Accreditation Institution, which evaluates all Danish HEIs and provides formal authorisation to establish new educational institutions.[12] A new HEI should meet two overall requirements to be approved: it should fulfil certain international quality norms and should have relevance for Danish society. The Danish Accreditation Institution evaluates the application and writes a report with recommendations to the Accreditation Council, which decides whether to approve or reject the application. Finally ministerial approval must be sought.

Regarding the closure of an existing HEI, there are two ways in which this can happen. One way is for the HEI to decide to close down a study place, usually because of small and declining application numbers. Another way is when the Accreditation Council considers the quality of the education at the HEI to be so poor that is does not live up to international standards. Ultimately

[11] There are a few exceptions, such a some HEIs that offers educational programmes in the fine and performing arts. These institutions are regulated by the Ministry of Culture.

[12] The procedure of endorsement is equal for all types of higher education (academy profession programmes, professional bachelor's programmes and university-level programmes) and follows a general procedure (Ministry of Higher Education and Science, 2013).

this can lead to the HEI's closure. However, at the time of writing there have been no cases of this happening in Denmark.

Location of HEIs

In Denmark, HEIs are widely geographically dispersed. Figure 10 below shows the geographical location of higher education institutions in 2020. The university colleges and the academies of professional higher education are included under the same category in the figure, namely "Colleges". The universities have their own distinct category. While historically universities have been located in the four largest cities in Denmark, the university colleges and the academies of professional higher education have had a broader geographical dispersal. By law the university colleges are obliged to provide their local region with education and to ensure access to higher education in peripheral areas. This requirement is rooted in their obligation to provide "public welfare" education (such as elementary school teachers, nurses and childcare workers) in every corner of the country.

Overall, the geographical dispersal of HEIs is good in Denmark, with only a few areas having poor geographical access to them. The average distances from home to the nearest college and university for upper secondary school leavers were 11.6 km and 25.4 km respectively in 2014 (Sørensen, 2021). Furthermore, 90 percent of upper secondary school leavers had access to a HEI within 30 km.

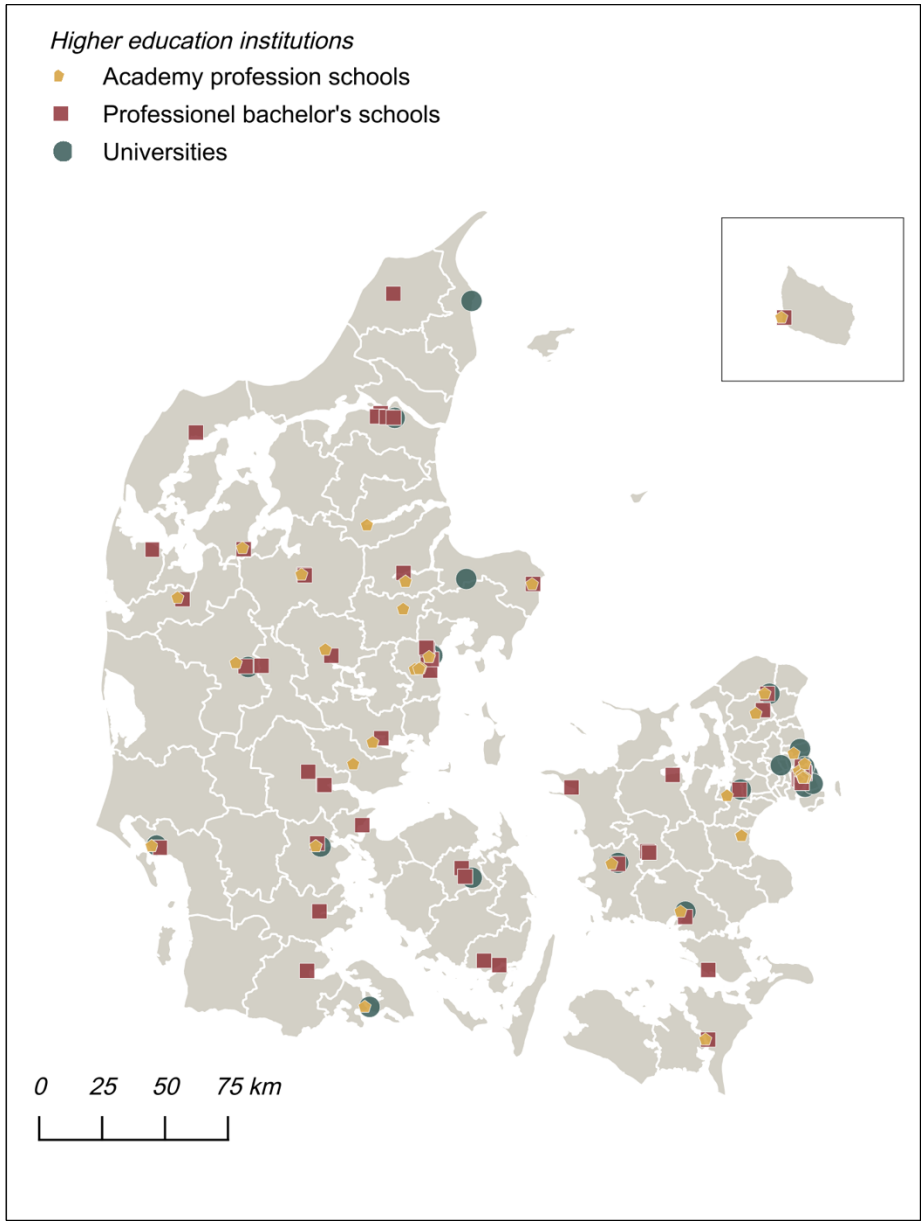

Figure 10. Location of higher education institutions in Denmark

4.4. Geography of human capital

This section considers the regional disparities in youth education and higher education in Denmark. First, I describe geographical variation in the choice of upper secondary education by the young before describing the geographical disparities in the proportion of residents who hold a university degree.

Regional disparities in youth education

Figures 11, 12 and 13 show the spatial variations in the share of 21-year-olds with no qualifications beyond compulsory schooling, a general upper secondary education, and a VET education respectively. Figure 11 shows that the share of young adults (age 21) with no qualifications beyond compulsory schooling varies substantially by municipality of upbringing, defined at the municipality where the individual was resident at the age of fifteen. The proportion without post-compulsory education ranges from 13 percent to 36 percent of elementary school leavers in some municipalities, typically those located in the more rural regions of Western Zealand and the three islands south of Zealand: Stevns, Falster and Lolland.

There is therefore a temptation to infer that the relatively low participation rates in the peripheral areas are caused by less geographical accessibility to post-compulsory education. However, there are also exceptions to this association. Copenhagen municipality and some of the suburban municipalities to the west and south of Copenhagen have relatively low participation rates in post-compulsory education, despite the fact that they are located close to or actually in urban centres with good access to upper secondary education. Common to all these municipalities is the fact that they also have a relatively high share of lower income families, families with lower educational levels and immigrants. Thus the map suggests a "socio-geographical" component, like that found in international studies of education and social heritage (Dickerson & McIntosh, 2013; Gibbons & Vignoles, 2012).

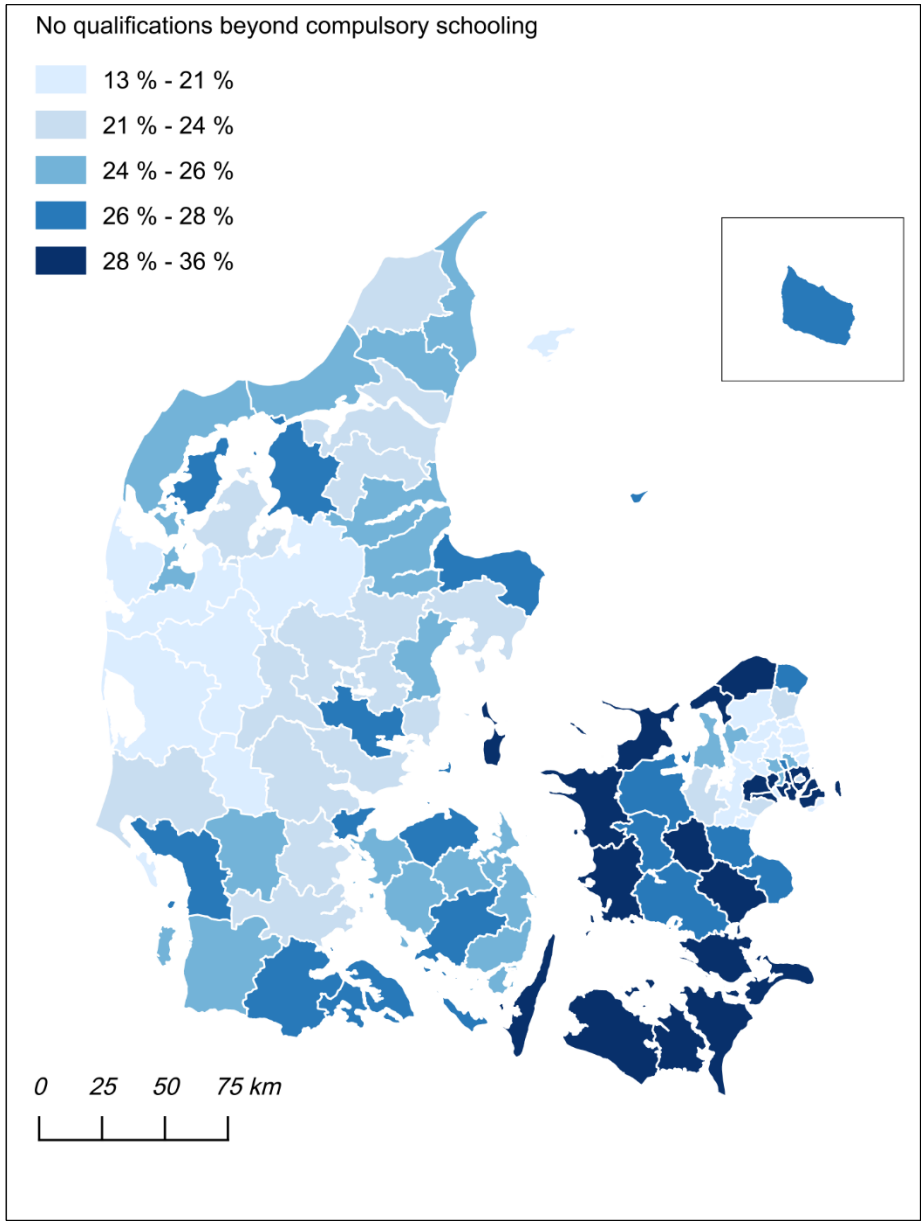

Figure 11. Share of the population without any post-compulsory education at age 21 by municipality of upbringing at age 15, 2013 – 2018. The figure illustrates the average educational attainment of six birth cohorts born between 1992 and 1998 who were 15 years old between 2007 and 2012. Hence the figure shows their educational attainment between 2013 and 2018.

Figure 12 illustrates the geographical differences in the choice of general upper secondary education by municipality of upbringing. In the Capital region, namely the municipalities north and west of Copenhagen, and in the second largest city, Aarhus, average participation in general upper secondary education is above 63 percent. In the country's most wealthy municipalities north of Copenhagen the participation rates rise to 83 percent. In the municipalities that are located in the more peripheral areas of Denmark, especially the western part of Zealand and Funen, the participation rates are between 46 percent and 55 percent.

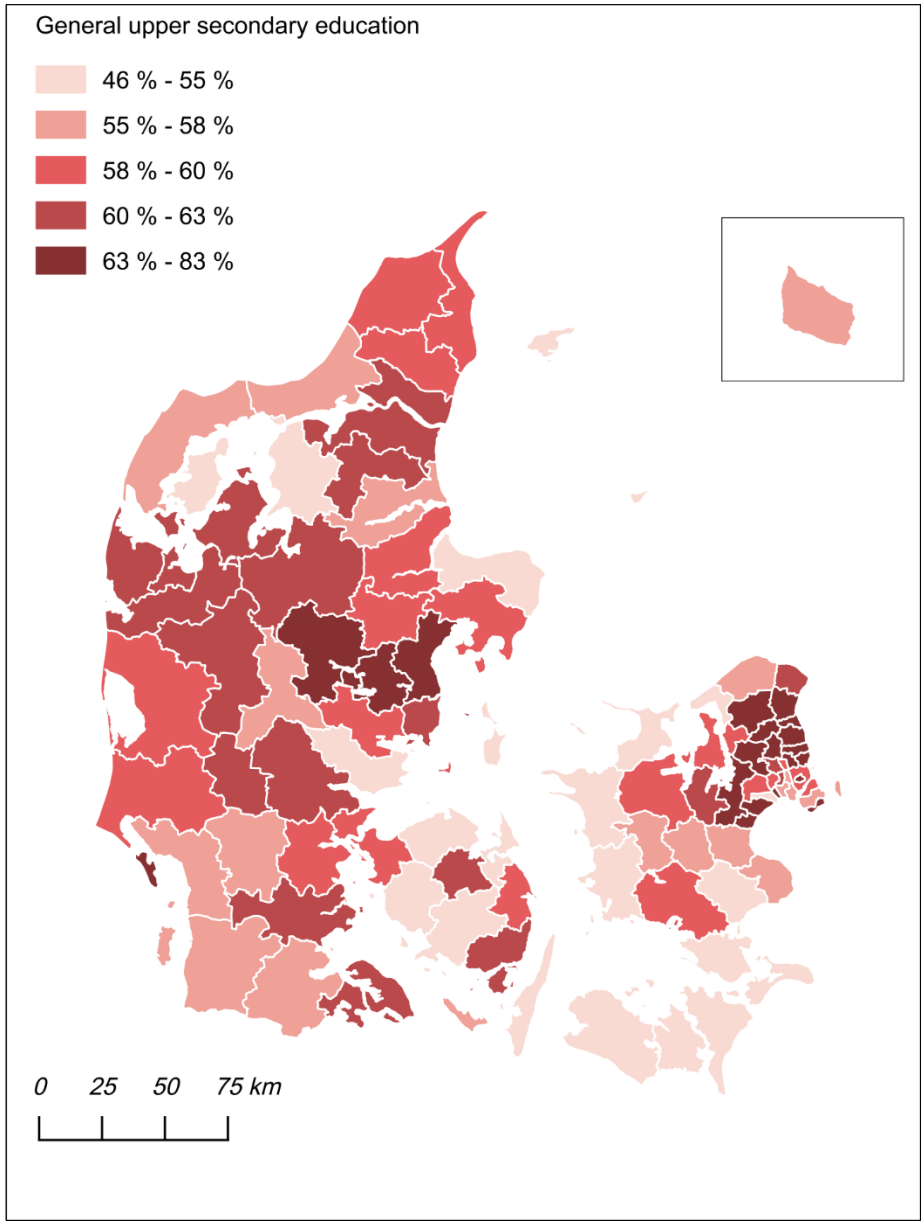

Figure 12. Share of the population who completed a general upper secondary education at age 21 by municipality of upbringing at age 15, 2013 to 2018. The figure shows the average educational attainment of six birth cohorts born between 1992 and 1998 who were 15 years old between 2007 and 2012. Hence the figure shows their educational attainment between 2013 and 2018.

Figure 13 shows the participation rate in VET programmes by municipality upbringing. The spatial pattern is nearly an opposite mirror image of the pattern we saw in Figure 12. The highest VET participation rate is seen in the peripheral areas in the western and northern parts of Jutland, which have a long tradition of small and medium-sized production firms and farming. In these regions roughly between 18 to 26 percent of the 21-year-olds completed a VET programme. In contrast, only 3 to 12 percent completed a VET programme in Denmark's two largest cities and in the municipalities north of Copenhagen. In sum, the geographical disparities in the choice of upper secondary education are huge. A comparison of the two maps reveals a spatial pattern in which VET programmes are most attractive in the rural and peripheral areas, and general upper secondary education programmes most popular in urban and suburban areas.

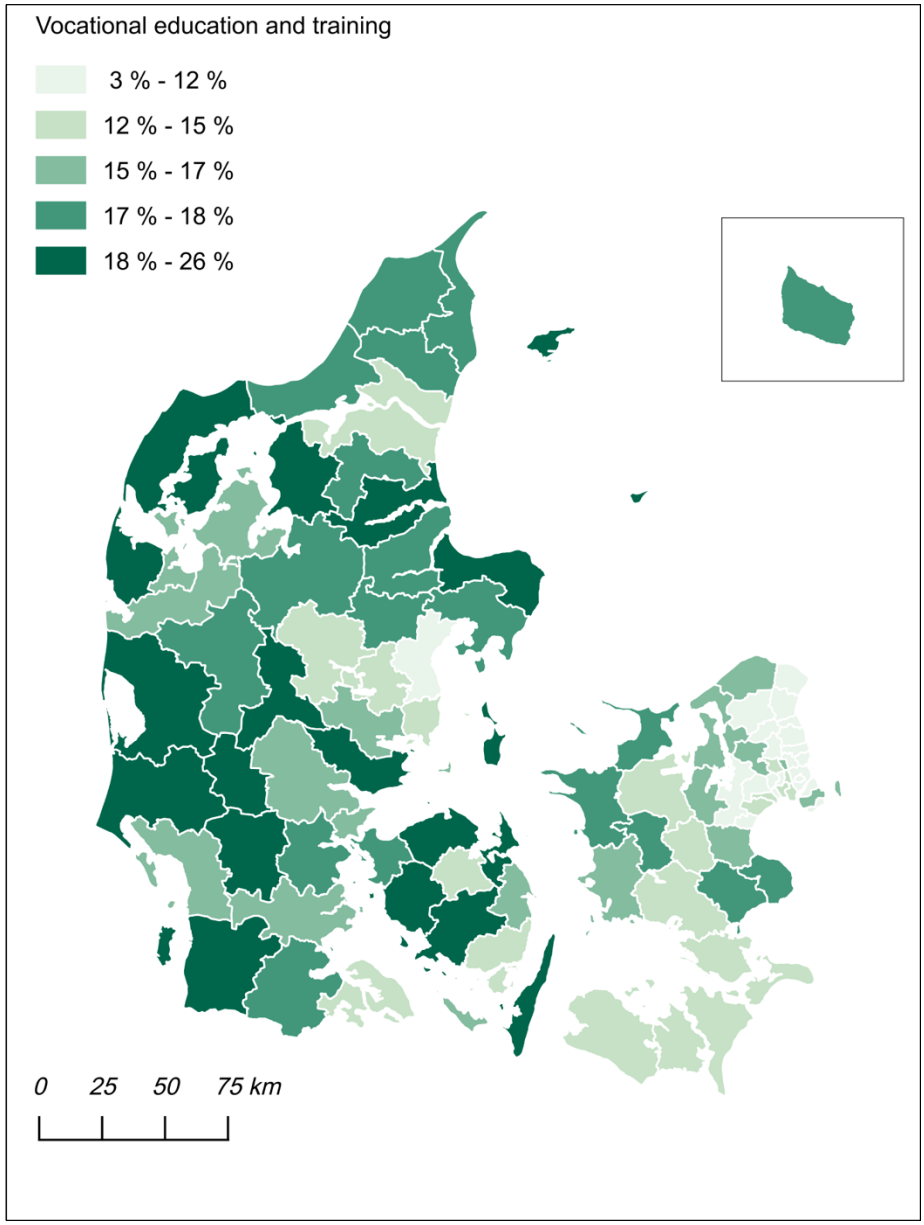

Figure 13. Share of the population who completed a vocational education and training (VET) programme, at age 21 by municipality of upbringing at age 15, 2013 – 2018. The figure shows the average educational attainment of six birth cohorts born between 1992 and 1998 who were 15 years old between 2007 and 2012. Hence the figure shows their educational attainment between 2013 and 2018.

Geographical disparities in human capital

In this section, I describe the geographical disparity of high levels of human capital in the labour force, namely the share of the adult population who held a university degree. Figure 14 shows the share of 15 to 69-year-olds who held a university degree in 2018 in each municipality in Denmark. There is huge spatial variation in university attainment, ranging from 2 to 5 percent of the population in some peripheral municipalities to approximately 30 percent in some of the municipalities in the Copenhagen city region. We see a spatial pattern in which the highly educated are concentrated in the four biggest university cities in Denmark: Copenhagen, Aarhus, Odense and Aalborg, and their neighbouring municipalities. The Copenhagen labour market is widely dispersed in spatial terms, and it is common for individuals with high levels of human capital who work in Copenhagen live in one of the municipalities north of the city and commute to work.

Overall, the finding that highly educated individuals are concentrated in the largest urban areas of Denmark is not surprising, as it is a pattern that is found all over the world. Education level is found to be one of the most reliable predictors of urban growth (Glaeser, 2011). In addition, the productivity of an urban area increases significantly with city size, but only if the city has a high level of human capital; otherwise it does not (Glaeser, 2011). Thus, in the next section I will consider the descriptive association between the increase in high levels of human capital and the increase in economic prosperity regionally.

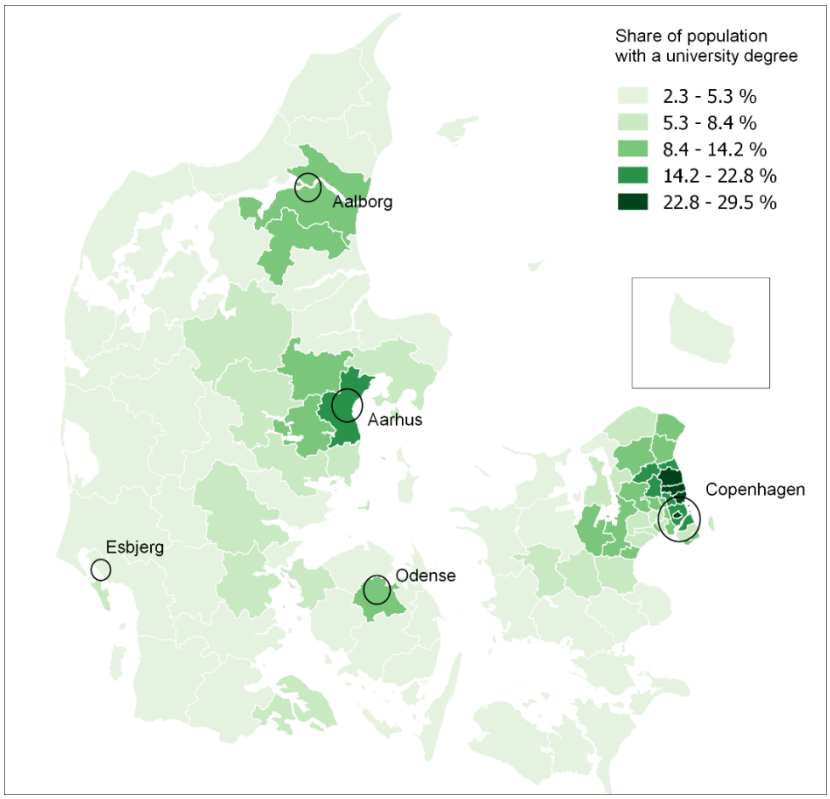

Figure 14. Share of highly educated by Danish municipalities, 2018
Source: the figure shows the share of 15 to 69-year-old adults who held a university degree in 2018. The numbers are my own calculations based on administrative data from Statistics Denmark.

4.5. Human capital and regional growth

This section describes the development of regional economic prosperity and human capital in Denmark between 1993 and 2016. First, I assess the regional differences in economic prosperity in 2016. I use GDP per head as a summary indicator of economic prosperity. Figure 15 shows the level of GDP per head compared to the national average in eleven Danish regions (NUTS 3 level). Following the work of Iammarino, Rodriguez-Pose, and Storper (2018), I distinguish between four categories of economic development: 1) very high GDP per head region (VH), with 150 percent of the national average or greater 2) high GDP per head region (H), with 120 to 149 percent of the national average; 3) medium GDP per head region (M), with 75 to 119 percent of the

national average; and 4) low GDP per head region (L), with less than 75 percent of the national average.

Figure 15 shows that there were substantial differences in the level of economic prosperity across the Danish regions in 2016. The VH category includes the "Greater Copenhagen area", where GDP per head was about 160 percent of the national average. The H category includes "Copenhagen city", where GDP per head was 143 percent of the national average. In the remaining nine regions of Denmark, GDP per head was close to or below the national average. The exact figures are given in Table A1 in the appendix. It is evident from Figure 15 that Copenhagen and the surrounding regions account for the bulk of Danish GDP, which is in line with the thesis of the agglomeration benefits of larger cities (Glaeser, 2011).

In that connection, it is striking that the numbers provides no descriptive evidence of agglomeration effects in Denmark's second and third largest cities, Aarhus and Odense, located in the Eastern Jutland and Funen Regions respectively. However, the figures in Table A1 in the appendix show that the province of Southern Jutland, located next to Germany, has a GDP per head just above the national average, which could indicate some positive economic spillover effects from the huge German economy in the south.

A comparison of regional patterns in respect of high levels of human capital (in Figure 14) and of economic prosperity (in Figure 15) reveals a striking resemblance in the Greater Copenhagen region, which has a high share of both high human capital and high GDP per head. However, the same pattern is not seen around Aarhus and Odense, the second and third largest cities. This might be due to the fact that the geographical region category is too broad geographically to capture the effect of having higher shares of highly educated in city centres. Another possibility could be that these two cities are too small (approximately 350,000 and 200,000 inhabitants respectively) for agglomeration effects to be present. In the next section, I will examine whether the regional disparities in economic prosperity have been constant over time, have converged or have diverged since the global financial crisis of 2008.

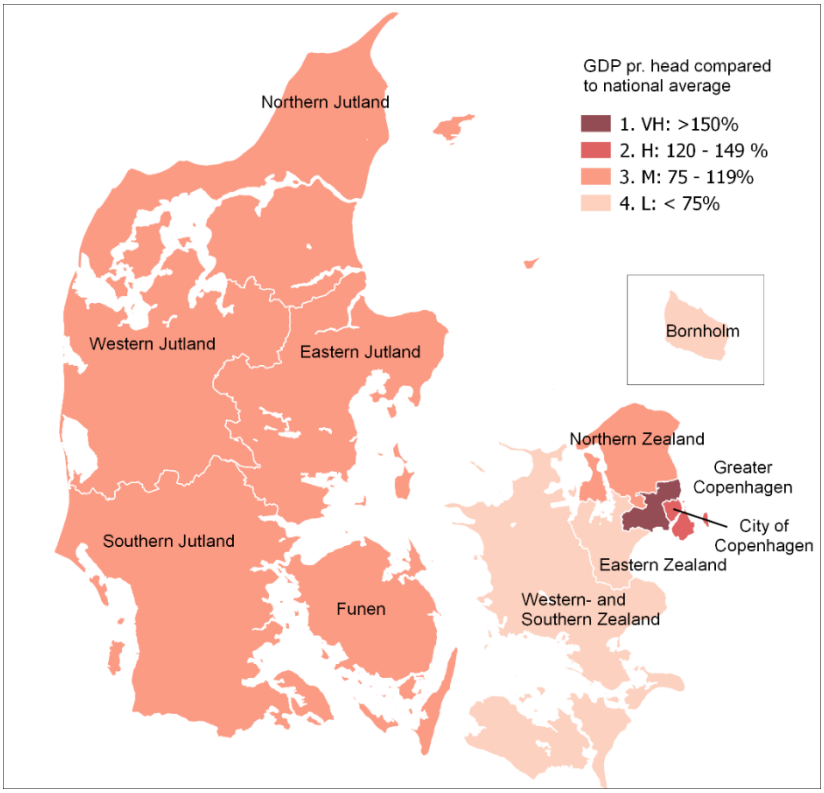

Figure 15. Classifying Danish NUTS 3 regions according to GDP per head compared to the national average, 2016.
Source: own calculations based on administrative data from Statistics Denmark

Prosperity and human capital: concentration over time

Figure 16 shows the historical development in GDP per head between 1993 and 2016 in five larger geographical regions in Denmark (NUTS 2-level). I have grouped the eleven smaller regions (NUTS 3-level) into five larger regions (NUTS-3 regions) for reasons of simplicity. Figure 16 confirms the regional pattern we saw on Figure 15, namely that economic development in the Capital Region is above the national average and in the other regions is below the national average. Furthermore, we see that the "rank order" of the regions has remained unchanged since 1993.

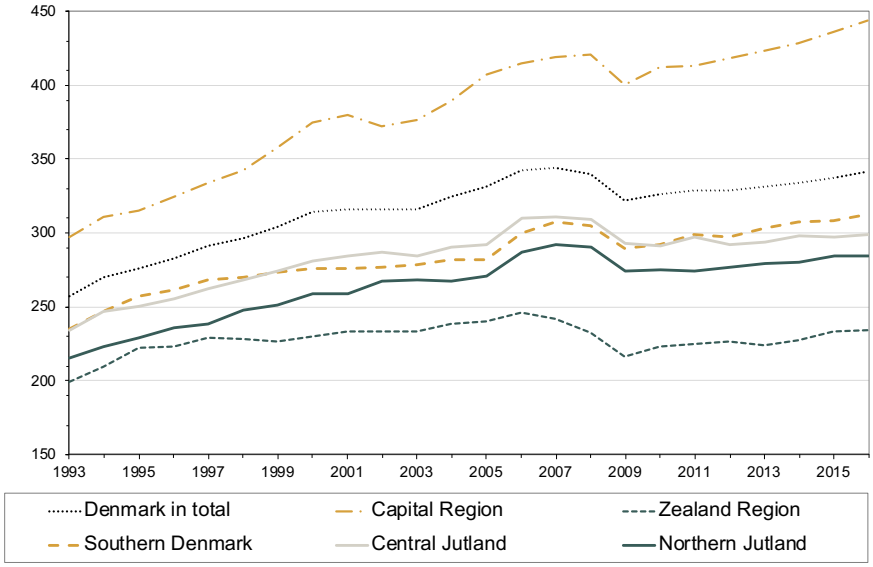

Figure 16. Development in GDP per head in the five geographical regions in Denmark (NUTS 2) between 1993 and 2016. GDP per head is quoted in DKK 1,000. Prices as of 2010.

Figure 17 shows the relative change in GDP per head between 1993 and 2016 in percentage points for each NUTS 2 region in Denmark (prices as of 2010). All five regions have experienced growth in GDP per head, but some regions have had higher growth than others. The Danish economy experienced economic recovery in the 1990s, during which period the percentage growth in GDP across the Danish regions was almost the same. The regional divergence started in around 2002, when Denmark experienced an economic recession, like other economies globally. However, the largest regional divergence took place one year before the financial crisis of 2008 and the following recession period.

Overall, regional economic disparities increased in Denmark between 1993 and 2016. The Capital region, which produced the highest share of GDP per head back in 1993, increased GDP per head by approximately 50 percentage points between 1993 and 2016. By comparison, the Zealand Region, which had the lowest GDP per head in 1993, increased its GDP per head by approximately 18 percentage points. The figure shows that the Zealand Region never came back to the same economic level as before the financial crisis of 2008.

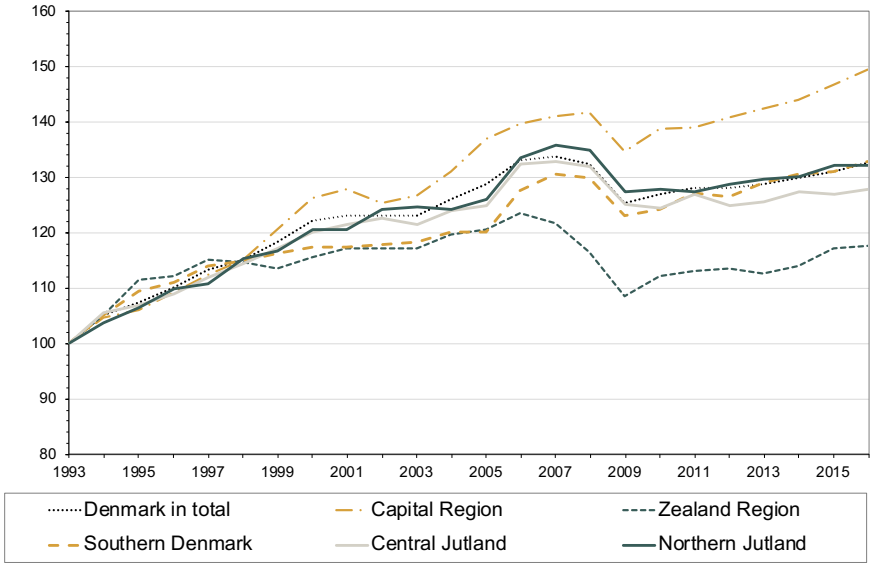

Figure 17. Changes in GDP per head in Danish NUTS 2 regions in percentage points between 1993 and 2016. GDP per head is quoted in DKK 1,000. Prices as of 2010.

Finally, I move on to a description of the growth in high levels of human capital in the Danish regions and compare this development with regional economic growth. Figure 18 shows the regional growth in individuals between 15 and 69-year-olds who held a university degree between 2006 and 2018. The percentage increase was greatest in the Capital Region (approximately 90 percent) and smallest in the Zealand region (approximately 55 percentage points). It is striking that the growth in human capital in the Zealand region is significantly below the other four regions, where the differences is minor. Thus, regional development in respect of human capital mirrors the patterns of economic growth seen in the Figures 17 and 16.

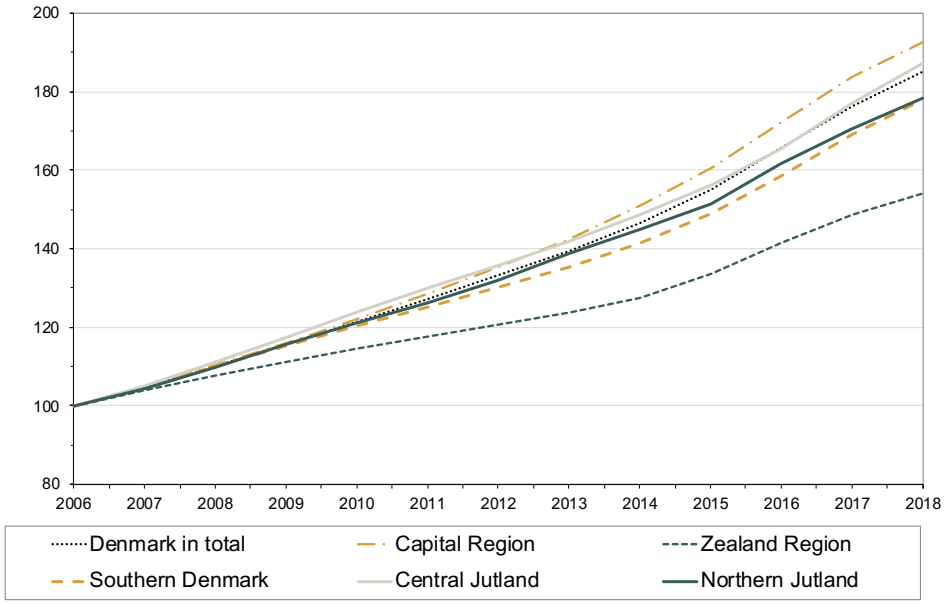

Figure 18. Regional development in respect of share of population with a university degree, 2006 – 2018 (2006 = index 100).
Note: own calculations based on administrative data from Statistics Denmark

4.6 Summary

This chapter has described the Danish educational system and the level of educational attainment from an international perspective. Furthermore, is has provided stylized facts about geographical access to upper secondary institutions and HEIs, regional disparities in levels of human capital and regional disparities in economic growth. So, what have we learned from this? First, regarding levels of educational attainment, we saw that the share of Danish younger adults (25 – 34 year-olds) with no education above the level of elementary education is relatively high compared to other OECD countries: in Denmark, about 20 percent of younger adults obtained no further education above elementary education in 2019, when the average for all OECD countries was 15 percent.

At the same time, the share of younger adults who completed a higher education was approximately 5 percentage points higher than the OECD average at 45 percent. This suggests that the disparities regarding educational attainment are greater in Denmark compared to the OECD countries, in the

sense that the groups at the bottom and the top of the educational distribution make up a larger share of the younger adult population. Hence, the proportion who obtained a vocational upper secondary education is relatively small from the international perspective.

Secondly, I described the process of the creation and closure of an education institution, and mapped the geographical location of upper secondary institutions and HEIs in Denmark. Only existing HEIs can apply to the minister to establish a new place of study, either at the same location or at a new location. It is evident that the process of creating a new educational institution is both demanding and expensive for the institution concerned, leading to the creation of a new study place at a new location rarely occurring. Mapping the institutions showed that overall few parts of Denmark have poor geographical access to education. The regional dispersal of upper secondary and higher education institutions is relatively good: on average elementary school leavers lived approximately 6.5 km from the nearest upper secondary school, while upper secondary school graduates had approximately 11 km to go to the nearest HEI.

Thirdly, I considered the regional disparities in education and concluded that there were large geographical differences in choosing and completing upper secondary education. The highest VET participation rates were found in rural and peripheral municipalities in western Denmark, while the highest participation in general upper secondary education was seen in the municipalities that surround Copenhagen and Aarhus. Looking at the adult population, we saw that individuals with a university degree were concentrated in the largest urban areas.

Fourthly, the figures also showed large regional differences in economic development. The economic growth centre was located around the capital, with GDP per head, about 50 percentage points above the national average, and the remaining regions were lagging behind, with GDP per head below the national average. Regional economic development between 1993 to 2016 largely followed the same growth rate up until the first economic recession in 2002. After 2002 growth rates diverged so that the regions with the highest GDP per head experienced the largest growth rates. One year before and in the years after the financial crisis of 2008, the divergence between the regions increased. The growth in the Capital Region clearly fell off, and the Zealand Region in particular fell behind.

Several studies have emphasized the importance of high levels of human capital for regional economic growth and development (Gennaioli, La Porta, Lopez-de-Silanes, & Shleifer, 2013; Glaeser, 2011; Iammarino et al., 2018). There is evidence that a skilled labour pool generates knowledge spill-overs and externalities, thus increasing productivity (Moretti, 2004). When compare the regional concentration and growth of high levels human capital in Denmark is compared with the regional magnitude and development of economic growth, there are clear associations between the two factors. Thus, this simple comparison at the macro-development level to a large extent confirms the theoretical and empirical evidence described in the literature.

4.7 References

Dickerson, A., & McIntosh, S. (2013). The Impact of Distance to Nearest Education Institution on the Post-compulsory Education Participation Decision. *Urban Studies*, *50*(4), 742–758.

Gennaioli, N., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2013). Human Capital and Regional Development. *The Quarterly Journal of Economics*, *128*(1), 105–164.

Gibbons, S., & Vignoles, A. (2012). Geography, choice and participation in higher education in England. *Regional Science and Urban Economics*, *42*(1–2), 98–113.

Glaeser, E. (2011). *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier.* New York: Penguin Press.

Iammarino, S., Rodriguez-Pose, A., & Storper, M. (2018). Regional inequality in Europe: Evidence, theory and policy implications. *Journal of Economic Geography*.

Ministry of Children and Education. *Bekendtgørelse af lov om institutioner for almengymnasiale uddannelser og almen voksenuddannelse m.v.*.LBK nr 596 (2019).

Ministry of Higher Education and Science. *Forslag til lov om akkreditering af videregående uddannelsesinstitutioner*. Lovforslag nr. L137 (2013).

Moretti, E. (2004). Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, *121*, 175–212.

OECD. (2020). *Education at a Glance 2020: OECD Indicators*. OECD.

Regeringen. (2011). *Et Danmark, der står sammen*. Regeringsgrundlag 2011.

Sørensen, E. S. (2021, April). *Gaps in Higher Education Choice and Participation: Understanding the Role of Distance to Higher Education*, Copenhagen.

The Ministry of Higher Education and Science. (2016). *The Danish Education System*.

4.8 Appendix

Table A1. Classifying Danish provinces, according to GDP per head compared to the national average, 2016

Province	GDP per head	GDP per head compared to the national average in percentage	Group
Copenhagen City	490	143.70	H
Greater Copenhagen	536	157.18	VH
Northern Zealand	274	80.35	M
Bornholm	251	73.61	L
Eastern Zealand	234	68.62	L
Western and southern Zealand	233	68.33	L
Funen	264	77.42	M
Southern Jutland	345	101.17	M
Eastern Jutland	290	85.04	M
Western Jutland	318	93.26	M
Northern Jutland	284	83.28	M

CHAPTER 5

Conclusion

In this chapter, I discuss the main findings of the four papers that make up this thesis. First, I provide a short recap of the research questions in Section 5.1. Then, in Section 5.2, I will answer the research questions and summarise the main finding of the papers. I discuss the policy implications of this study in Section 5.3. and draw three overall conclusions in Section 5.4. Finally, Section 5.5 provides directions for further research.

5.1 Research goal and research questions

The overall aim of this thesis is to clarify to what extent the creation and relocation of new educational institutions will reduce the regional disparities in human capital and regional development. More specifically, the thesis answers two central research questions, which can help inform the debate about the location of institutions:

1. Does proximity to educational institutions increase the creation of human capital locally?
2. To what extent do levels of human capital influence inter-regional migration and settlement in urban areas?

Both the regional creation of human capital and region migration patterns could potentially amplify the regional disparities in human capital in several ways. In the next section, I will provide a brief description of the main findings in the thesis by answering these two questions.

5.2 Main findings

Research question 1

To examine whether proximity to educational opportunities plays an important role in participation in education, I analyse distance to youth education and higher education institutions. The results suggest that distance from home to places of education is most important for enrolment at age 15, when young individuals make their first educational choices whether to continue in post-compulsory education. However, the effect of distance are inconsiderable at age

18, when upper secondary graduates must decide whether to continue in higher education or not. Hence, the effect of distance to school varies across the life-cycle.

Both studies show that the influence of distance over enrolment varies significantly with family background, being most profound for young individuals from less-educated families and negligible if one or both parents have a higher education. This suggests that parental background moderates the effect of distance: highly educated parents are likely to have greater expectations regarding their children's educational paths and therefore encourage them to pursue further education, regardless of distance to it. Hence, a reasonable explanation is that the potential barrier of long geographical distance to school, can be overcome when family encouragement and aspirations pushes young people to enter further education.

Research question 2

The relationship between human capital and inter-regional migration is also studied at two distinct times in the life-cycle, namely after the completion of youth education and after graduation from higher education. Overall, the results suggest that human capital plays an essential role in patterns of inter-regional migration and residential settlement. The two studies in the present thesis on inter-regional migration are quite different in the population studied, the methodology and the set-up. I will therefore summarise the findings separately in the following, starting with interregional migration after youth education before moving on to migration patterns after higher education.

The first paper on human capital and inter-regional migration focuses on youth migration from rural to urban areas. First, the study provides descriptive evidence that youth migration from rural to urban areas increased significantly between 1989 and 2017, while out-migration in all other age-groups did not change. Hence, it is relevant to examine the determinants of migration among young people to arrive at a deeper understanding of overall net migration out of rural areas. The results show that completion of a general upper secondary education increases the probability of one moving to an urban area significantly. In addition, the analysis shows that the effect of general upper secondary education on out-migration varies with level of parental education in the sense that young adults from low socioeconomic backgrounds are most affected by their choice of youth education.

The causal evidence suggests that the rise in the completion of general upper secondary education by young adults, from 30 to 60 percent between 1982 and 2019, is a central driver of the increasing youth migration out of rural areas. This suggests that the increase in human capital has shaped the growth of urban areas, as we have seen in Denmark in recent decades.

The second paper on human capital and inter-regional migration examines the graduate retention rates of higher education graduates in Denmark. The sample population comprises newly educated graduates, and we analyse their migration decisions up until 10 years after graduation. The study provides descriptive evidence of how graduate migration is associated with the geographical location of institutions of higher education, educational levels, previous migration, and other individual characteristics. Overall, the results show that most graduates have left the city where they received their degree 10 years after graduation. However, there are large geographical disparities between city sizes and regions. Copenhagen differs significantly from the other cities with HEIs. In Copenhagen about 65 percent of graduates were still living in the city 10 years after graduation, compared to only 22 percent in the smaller towns with HEIs. Based on these results, we conclude that regional location and the size of the local labour market are crucial to the retention of highly educated graduates. Consistent with human capital migration theory, the results suggest that migration is associated with level of education, though the association between field of study and migration is smaller than expected.

Overall, the results indicate that highly educated individuals tend to live in the largest urban areas. If the educational institution is located in the capital, Copenhagen, we find that most of its graduates are still living in the area ten years after graduation. However, if the HEI is located in a smaller town, the majority of its graduates will move away after graduation. This suggests that re-locating HEIs to smaller towns will not necessarily increase the local stock of highly educated individuals substantially, as the majority will most likely migrate to the largest cities afterwards. Among those graduates who moved to study in another town, about 20 percent return to their region of origin within 10 years of graduation. This suggests that a local sense of belonging also influences residential migration patterns, but that the force of attraction to the capital area seems strong and difficult to change. However, the present global Covid-19 pandemic may change this pattern and make smaller towns and rural areas more attractive again. We still need to wait to determine the long-term impacts of the pandemic on residential settlement patterns.

5.3 Implications for policy

As the regional inequalities in economic prosperity have increased in Denmark in recent years, the political tools for slowing down this development are taking up space in the public debate. The geographical location of public institutions, in particular educational institutions, has become a political hot topic in recent years in Denmark. There are several arguments for focusing on the decentralisation of public educational institutions. In this section, I will present the most common arguments and subsequently describe how the papers in this thesis answer, shed new light on or inform these arguments.

First, from a welfare perspective, it is important that citizens have reasonably equal access to services and amenities, including education, despite deciding where to live (Andersen & Etzerodt, 2018). Papers 1 and 3 and Chapter 4 in the present thesis, provides descriptive evidence showing that the current geographical dispersal of both youth- and higher education is relatively good in Denmark and that most young people have access to an educational institution within a relatively short distance. Hence, from a welfare perspective, the present location of both youth and higher education is not a significant problem in Denmark.

Second, a central argument concerns the right to equal opportunities for all citizens, including the opportunity to harness human resources. If distance to education acts as a barrier to participation in education, it becomes both a social problem and a failure to utilize a source of human capital. Papers 1 and 3 address this question. Paper 1 shows that distance to general upper secondary schools has an impact on young adults' decisions over whether to enrol in education after compulsory education, especially among young people from less-educated families. However, distance to higher education has no substantial effect on students' chances to pursue higher education, only on their choice of study programme. Hence, the creation of new HEIs in peripheral areas would most likely not increase enrolment in higher education at all significantly.

A third argument is that the location of educational institutions might affect residential choice and internal migration. Especially among families with children, it can be expected that access to local public schools and upper secondary education is important when deciding where to live. The papers in the thesis do not examine educational institutions directly as a "pull" factor influencing residential choice. However, in studying rural– urban migration among young individuals (Paper 2), it is evident that the location of HEIs in the largest urban areas is implicitly a pull factor among young people who

completed a general upper secondary education. However, as shown in Paper 3, distance to higher education is not a crucial factor in their decision to pursue higher education. Many young individuals apply to enter an HEI *because of* its location in Copenhagen. That is, they want to move to Copenhagen because of the city's amenities, culture, and youth orientation (Florida, 2002) the choice of educational provision being secondary, not the other way around. However, we need more empirical evidence to properly test the hypothesis of whether young people move to the cities for education or for fun.

Fourth, public institutions such as universities can have an impact on local economic development in various ways, among others through knowledge spill-overs to local firms and the creation of human capital (Drucker & Goldstein, 2007). The empirical evidence of Paper 4 shows that the retention of highly educated graduates is associated with three central factors. First is the age and family formation of the graduate: older graduates with a spouse and children are significantly less migratory. Secondly, there is the regional sense of belonging: students who moved to another region to study are twice as likely to move away again. Thirdly, the most important factor above all is the geographical location of the institution: the larger the city size, the more likely graduates are to stay after graduation. This points to evidence of the importance of local labour-market opportunities for the residential choice of highly educated individuals (Venhorst & Cörvers, 2018). However, further empirical evidence of local employment patterns among local graduates is needed to fully understand this relationship in the Danish context.

A final argument is that local educational institutions might play a fundamental role in the identity and cohesion of local areas through interaction with civil society (Andersen & Etzerodt, 2018). This is a central and important argument, but I do not engage with it in this thesis. Other, more qualitative research designs are needed in order to evaluate how important this is.

5.4 Concluding remarks

To conclude, the aim of the thesis is to gain insights into how the geographical location of educational institutions influences human capital creation and inter-regional migration in Denmark.

The findings can be condensed into three main conclusions:

1) A long distance to school is not an obstacle to young people in reaching their educational goals if they have sufficient encouragement and support from their parents.

2) Young people do not leave their hometown because of a lack of educational opportunities but *because of* local educational opportunities. Educational achievement creates new opportunities and widens one's view. Hence, after having completed a general upper secondary education, the majority of students are inclined to move to a larger city in search of more educational opportunities and job opportunities, having already acquired a "taste of" widening one's horizons.

3) HEIs only have an impact in attracting highly educated graduates to the local labour market if they are located in regions that are already thriving economically.

It is my hope that the findings in this thesis will help inform current and future policy decisions about the location of educational institutions in Denmark.

5.5 Directions for further research

After writing this thesis, I am left with numerous ideas of further research and specific wishes to improve the research papers. I will highlight some of them in this final section. First and foremost, it would be interesting to extend our study of the migration patterns of young adults aged 18 to 24 by adding more years to the dataset. By following young persons' migration paths over a longer time span, it would be possible to evaluate the overall effect on society of youth migration to an urban area. It would also be valuable to know how large a share of the birth cohort return to their regions of upbringing again, after living in the city for some years. In addition, it would be interesting to examine in which parts of the country young adults make this return and which parts they do not return to. Furthermore, a comparative study of the young adults who grew up

in urban areas would help provide a fuller picture of inter-regional migration among the young.

Considering the methodological improvements to the study, I would like to mention a few natural improvements one could make. Firstly, one could examine the effect of other definitions of rural areas and focus more on the role of different city sizes. Secondly, definitions of internal migration, such as distance moved, could also help improve the study. Thirdly, the identification strategy could be improved by taking distance to VET into account as well. Finally, the use of openings and closures of upper secondary schools as a source of exogenous variation in distance would also strengthen the analysis.

There are also several directions for further research in the analysis of highly educated graduates. Firstly, it would be interesting to compare the retention rates of graduates who completed their higher education in different economic cycles. How does an economic recession influence graduate migration in comparison with cycles of economic growth? Secondly, more detailed analyses of graduates' sequential migration patterns within Denmark could be valuable, in particular to examine which regions do attract return migrants.

Regarding possible methodological improvements, a competing risk model with three outcomes could help clarify the individual characteristics of the return migrants compared to the repeat migrants and the university stayers. In addition, the application of a discrete hazard rate model would improve the latter's flexibility. Finally, one possible improvement is to allow for the repeating of spells (that is, repeated moves) in the statistical model.

5.6 References

Andersen, J. G., & Etzerodt, S. F. (2018). Placering af offentlige institutioner (Location of public institutions). *Økonomi & Politik*, *91*(4), 34–49.

Drucker, J., & Goldstein, H. (2007). Assessing the Regional Economic Development Impacts of Universities: A Review of Current Approaches. *International Regional Science Review*, *30*(1), 20–46.

Florida, R. (2002). *The rise of the Creative Class: And row it's transforming work, leisure, community and everyday life*. New York: Perseus Book Group.

Venhorst, V. A., & Cörvers, F. (2018). Entry into working life: Internal migration and the job match quality of higher-educated graduates. *Journal of Regional Science*, *58*(1), 116–140.

Paper 1

PAPER 1

Youth Participation in Education: The Impact of Distance to School and Parental Education

Elise Stenholt Sørensen

Abstract

Geographical distance between parental home and upper secondary school could be a potential barrier to the decision to participate in upper secondary education. This paper estimates the impact of distance to the nearest general upper secondary education school on enrolment in and graduation from general upper secondary education. Location data on Danish upper secondary schools, which increased in number between 1982 and 2014, are used as a source of exogenous variation in distance to school. This allows me to estimate the causal effect of distance by using changes in distance from home to school, which previous studies have not done. The results show that distance from home to the nearest general upper secondary education institution has a small overall impact on enrolment in and graduation from general upper secondary education. However, there is significant heterogeneity in the association between distance and participation. The association is largest for young people from less-educated families and negligible if one of the parents holds a university degree. In addition, the association between distance and the decision to participate in upper secondary education has become more significant in recent years: the effect was statistically insignificant between 1982 and 1987 and more significant from 2007 to 2019. The small, but potentially important, impact of distance on participation decisions, could, in the long run, maintain spatial disparities in human capital and labour market outcomes.

Keywords

Educational choice, youths, regional development, human capital, distance to upper secondary education.

1. Introduction

In people's decision making on their educational paths, distance to educational institutions is associated with potential costs, both in terms of monetary costs, information costs and emotional costs (Gibbons & Vignoles, 2012). The monetary costs occur if an individual has a long commute to school or has to leave home to study. Longer distance increases information costs and there might be emotional costs associated with moving away from friends and family. These costs can influence educational choice, particularly for young adults from less-educated families, where parental incentives and aspirations might be lower (Dickerson & McIntosh, 2013; Falch, Lujala, & Strøm, 2013). Thus, one concern is that geographical distance between parental home and upper secondary school poses a potential barrier to enrolment in upper secondary education.

Another related concern that arises is that geographical distance between parental home and upper secondary school may deter young people from choosing "the right" upper secondary education track, if their favourite educational track is further away (Gibbons & Vignoles, 2012). This could lead to efficiency issues, because those facing the highest costs (facing the longest distance to school) may have higher potential returns from completing a general upper secondary education programme. If distance costs vary according to families' socio-economic backgrounds, variations in the geographical accessibility of youth education would influence how differences in education choices for different socio-economic groups are reproduced. This has implications for inequality in educational outcomes, earnings and life chances (Gibbons & Vignoles, 2012).

A study by (Heinesen & Sørensen, 2021) shows that there is a long-term income gain (at the age of 40) related to completing general upper secondary education programme instead of a vocational education programme in Denmark. Thus, small differences in access to upper secondary education could, in the long run, cause and retain spatial disparities in human capital and labour market outcomes. Based on these concerns, many local authorities and politicians have discussed the need for enhancing the geographical dispersion of upper secondary education in Denmark (Danske Gymnasier, 2019; Danske Regioner, 2019; Jakobsen, 2020; Toft, 2020). However, these policy proposals are based on weak empirical evidence.

This paper assesses the effect of distance from home to the nearest general upper secondary education institution on youths' enrolment in and completion of general upper secondary education programmes. The contribution of the paper is twofold. Firstly, it estimates the causal effect of distance from home to the nearest general upper secondary education institution on participation in upper secondary education. Secondly, the paper presents a novel research design that uses exogenous variation in distance to schools to estimate causal effects of distance to education. Specifically, the paper examines three research questions:
1) To what extent does distance from home to general upper secondary education institutions influence participation [1]?
2) Are young individuals from disadvantaged backgrounds more affected by geographical constraints?
3) Has the effect of distance from home to the nearest upper secondary education institution changed during the last three decades?

Using rich administrative data on the total population of 15-year-olds in Denmark, from 1982 – 2018, this study is the first to use location data on general upper secondary schools, which have increased in number since 1982, to examine the causal effect of distance from home to the nearest general upper secondary education institution on participation in post-compulsory education.

The international evidence of the impact of distance on post-compulsory education participation is limited. To my knowledge, only two studies have assessed the impact of distance from home to educational institutions on participation in post-compulsory education (Dickerson & McIntosh, 2013; Falch et al., 2013), which stands in contrast to a large literature examining the effect of distance to higher educations (see among others (Denzler & Wolter C., 2011; Frenette, 2009; Gibbons & Vignoles, 2012; Spiess & Wrohlich, 2010). Both studies on the impact of distance from home to educational institutions on participation in post-compulsory education conclude that geographical constraints have a small overall impact on enrolment and on-time graduation from upper secondary schools (Dickerson & McIntosh, 2013; Falch et al., 2013).

[1] The present study considers two measures of participation: a) enrolment in a general upper secondary education before age 18 (y1), and b) graduation from a general upper secondary education before age 21 (y2).

The vast majority of the empirical studies have not paid close attention to issues of causality (Gibbons & Vignoles, 2012). Estimation of the causal effects of distance from home to place of education on participation is beset by problems of spatial heterogeneity and residential sorting. These problems arise because families with different incomes, abilities and preferences sort into residential locations with different distance to education institutions. It is therefore easy to mistakenly infer a causal linkage between distance and participation, when the association between distance and participation is caused by family background and residential sorting (Gibbons & Vignoles, 2012).

This study contributes to the literature in several ways. Firstly, it adds new knowledge to the sparse evidence on the effects of distance from home to place of education for youths' education decisions. Secondly, this study provides knowledge about the time variation in the effect of distance to school by analysing distance data from 1982 until 2014, in total 33 birth cohorts. Finally, the paper applies a new research design for deriving causal estimates of distance to school. I selected siblings who lived at the same geographical location, but had different distance to general upper secondary school because schools had opened up or closed down. This empirical strategy could potentially be used as an instrumental variable approach in future studies concerned with the influence of upper secondary education on other outcome variables such as wage and labour market outcomes.

The remainder of the paper is organised as follows. In section 2, I describe the international literature about geographical accessibility of education and its impact on educational choice. In section 3, I describe the statistical methods used in the analysis. In section 4, I briefly describe Denmark's educational system, where I focus on admission to upper secondary education programmes and their geographical location. Subsequently, the data sources and variables are described in section 4. The first part of the analysis, which considers the relationship between distance of parental home to upper secondary schools and participation, is described in section 5. The second part of the analysis, which examines the effect of changes in distance to education, is presented in section 6. This is followed by conclusions in section 7.

2. Geographical constraints, Participation and Choice of Education

There is a large literature studying the impact of geographical constraints on participation in and choice of higher education. The majority of the studies are from the US; Card (1995) and Kane & Rouse (1995) were the first to provide empirical evidence of distance being an important determinant of college participation. Since then, empirical studies from European countries have provided evidence on the impact of geographical constraints on participation in and choice of higher education from Germany, the Netherlands, Switzerland, Sweden, Denmark, Ireland, and the UK (Cullinan, Flannery, Walsh, & McCoy, 2013; Denzler & Wolter C., 2011; Gibbons & Vignoles, 2012; Öckert, 2012; Sá, Florax, & Rietveld, 2006; Sørensen, 2021a; White & Lee, 2020). The evidence is rather mixed and there are large country-specific differences with regard to data sources, the spatial distribution of higher education institutions, proximity measures and general access to higher education. The following description will focus on empirical evidence from studies in European countries, unless otherwise stated, as this seems most relevant to the Danish context.

There are two main research questions addressed in the literature. The first assesses the effect of distance from home to place of education on participation (whether someone enrols in higher education or not) and the second examines the effect of distance from home to place of education on "field-of-study" decisions among young people who have enrolled in higher education. Empirical evidence on the association between distance from home to place of education and participation shows that the distance to higher education has a very small or no impact on the decision to participate in higher education (Cullinan et al., 2013; Gibbons & Vignoles, 2012; Öckert, 2012; Sá et al., 2006; Sørensen, 2021a). Studies concerned with how distance from home to place of education effects choice of "field-of-study" agree that proximity to university influences a student's choice of study, conditional on enrolment (Denzler & Wolter C., 2011; Gibbons & Vignoles, 2012; Sá et al., 2006).

Evidence on the impact of geographical accessibility on participation in upper secondary education is, however, limited. Falch, Lujala and Strøm (2013) find that geographical constraints have an impact on on-time graduation from upper secondary schools in Norway. Using administrative data on one complete

birth cohort of young individuals in Norway, they examine how travel time from students' homes to their nearest upper secondary school affects graduation. They conclude that "reduced travel time has a positive effect on graduation" and that the effect is strongest for students at the margin of graduation (those in the middle of the grade-distribution, the second and third quartile) (Falch et al., 2013). However, they find that the effect of travel time does not differ across parental education.

Another study by Dickerson and McIntosh (2013) uses survey data from England to examine the effect of distance on 16-year-olds' decisions to enrol in upper secondary education. They find a small overall effect of distance to the nearest academic institution on participation in academic upper secondary education. On average, a one-kilometre increase in distance to academic upper secondary education reduces the likelihood of participating by 1.5 percentage points. Furthermore, their results suggest an even larger effect of distance for a group of young people who are on the margin between participating in academic upper secondary education.

In the Danish context, Jensen (2019) has examined the impact of the proximity of a technical track in general upper secondary education (within 20 km of a potential student's home) on the choice of upper secondary education. The study concludes that the introduction of technical general upper secondary education tracks "significantly changed the distributions of students across upper secondary education choices" (Jensen, 2019). In addition, two descriptive reports have considered the role of distance from home to places of upper secondary education in Denmark. The first, a conference paper by Nielsen (2011) that examined data at the municipal level, found a correlation between the average transportation time from home to place of education and the average participation rates in upper secondary education. The second, a report by the Danish thinktank DEA, found that distance from home to place of upper secondary education was uncorrelated with the share of young individuals who graduated from a general upper secondary education programme and correlated positively with the share who graduated from vocational upper secondary education programmes, conditional on enrolment (DEA, 2019). However, both analyses lack a broad range of individual and family controls that would be able to take residential sorting into account.

The majority of the previous empirical studies have not paid close attention to issues of causality and the possibility that unobserved factors, such as parental aspirations, could affect both residential location and participation and

therefore bias the results. Frenette's (2009) empirical analysis from Canada is an exception. He uses the creation of five new universities (or university departments) in Canada from 1982 – 1995 to study university participation rates following the establishment of a university in a city where there was previously no university. Frenette (2009) concludes that the creation of a local university is associated with an increase in university enrolment among the local population of the area the university is placed in. However, the increase in university enrolment came at the expense of other types of tertiary education, for example college participation, in most of the cities. Therefore, there was little change in overall participation rates in tertiary education.

In sum, the existing evidence suggests that distance from home to place of education has little or no impact on enrolment rates in higher education, conditional on the completion of an upper secondary education. The existing evidence also suggests that there is a more significant effect of distance from home to place of education on youths' choice of type of tertiary education, conditional on enrolment in a higher education. The few studies on distance from home and travel time to upper secondary education institutions suggest that the effect of distance on the decision to participate in upper secondary education is stronger among individuals aged 16 who have recently left elementary school (Dickerson & McIntosh, 2013). One might expect that sensitivity to distance is larger among young individuals who still live at their parental home, and hence are more constrained with regards to re-location. However, additional empirical evidence is needed to draw this conclusion.

3. Methodology

3.1 Research design

The aim of the empirical analysis is to examine how young people's educational decisions are influenced by distance between their parental home and the nearest general upper secondary education institution. Despite the fairly simple question, it is challenging to measure the causal effect of distance, empirically. The challenge is that distance to school is not random. Choice of residential location is determined by household characteristics (e.g., income and parental education), which may also influence a child's educational decisions. This implies that households that are located close to upper secondary schools will likely differ in substantial ways from those located far from upper secondary schools. For example, highly educated parents would be more likely to live in urban areas where distance to upper secondary school is shorter, on average. Thus, a research design that compares the decisions of individuals who face different home-to-school distance patterns according to where they live, could be problematic. The difference in participation rates could be due to residential sorting and not (solely) distance to school.

Many previous studies discuss this potential problem (see among other (Frenette, 2009; Gibbons & Vignoles, 2012; Spiess & Wrohlich, 2010) but only a few of them have had access to data sources that have enabled them to take the endogeneity problem into account. Gibbons and Vignoles (2012) rely on an extensive control-variable-based regression strategy and Falch et al. (2013) use several IV-estimation techniques. However, previous studies lack exogenous variation in the location of institutions over time. The present study is the first to use historical openings and closures of general upper secondary schools as a source of exogenous variation in distance to upper secondary school. I use changes in distance to upper secondary education schools within families[2]. Hence, I compare siblings who grew up in the same family[3] at the same

[2] Despite the fact that I have access to rich panel data, I cannot use the panel data structure on the individual level. I face the problem that young people only have to choose their upper secondary education once. A person only finishes elementary education once and distance to upper secondary education is only relevant to someone once in their life – when they are aged 15 and graduate from elementary school.

[3] Individuals are defined as siblings if they have the same mother, not necessarily the same father.

residential location. Using a fixed effects (FE) estimation technique, I estimate the effect of changes in distance (due to openings and closures) on the probability of participation in general upper secondary education within families.

The following empirical analysis consists of two different parts. The first part (section 5) consists of results based on the full population of 15-year-olds. Section 5 describes the association between distance from home to general upper secondary education and participation. In addition, heterogenous effects of home-to-school distance (regarding time period and parental education) are examined and discussed. The second part of the empirical analysis (in section 6) consists of results based on a sub-population of siblings. The results in section 6 describe the effect of *changes in distance* from home to general upper secondary education within siblings in the same family. These results add new knowledge to the overall results about the causal effect of distance from home to school. The following description of the empirical models is organised in a similar way. Thus, first the "full population" model is described and then the "sibling population" model is presented.

3.2. Distance from home to school and participation in upper secondary education

In the first part of the analysis, I apply a simple binary choice model, estimating the probability of enrolment before age 18 (y1) or graduation before age 21 (y2) with respect to distance from home to the nearest general upper secondary school. For reasons of simplicity, I consider y1 and y2 under one common term "participation (y)" in the following description of the empirical model. The binary variable y_i is equal to one if the individual participates in a general upper secondary education programme. If the individual participates in multiple upper secondary educations, the first choice is used.

I estimate the following logit model:

$$P(y_i = 1) = \frac{\exp(V_i)}{(1+\exp(V_i))}, \qquad \text{where } V_i = \alpha + \delta D_i + \mathbf{x}_i'\boldsymbol{\beta} \qquad (1)$$

where subscript i denotes individuals and where $\alpha, \delta, \boldsymbol{\beta}$ are parameters to be estimated. The continuous variable D indicates distance from home to the nearest general upper secondary school in kilometres, while $\mathbf{x}$ is a vector of

observable control variables. I control for observable individual characteristics (gender, immigration status, and birth year), geographical characteristics (residential region, and city size), and family characteristics (parental education, and the natural logarithm of disposable annual household income). I return to a description of the control variables in section 4.5. By assumption, the error term e is independent of $\mathbf{x}$ and has the standard logistic distribution. The parameters are estimated by maximum likelihood methods.

This estimation strategy relies on the assumption that I have information about all relevant control variables that are correlated with participation and distance. This might be a problematic assumption due to unobserved residential sorting, as discussed in section 3.1[4]. This implies that the parameter estimate of distance cannot necessarily be interpreted as the causal effect of distance. To circumvent the possible endogeneity of distance D, I employ an additional estimation approach using changes in distance to the nearest general upper secondary school.

3.3. Changes in distance from home to general upper secondary school

I use the fact that some new general upper secondary schools were created while others closed down between 1982 and 2014. Table A1 in the appendix presents the historical openings and closures. These openings and closures give exogenous variation in distance. The openings and closures imply that distance to general upper secondary school, in some cases, changed between siblings within the same family. I exploit the fact that I can link individuals to their parents and their siblings in the data. I sample siblings who lived at the same geographical location, with the same parents, but who were exposed to different distances to the nearest general upper secondary school because the nearest school opened up or closed down between the two siblings' 15th birthdays. If parents cannot forecast the openings and closures of education institutions, the change in distance for siblings in the same family is arguably an exogenous shock. I use this to estimate a family FE model, where subscript f denotes families and subscript s denotes siblings within the same family.

[4] I conduct the estimation strategy, so my results are comparable with previous empirical studies in the literature that are estimated in a similar way.

$$\ddot{y}_{fs} = \ddot{\boldsymbol{x}}'_{fs}\boldsymbol{\beta} + \delta\ddot{D}_{fs} + \ddot{u}_{fs} \qquad (2),$$

Where $\ddot{y}_{fs} = y_{fs} - \bar{y}_f$, $\ddot{\mathbf{x}}_{fs} = \mathbf{x}_{fs} - \bar{\mathbf{x}}_f$, $\ddot{D}_{fs} = D_{fs} - \bar{D}_f$ and $\ddot{u}_{fs} = u_{fs} - \bar{u}_f$

Thus, subtracting the sibling averages within each family from each sibling's number yields the Family Fixed Effect model in equation (2). The estimation results should therefore be interpreted as the effect of change in distance from home to school on changes in participation rates. As estimates are based on changes in distance, only those siblings who experience changes in distance enter into the calculations. It is therefore relevant to describe the magnitude and frequency of changes in distance from home to upper secondary school in Denmark (see section 6) and also to give a brief introduction to the Danish education system (see section 4).

4. Institutional context and data

4.1. Institutional background

This section gives a short introduction to the Danish education system. For a more thorough description of the Danish education system see Chapter 4 in the Introduction of this dissertation. Compulsory education in Denmark consists of 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 education system, continue in 10th grade, or enter upper secondary school. The upper secondary school comprises two main tracks of education: a vocational education and training (VET) track and a general upper secondary education track.

The VET track consists of many different types of vocational programmes, e.g. commercial field, technology, construction, craftmanship, food production, mechanical engineering and service. General upper secondary education (equivalent to high school in the US) consists of three main types of education programmes: a general academic track (HF and STX), a mercantile track (HHX), and a technical track (HTX). General programmes are typically three-year programmes[5]. They are a formal requirement for admission to all tertiary

[5] However, some programmes (HF) are only two years.

educations, but they do not provide qualifications for jobs in the labour market. The general upper secondary education institutions and the VET institutions are not usually located at the same geographical location.

The education system in Denmark is universal (The Ministry of Higher Education and Science, 2016). There are no tuition fees in upper secondary or in higher education. Students who have attended the same lower secondary school together will typically enrol in different upper secondary education institutions depending on which upper secondary school and study track they prefer. There was no grade requirement to enter general upper secondary education in the empirical period of the present analysis[6]. If the number of applicants exceeded the number of student seats in a specific general upper secondary education institution (typically in the largest cities), the general upper secondary education institution had the possibility to choose between candidates based on their letter of application and their GPA from compulsory education. If the students did not send an application letter (this was not mandatory) the decision would often be based on the distance to the institution from the family home. The students who lived closest to the institution took precedence over students who lived closer to other upper secondary institutions. Overall, general upper secondary education institutions are very homogenous in Denmark in terms of quality of education, the composition of pupils and the obtained skills.

4.2. Data

The analysis is based on a composite data set linked together from a number of administrative data sources. The core of this dataset is 33 birth cohorts of Danish youths at the age of 15. The cohorts were born from 1966 to 1998, which means that they were 15 years old in 1982 to 2014. The sample is limited to youths with a residential address in Denmark at age 15 years (in total 2,213,154 individuals). The administrative population register (BEF) from Statistics Denmark is used to acquire information about birth year and residential address. Youths who lived on an island without a bridge to the mainland are excluded from the analysis. This is done to ensure similar distance

[6] New admission requirements for enrolment in general upper secondary education became effective from March 2019. The applicant had to have passed the 9th form-level school-leaving examination (of the Danish Folkeskole) and have a GPA from the school-leaving examination of 5.0.

measures through the road network[7]. In total, 0.3 percent of the sample is excluded due to this. I also exclude persons with missing education information (about 2.8 percent) and persons with no distance information (about 3 percent). The total regression sample is in total 2,061,470 individuals. The population register (BEF) is also used to link individuals to their mothers and siblings via a unique identifier for each mother and father. I define siblings as persons who have the same mother, but not necessarily the same father. About 80 percent of the regression sample have a sibling within the analytical time frame.

4.3 Dependent variables

Participation. I analyse two different measures of education choice: enrolment in general upper secondary education, up until age 18 (y1) and graduation, up until age 21 (y2). In those cases where an individual has enrolled in, or graduated from, several upper secondary education programmes after primary school, I use information on youths' first choice of upper secondary education or first completed upper secondary education. Those individuals who never enrol/graduate from an upper secondary education programme are registered as "none" for education. Information about enrolment in upper secondary education programmes is provided in the "Pupils register" (KOTRE) by Statistics Denmark. Information about educational attainment is provided in the register of "Highest completed educations" (UDDF) by Statistics Denmark. Table 1 shows that the majority (55 percent) of the youths choose general upper secondary education programmes as their first choice of upper secondary education after elementary school (y1), and that 51 percent of the sample graduate from general upper secondary education before turning 21 (y2). There has been a significant change in educational choices among youths in Denmark during the past 30 years – see Figure 6 in Chapter 4 in this thesis.

4.4. Distance measures

Distance. The location information of all education institutions is provided by the Danish Ministry of Children and Education. I have calculated the distance between youths' homes and all general upper secondary institutions for each year in the analysis. The distance between home and all institutions was

[7] Among youths living on islands, the distance measure comprises the distance of the ferry route. The travel time with ferry is longer and more costly, thus the two types of distance measure are not comparable.

measured through the road network, using geographic information system (GIS) software. To limit the number of distance calculations, each residential address was located in a 100 x 100-meter geographically defined quadrant. The distance was measured from the bottom-left corner of each quadrant to all of the education institutions. This implies a possible measurement error of maximum 141 meters, which is so small that it would hardly have an impact on the results.

Overall, Danish youths live within relatively short distances to a general upper secondary education institution. The average distance to the nearest general upper secondary education institution was 7.13 kilometres and it ranged from 0 to 49 kilometres (as shown in Table 1). Table A4 in the appendix shows that 53 percent had less than 5 kilometres and 72 percent had less than 10 kilometres to the nearest institution as shown in. In addition, the distance histogram, shown in Figure A2 in the appendix, illustrates that only a small minority had above 20 kilometres.

Figure 1 shows the average distance each year from 1982 to 2014. The average distance to institution decreased from about 7.15 km in 1982 to 6.70 km in 2014. This is a small reduction, however, the overall trend is that the average distance has continuously become shorter since 1995. The "peak" in average distance shown in Figure 1 between 1992 to1994, seems like an outlier that is due to administrative changes. This peak needs to be examined further, which is beyond the scope of the present paper. In sum, the average distance from home to upper secondary education institution is relatively short and has not changed substantially during the past four decades.

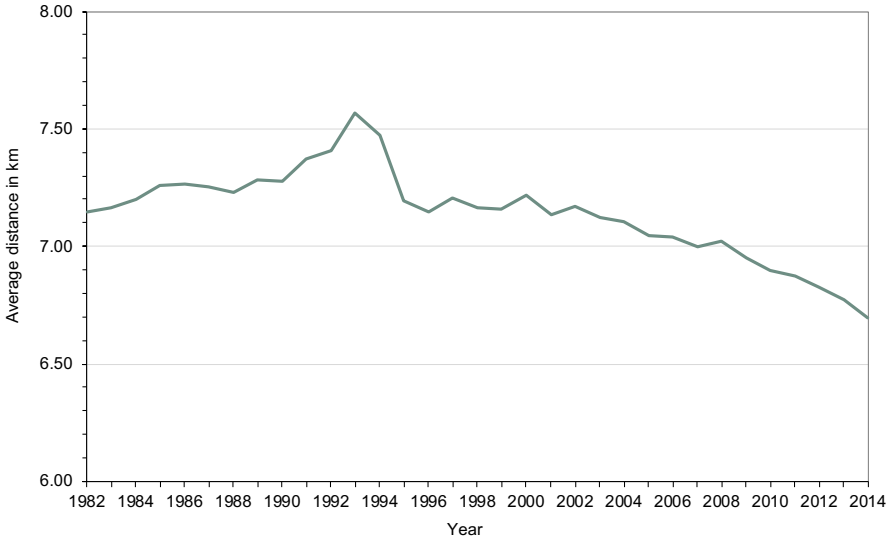

Figure 1. Average distance from home to the nearest general upper secondary education institution, 1982 – 2014.

Figure 2 shows the distance distribution divided by parental education background. As expected, we see that individuals with highly educated parents lived closer to education institutions than their peers from families with lower levels of education. There is no substantial difference in the distance distribution between the remaining three groups.

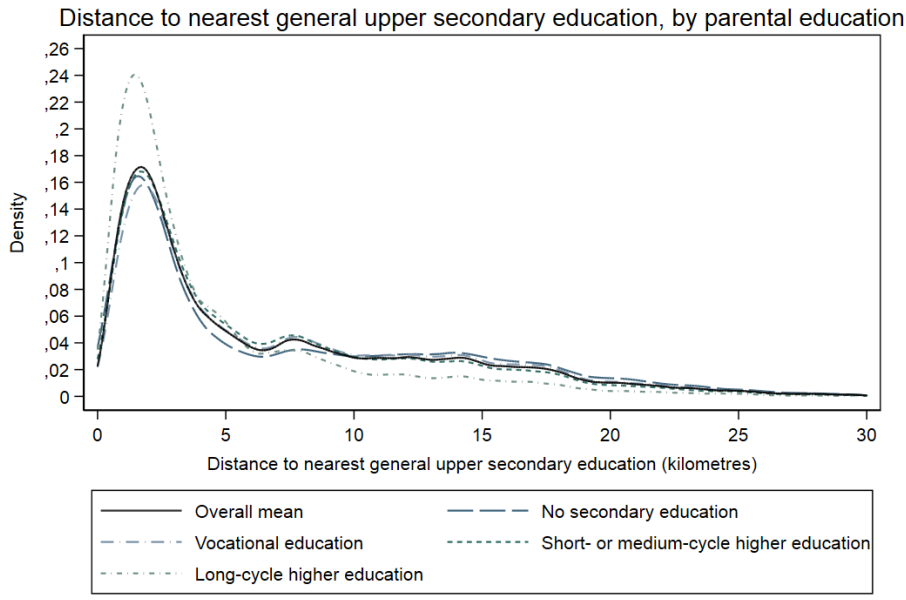

Figure 2. Average distance from home to the nearest general upper secondary education institution, 1982 – 2014.

Openings and closures of institutions. The geographical location of general upper secondary education institutions in Denmark, is derived from the "Institution register" (INST) of the Ministry of Children and Education (https://www.uvm.dk/institutioner-og-drift/institutionsregisteret). The dataset comprises information about all educational institutions in Denmark and their departments. It contains information about the exact address and geo-coordinates of every institution as well as the name of the institution and a unique identifier for each institution – INSTNUMMER. I link this identifier to individual pupils (age 14-17) through the Pupil dataset (KOTRE) for every year (1982 – 2014). This is to ensure that the institutions in my dataset are used for education and teaching of pupils (and not solely administrative tasks). In addition, I can detect which educational departments opened up or closed from 1982 to 2014. I delete institutions with less than 10 pupils enrolled in any given year.

4.5. Individual and family controls

The main goal of the analysis is to examine to what extent distance from home to school affects youths' first choice of education. However, it is evident from

previous studies (Dickerson & McIntosh, 2013; Falch et al., 2013) that distance to school is correlated with several other characteristics which also affect choice of education. In order to derive a good estimate of the effect of distance itself, I take several background variables into account in the analysis.

First, I include individual background characteristics in terms of dummy variables indicating gender, immigration background and year of birth. Previous research shows that females are much more prone to enrol in general upper secondary education than males, who are more represented in vocational education programmes (Heinesen & Sørensen, 2021). Secondly, I add geographical indicators to the model. That is residential region and a dummy indicator of whether the individual grew up in a rural area or not. It is natural to think that distance is more important in rural regions where the transport infrastructure is less developed and access to public transportation is less accessible. Thus, the transportation costs might arguably be higher in rural areas. In this study, a rural area is defined as a village or town with less than 20,000 inhabitants. Third, I take parental education and household income into account.

Table 1 presents the sample characteristics of the full sample and the sibling sample in the analysis. The two samples to not seem to differ substantially. About 51 percent are males, and 6 percent are either first- or second-generation immigrants. About 7 percent of the fathers and 4 percent of the mothers hold a university degree. All explanatory variables are measured at the age of 15, that is from 1982 – 2014.

Table 1
Sample characteristics for full sample and siblings.

	Full Sample		Siblings	
Variables	Mean	Std. Dev	Mean	Std. Dev
Enrolment (y1)	0.55	0.50	0.58	0.49
Graduation (y2)	0.52	0.50	0.56	0.50
Distance to nearest general upper secondary education	7.13	6.67	7.47	6.70
Male dummy	0.51	0.50	0.51	0.50
Immigrant status				
Native (ref.)				
Foreign born	0.03	0.17	0.02	0.15
Second generation immigrant	0.03	0.17	0.03	0.17
Urban				
Urban dummy	0.45	0.50	0.42	0.49
Geographical region (NUTS 2)				
Northern Jutland (ref.)				
Central Jutland	0.24	0.42	0.25	0.43
Southern Denmark	0.23	0.42	0.24	0.43
Capital Region	0.26	0.44	0.24	0.43
Zealand Region	0.16	0.36	0.15	0.36
Family income				
Ln (Annual Household income)	12.48	1.06	12.57	0.75
Richest 1 percent (dummy)	0.01	0.10	0.01	0.11
Poorest 1 percent (dummy)	0.01	0.09	0.00	0.06
Father's educational level				
Primary school	0.25	0.43	0.24	0.43
High school	0.03	0.17	0.03	0.17
Vocational education	0.41	0.49	0.43	0.49
Short-cycle HE	0.04	0.19	0.04	0.19
Medium-cycle HE	0.12	0.32	0.13	0.33
Long-cycle HE	0.07	0.26	0.08	0.27
Mother's educational level				
Primary school	0.32	0.47	0.30	0.46
High school	0.03	0.18	0.03	0.18
Vocational education	0.35	0.48	0.36	0.48
Short-cycle HE	0.03	0.17	0.03	0.18
Medium-cycle HE	0.19	0.39	0.21	0.41
Long-cycle HE	0.04	0.19	0.04	0.20
Observations	2,074,077		1,254,735	

Note: All explanatory variables are measured at the age of 15 (t0). Enrolment (y1) is measured at age 18 (t+3) and graduation (y2) at age 21 (t+6).

5. Empirical analysis: Distance from home to general upper secondary education institution

5.1. Descriptive statistics

In this section, I present descriptive statistics of distance from home to general upper secondary education institutions. I consider the following questions: How geographically accessible is general upper secondary education to Danish youths when measured by distance to their homes? How has this changed over time? And is geographical accessibility to general upper secondary education equal across family background? In addition, I consider the overall association between distance from home to upper secondary education institution and participation rates in upper secondary education.

First, I consider the geographical accessibility of general upper secondary institutions from young people's homes. Figure 3 maps the average distance from home to the nearest general upper secondary education institution for 94 municipalities in Denmark[8]. Distance is calculated from each person's residential address (in a 100-meter square raster), and then averaged at the municipality level[9]. Figure 3 shows that very few Danish municipalities have poor geographical access to general upper secondary education institutions. As expected, distance to general upper secondary education institutions is shortest in urban areas. In figure 3, the lightest yellow municipalities (with access to a general upper secondary education institution within 5 km) tend to delineate the three major cities in Denmark, Copenhagen, Aarhus and Odense. However, the lightest yellow areas reach far north and west of the City of Copenhagen, showing that the geographical access to schools is under 5 kilometres in the large capital area surrounding Copenhagen. Figure 3 illustrates that distance varies across Denmark; however, the distances are relatively small. In the majority of the municipalities, the average distance is less than 15 kilometres, and only four municipalities have an average distance above 15 kilometres. Figure A1 in the Appendix maps the location of general upper

[8] Denmark consists of 98 municipalities (LAU 1) in total. Four out of 98 municipalities are excluded because they are small islands with no general upper secondary education institutions and with no road-network connection to the mainland. The four excluded municipalities are Ærø, Fanø, Samsø and Læsø.

secondary education institutions and illustrates that the geographical dispersion of general upper secondary education institutions in Denmark is relatively good.

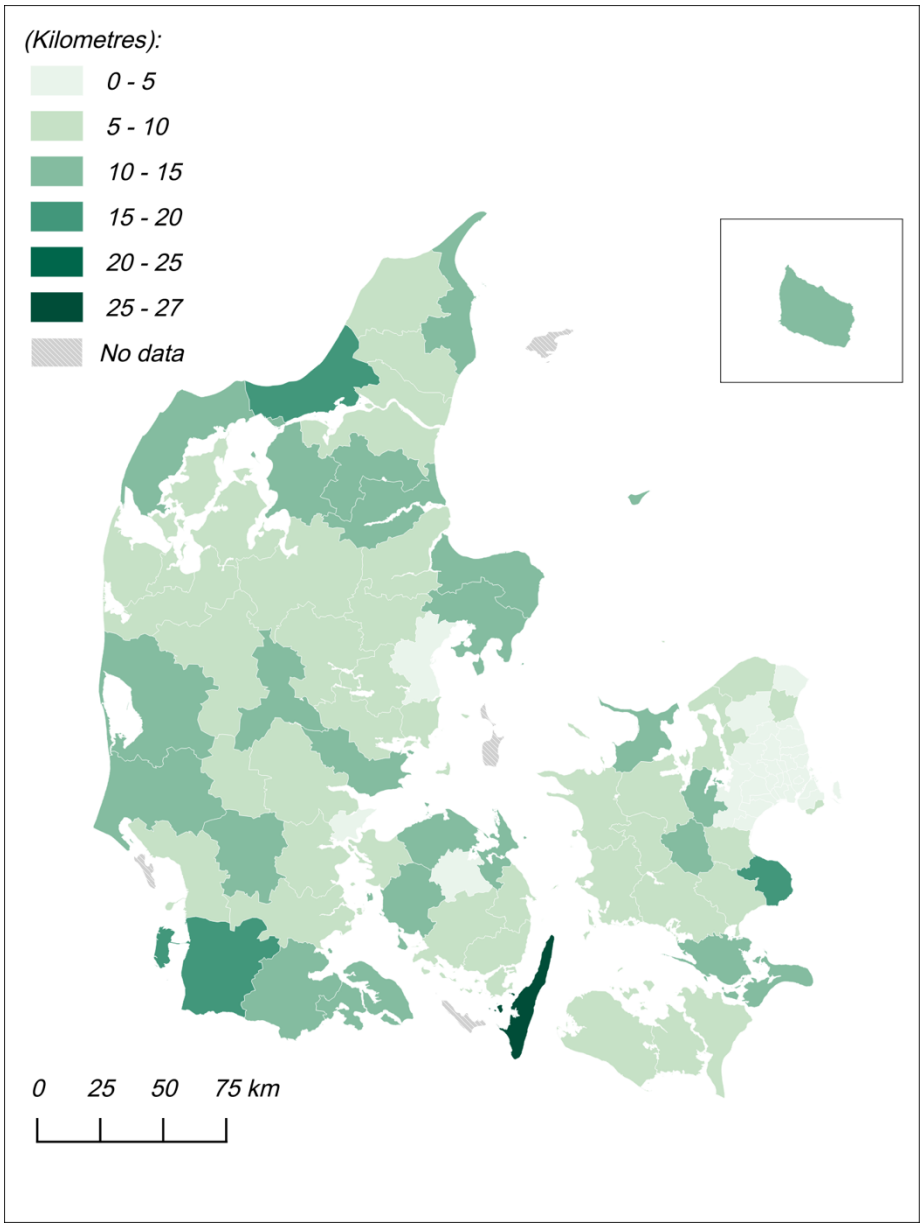

Figure 3. Average distance from home to general upper secondary education institution, by municipality (year 1982 - 2014). The figure shows the mean distance through road network, from home to the nearest general upper secondary education institution. The distance calculation is based on exact addresses of residence, of age-15 population in 1982 – 2014.

5.2. Descriptive association between distance and participation

In this section, I consider the descriptive association between general upper secondary education participation at the individual level and distance from home to the nearest general upper secondary education institution. Table 2 shows the mean distance to the nearest general upper secondary education institution and standard deviation in parentheses below. The mean distance is summarised by time period, parental highest education and general upper secondary education participation status (y1).

Overall, we see that young people live, on average, relatively close to a general upper secondary education institution (between 4.6 – 8.5 kilometres). It is also evident that participants in upper secondary education, on average, have shorter distance to school, compared to non-participants. This association is present within all family types and across time. On average the difference is about 0.9 kilometres, which is not large.

Looking at parental education, we see that distance to school is associated with parental highest education. The more educated the parents are, the shorter distance is, on average (see also Table A3 in appendix for details). Young people from highly educated families (Group 4) have 4.82 km to a general upper secondary education institution on average, compared to young people from low-educated families (Group 1), with 7.74 km on average.

Table 2 also indicates that distance from home to a general upper secondary education institution has decreased over time. The mean distance has fallen from 7.21 kilometres in 1982 – 1987 to 6.88 kilometres in 2007 – 2014 (see Table A2 in appendix for the overall averages). It is striking that the fall in average distance is not equal for all family groups. The largest drop in average distance is seen in families where none of the parents hold qualifications beyond compulsory schooling (Group 1), while there is almost no difference in average distance over time in families where at least one parent holds a tertiary education (Group 3 and 4). However, in the last group (Group 2, where one or both parents have a vocational education) we see the opposite tendency: in Group 2, the average distance has increased from about 6.9 km (among non-participants) to 8.2 kilometres (among non-participants).

This could potentially be caused by two different mechanisms – either a pattern where general upper secondary education institutions have primarily been opened in areas with lower proportions of families with vocational educations (e.g., city centres), or a pattern where families with vocational educations have moved out of city centres in suburban areas because of a rise

in urban housing prices. A possible explanation for the opposite association in Group 1 (where none of the parents have qualifications beyond compulsory schooling) could be that the ethnic composition in Group 1 have changed during the period analysed. In beginning of the 1980s, this group primarily comprised native Danish working-class families resident in suburban areas. By 2007 – 2014 the group comprised a larger share of immigrants that are predominantly resident in the largest urban areas, hence closer to general upper secondary education institutions.

Table 2.
Distances from home to general upper secondary education institution, summary statistics for participants and non-participants (y1), by time and parental education.

		1982 - 1987		1988 - 1994		1995 - 2001		2002 - 2006		2007 - 2014	
		No upper sec. edu.	Upper sec. edu.	No upper sec. edu.	Upper sec. edu.	No upper sec. edu.	Upper sec. edu.	No upper sec. edu.	Upper sec. edu.	No upper sec. edu.	Upper sec. edu.
Parental highest education											
1: Mother and father: no education	Mean	8.5	8.3	8.2	7.9	7.5	7.0	7.4	6.5	7.5	5.9
Distance to nearest institution	Std. Dev.	(7.3)	(7.2)	(7.3)	(7.2)	(7.2)	(6.9)	(7.2)	(6.9)	(7.1)	(6.4)
	N	92,635	35,014	70,194	31,300	48,776	25,651	31,477	16,304	41,938	31,390
2: Mother or father: vocational edu.	Mean	6.9	6.6	7.6	7.4	7.9	7.7	8.1	7.7	8.2	7.5
Distance to nearest institution	Std. Dev.	(6.6)	(6.4)	(6.8)	(6.6)	(7.0)	(6.9)	(7.0)	(7.0)	(7.0)	(6.8)
	N	108,394	73,039	114,294	91,595	83,593	79,780	66,353	60,086	102,119	129,919
3: Mother or father: tertiary edu.	Mean	7.0	6.7	7.2	7.0	7.0	6.7	7.1	6.6	7.2	6.5
Distance to nearest institution	Std. Dev.	(6.3)	(6.2)	(6.5)	(6.3)	(6.5)	(6.3)	(6.5)	(6.3)	(6.6)	(6.2)
	N	27,625	50,834	35,190	76,142	28,503	73,664	23,933	61,616	38,143	135,141
4: Mother or father: university	Mean	5.7	4.9	5.6	5.1	5.1	4.8	5.2	4.7	5.1	4.6
Distance to nearest institution	Std. Dev.	(5.9)	(5.2)	(5.7)	(5.4)	(5.6)	(5.2)	(5.6)	(5.1)	(5.4)	(5.0)
	N	3,294	19,715	4,019	28,758	3,778	29,641	3,554	27,230	6,094	63,352

Notes: Table reports means, standard deviations in parentheses and numbers of observations. Distances are in km. "No upper sec. edu." contains people who did not enrol in any general upper secondary education before age 18 (y1=0). "Upper sec. edu." contains people who enrolled in a general upper secondary education before age 18 (y1=1).

Before turning to the regression models, I consider the association between distance from home to the nearest general upper secondary education institution and participation. The aim is to gain more knowledge about the distance distribution and choose an appropriate functional form for the regression models.

Figure 4 illustrates the association between distance from home to the nearest general upper secondary education institution and participation. The figure shows that the association is negative, but not perfectly linear. Between 0 – 5 kilometres, the attendance rate is almost flat at around 0.58 - 0.59. Between 5 – 12 kilometres, the attendance drops significantly from about 0.59 to 0.5. Above 12 kilometres, the slope of the curve less steep. Above 20 kilometres the curve rises again, which indicates that above 20 kilometres distance will actually increase enrolment. This finding is surprising. It could suggest that if someone lives more than 20 kilometres away from their nearest general upper secondary education institution, distance does not have the same effect, because they need to commute with public transportation anyway. However, above 20 kilometres, the confidence intervals become larger (less efficient), because of fewer observations. Only 5 percent of the population has more than 20 kilometres to a general upper secondary education institution, and 0.5 percent has more than 30 kilometres.

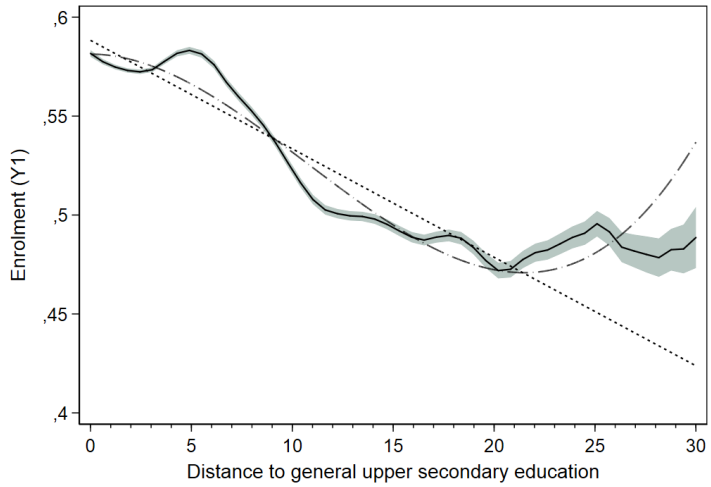

Figure 4. Association between distance to the nearest general upper secondary education and enrolment rate (without outliers > 30 km). Non-parametric association between distance to the nearest general upper secondary education institution and enrolment rates (1982 – 2014). In Figure 4, distances above 30 kilometres are left out to give a better visual picture of the association of the 99.5 percent of the population. See Figure A3 in Appendix for the association in the full range of the data (0 – 50 kilometres).

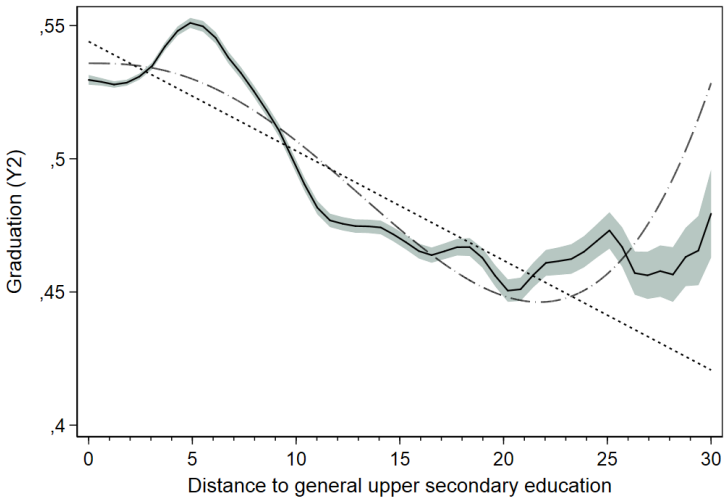

Figure 5. Association between distance to the nearest general upper secondary education institution and graduation rate (without outliers > 30 km). Non-parametric association between distance to the nearest general upper secondary education institution and graduation (1982 – 2014). Distances above 30 kilometres are left out to give a better visual picture of the association of the 99.5 percent of the population. See Figure A4 in Appendix for the association in the full range of the data (0 – 50 kilometres).

On the basis of a visual examination of the data, I chose to exclude outliers (above 30 kilometres) from the analysis. This gives a better model fit, which is verified by analysis of the residuals, shown in Figure A4 and A5 in the appendix. After careful consideration of various distance specifications, I chose a specification where distance is squared, to ensure flexibility of the model. Figure 4 confirm that a model with a polynomial distance specification is the best model.

These simple comparisons give some descriptive evidence that proximity to general upper secondary education institution could have a small influence on participation. However, I have not yet controlled for family income, immigration status, geographical region and other factors which could potentially affect both distance and participation. In the next section I use regression analysis to control for other observable personal and geographical characteristics.

5.3 Regression models of distance and participation

Table 3 reports the probability of enrolment and graduation from general upper secondary education with respect to distance from home to the nearest general upper secondary education institution. The estimates are derived from

a binary logit model (Equation 1.) described in section 3.2. Table 3 presents the average marginal effects (AME) of distance, using a polynomial distance specification. Other model specifications have been conducted (linear specification and logarithmic distance) and the results are very similar. All the presented models include a full set of individual, parental and geographical controls.

The first column (1) in Table 3 shows the overall result, using data from the whole time period (1982 – 2014). The result suggests that distance from home to the nearest general upper secondary education institution has a small but significant effect on both enrolment and graduation. When the distance increases by 10 kilometres, the probability of enrolment in general upper secondary education decreases by 2.6 percentage points on average. Columns (2) to (6) show the estimation results for each time period. They reveal that the association between distance and enrolment increases by time. In the beginning of the time period (1982 – 1987), the effect size is not statistically different from zero. In the second time period (1988 – 1994, in column (3)), the AME is -0.1, in the third period it changes to about -0.2 (column 4) and in fourth period (column 5) the AME is -0.4. The largest effect, around -0.0047 (in column 6), is seen in the last period (2007 – 2014). The effect size implies that shortening the distance to institutions by one kilometre increases the probability of enrolment by 0.47 percentage points.

The results suggest that distance from home to the nearest general upper secondary education institution has become more important for enrolment over time, on average. One possible explanation for this finding could be that the compositions of students in general upper secondary education programmes has also changed over time. We know that a larger share of students with lower-educated parents attend general upper secondary education today compared to 30 years ago (see Table A5 in the Appendix for exact numbers). If the influence of distance is biggest among individuals on the margin of participation, as shown in previous international studies (Dickerson & McIntosh, 2013), the increasing share of students from less-educated families could explain the increased importance of distance over time.

Table 3.
Association between distance from home to general upper secondary education institution and enrolment and graduation, by time periods. Table report the average marginal effects from a logit model.

	Enrolment (y1)						Graduation (y2)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	1982 - 2014	1982 - 1987	1988 - 1994	1995 - 2001	2002 - 2006	2007 - 2014	1982 - 2012	2007 - 2012
Distance to nearest institution in km	-0.0026***	-0.0002	-0.0010*	-0.0023***	-0.0041***	-0.0047***	-0.0014***	-0.0034***
	(0.0002)	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0004)	(0.0002)	(0.0004)
Distance to nearest institution in km, squared	0.0000***	-0.0000	-0.0000	0.0000	0.0001***	0.0001***	0.0000	0.0001**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Average participation/graduation rate	0.54	0.44	0.50	0.56	0.57	0.66	0.51	0.61
Controls included								
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,061,470	407,867	448,750	371,092	288,330	545,431	1,927,451	411,412

Notes: Logit models. Table report average marginal effects. Standard errors in parentheses. Significant estimates are indicated by: *** p<0.001, ** p<0.01, * p<0.05. The set of individual variables includes a male dummy, birth year, immigration status. The set of parental characteristics includes mother's and father's educational levels and household annual income in natural logarithms. The set of geographical variables includes an urban dummy (over 20,000 inhabitants) and residential region (NUTS 2 level). Individuals with more than 30 km to the nearest general upper secondary education institution are excluded.

5.4. Heterogenous effects of distance

To test if distance has a larger impact on young people from less-educated homes, I split the sample by parental education, and re-run the models[10]. The results are presented in Table 4. To save space, I only present the results from the latest time period (2007 – 2014). The effect sizes are largest in this period (2007 – 2014), however the patterns are the same in all time periods. Overall, the results suggest that the cost of distance to school is larger in less-educated families.

Column 4 in Table 4 reports the AME of distance from home to general upper secondary school on enrolment (Y1), and Column 5 reports the AME of distance on graduation (Y2). Overall, the AME's is larger for enrolment that graduation, but the association is relatively modest for all groups. The first row shows the AME for young people where none of the parents hold qualifications beyond compulsory schooling (SES group 1). The second row presents the estimates for SES group 2 (where one or both of the parents have a vocational education). Finally results for SES group 3 (where at least one parent holds a medium- or short-cycle higher education college education) and SES group 4 (where at least one parent holds a university degree) are presented in row three and four respectively. The results suggest that the association between distance from home to school and participation is largest for young people where none of the parents hold a tertiary education (SES group 1 and SES group 2). The AME's are between -0.004 and -0.006. If one or both parents hold a medium- or short-cycle higher education, the AME's are between -0.002 and -0.004, and if one or both parents hold a university degree the AME's are between 0.000 and -0.002.

The results in Table 4 confirm the findings from previous international studies (Dickerson & McIntosh, 2013; Falch et al., 2013) which suggest that mainly students who are on the margin of participation are affected by distance.

[10] I divide the population into four groups according to their parent's highest education, in the year the young person turned 15 years old.

Table 4.
Association between distance from home to upper secondary education institution and enrolment and graduation, by parental education, 2007 - 2014. The table reports the average marginal effects from two distinct logit models.

			Y1	Y2
Groups	Parental education	Variables	Enrolment	Graduation
SES 1	Both parents have no education beyond primary school	Distance to general upper secondary institutions	-0.0051***	-0.0050***
			(0.0011)	(0.0012)
		Distance to general upper secondary education, squared	0.0001*	0.0001
			(0.0000)	(0.0000)
		Participation rate	*0.43*	*0.38*
		N	72,964	56,206
SES 2	One or both parents hold a vocational education	Distance to general upper secondary institutions	-0.0059***	-0.0042***
			(0.0006)	(0.0007)
		Distance to general upper secondary education, squared	0.0002***	0.0001***
			(0.0000)	(0.0000)
		Participation rate	*0.56*	*0.52*
		N	230,551	176,584
SES 3	One or both parents hold a shorter- or medium cycle higher education	Distance to general upper secondary institutions	-0.0040***	-0.0023**
			(0.0006)	(0.0007)
		Distance to general upper secondary education, squared	0.0001***	0.0000
			(0.0000)	(0.0000)
		Participation rate	*0.78*	*0.74*
		N	172,578	128,626
SES 4	One or both parents hold a long-cycle higher education	Distance to general upper secondary institutions	-0.0017*	-0.0003
			(0.0007)	(0.0010)
		Distance to general upper secondary education, squared	0.0000	-0.0000
			(0.0000)	(0.0000)
		Participation rate	*0.91*	*0.87*
		N	69,338	49,996

Note: Logit models. Table reports average marginal effects of distance (in km) from home to upper secondary education institution on participation. Standard errors in parentheses. Significant estimates are indicated by: *** p<0.001, ** p<0.01, * p<0.05. The set of individual variables includes a male dummy, birth year, immigration status, mother's and father's educational levels and household annual income in natural logarithms. The set of geographical variables includes an urban dummy (over 20,000 inhabitants) and residential region (NUTS 2 level). Individuals with more than 30 km to the nearest general upper secondary education institution are excluded.

6. Empirical results: Changes in distance

6.1. Visualisation and descriptive statistics of changes in distance from home to general upper secondary education institution

During the past four decades, 37 general upper secondary education institutions (out of 305 institutions in total) have been opened or moved to a new location. In the same period, 35 institutions have closed or moved. Table A1 in the appendix gives an overview of the openings and closures of general upper secondary education institutions each year. The majority of openings and closures have happened in areas where the access to general upper secondary education was already good. This implies that the majority of the openings and closures did not change the geographical accessibility to general upper secondary education significantly.

Only a small part of the sibling population experienced that their nearest general upper secondary education institution either opened up or closed down. In total, 43,125 of the 701,613 siblings in the sample (about 6 percent) experienced a change in distance due to an opening or a closure. However only 13,340 (about 2 percent) experienced a change in distance above one kilometre (positive or negative) (see Table A6 in Appendix).

Table 5 below summarises statistics of changes in distance from home to the nearest general upper secondary education institution for each time period. A negative distance change is equivalent to decrease in distance to school and a positive change means that the distance has increased since the oldest sibling was 15 years old. The largest changes in distance happened from 2002 – 2014, where both minimum and maximum changes are largest (minus 17 and plus 23 km respectively). However, despite the fact that the mean is close to zero in the period 2007 to 2014, the minimum and maximum changes in distance are also substantial in magnitude between 2007 and 2014. The mean is close to zero in 2007 – 2014, because the averages of the positive and negative values tend to nullify each other out.

Table 5.
Summary statistics of changes in distance from home to the nearest general upper secondary education institution by time period.

Time period	Mean	Median	Min.	Max.	N
1982-1987	-2.39	-2.23	-3.88	-1.03	164
1988-1994	1.95	2.26	-3.88	4.45	1,031
1995-2001	-2.54	-1.77	-17.71	3.19	7,798
2002-2006	-1.27	-1.48	-17.25	23.86	1,667
2007-2014	-0.06	-1.01	-16.16	23.55	2,678

Note: Table shows summary statistics of changes in distance from home to the nearest general upper secondary education institution for siblings in the sample.

Figure 6 maps the geographical dispersion of changes in distance from home to the nearest general upper secondary education institution. The figure shows that two municipalities stand out. One is Vesthimmerlands municipality in the northern part of Jutland where the affected individuals experienced an average of 5.6 km increase in distance to school because of a general upper secondary school is closed. The other municipality that stands out is Ringsted municipality, where individuals had 11.7 km shorter distance on average because of the opening of a general upper secondary education institution. The majority of the municipalities have experienced average changes in distance close to zero. However, this could potentially cover both large increases and decreases in distance that nullify each other, hence I examine the minimum and maximum changes in distance in Figures 7 and 8.

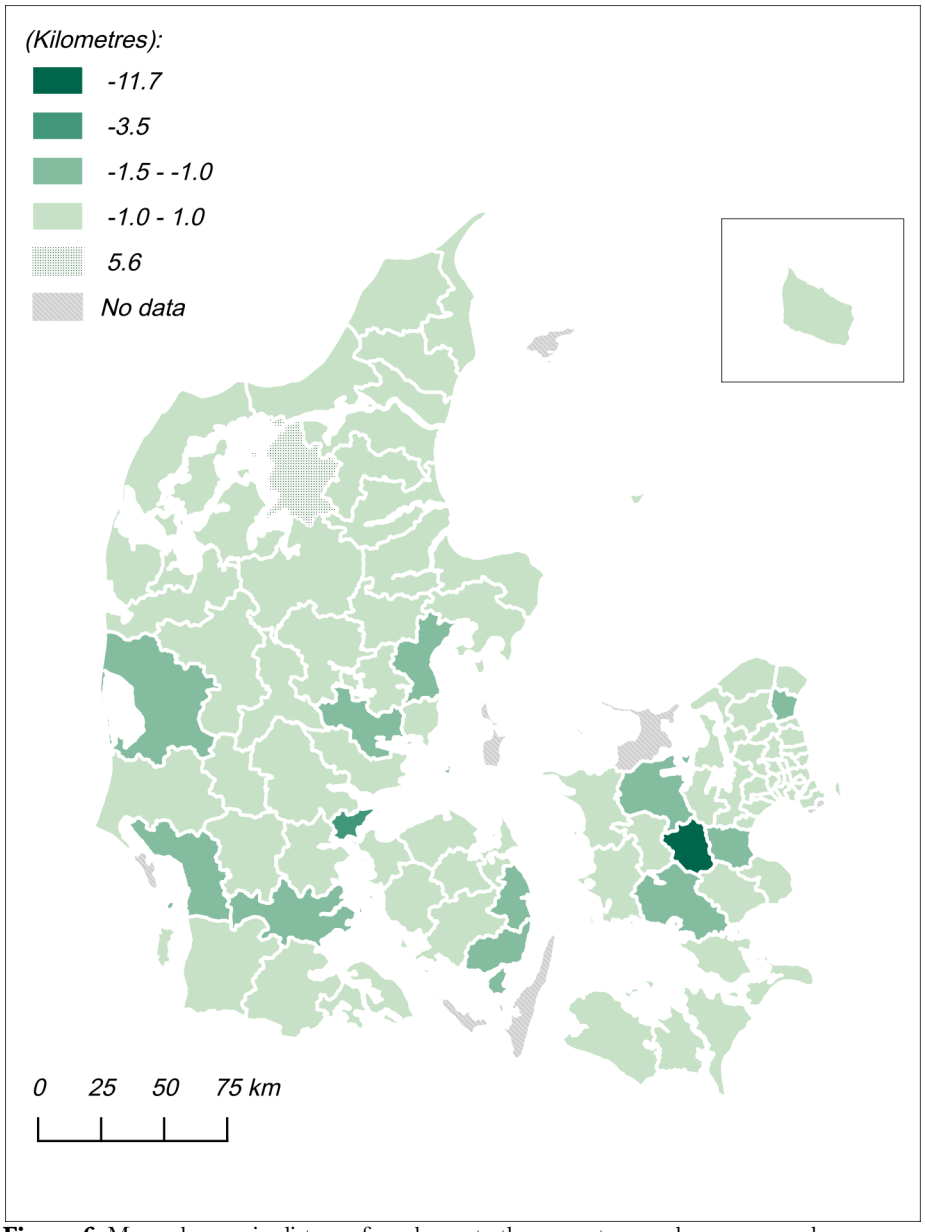

Figure 6. Mean changes in distance from home to the nearest general upper secondary education institution by municipality, caused by school closures and openings, 1982 – 2014. The figure shows the average change in distance from home to general upper secondary education institution among siblings who experienced different distances to general upper secondary education institution because of openings or closures of institution.

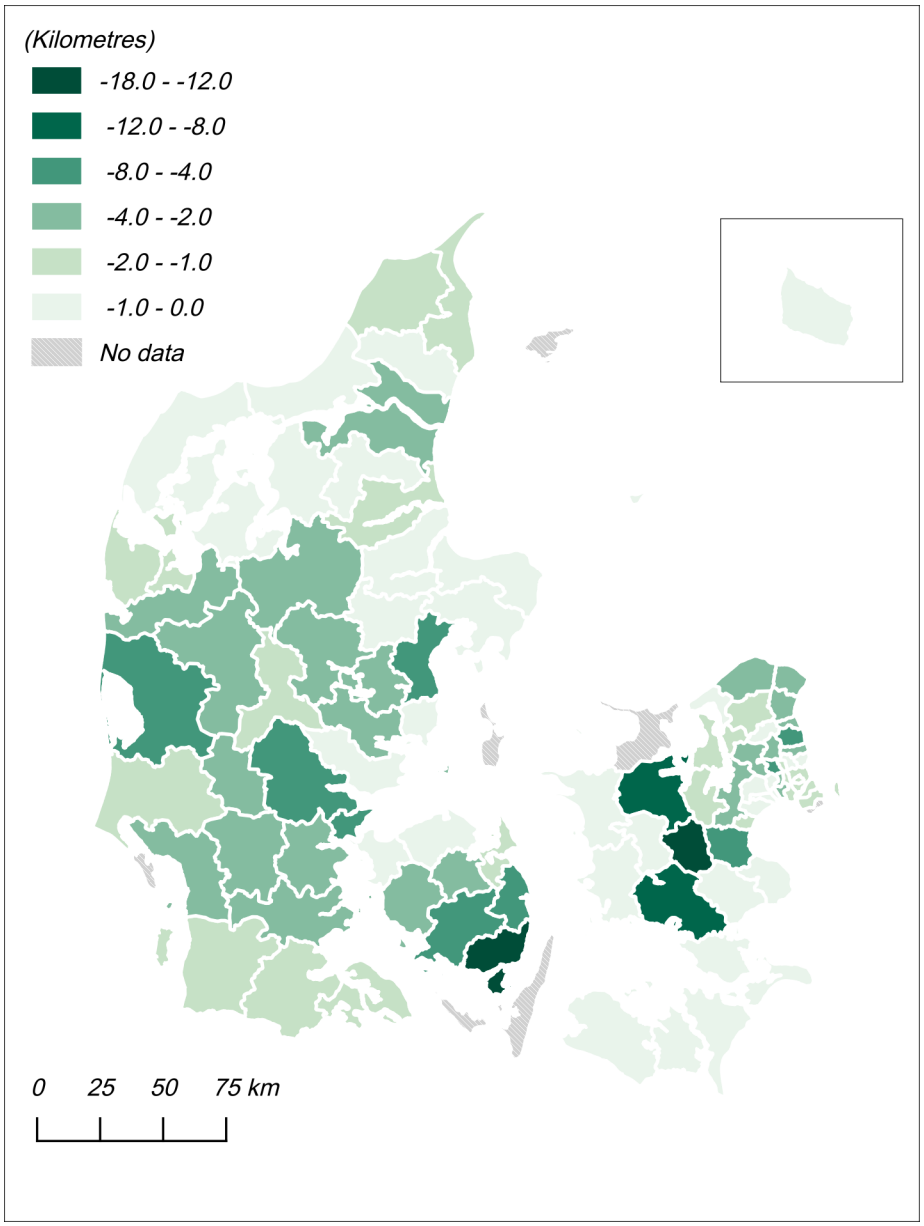

Figure 7. Minimum changes in distance from home to the nearest general upper secondary education institution by municipality, caused by school closures and openings, 1982 – 2014. The map shows the minimum change in distance from home to general upper secondary education institution among siblings who experienced different distances to general upper secondary education institution, 1982 – 2014.

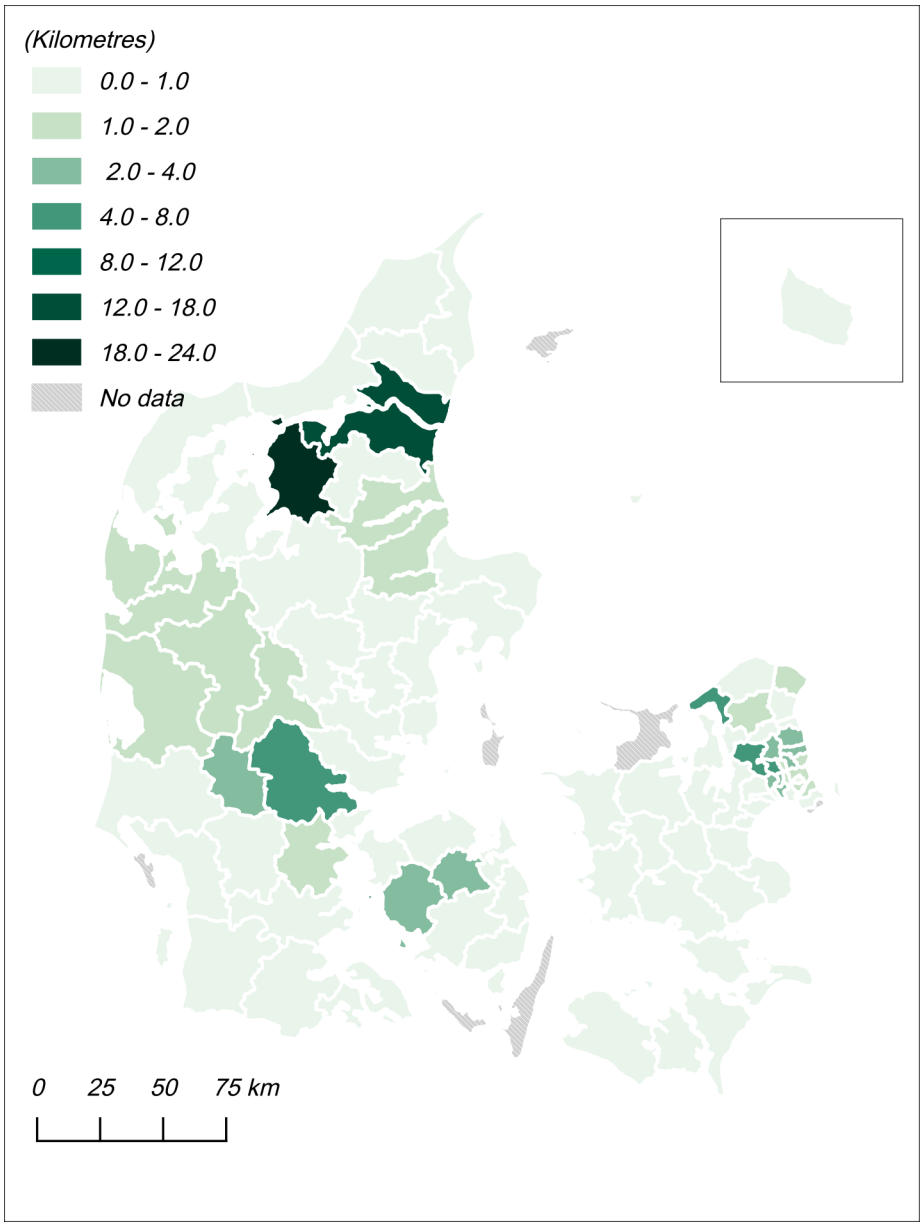

Figure 8. Maximum changes in distance from home to the nearest general upper secondary education institution by municipality, caused by school closures and openings, 1982 – 2014. The map shows maximum change in distance from home to general upper secondary education institution among siblings who experienced different distances to general upper secondary education institution, 1982 – 2014.

6.2. Regression models. Causal estimates from sibling FE models

Table 6 reports the probability of enrolment with respect to *changes in distance* to the nearest general upper secondary education institution. The estimates are derived from a sibling FE model (Equation 2.) (described in section 3.3). I use the sub-population of siblings who grew up in the same family and lived at the same address at their respective 15th birthdays, from 1982 – 2014. I control for gender and father's highest education. Other control variables are irrelevant as they are constant within families.

The first column (1) shows the overall result, using data from the whole time period (1982 – 2014). The parameter estimate is very small and not statistically significant from zero. This result is confirmed in the other time periods (column (2) to column (5)). The only exception is the most recent time period (2007 – 2014) in column (6). The results suggest a substantial effect of changes in distance to school. It shows that a one-kilometre increase in distance reduced the probability of enrolment by about 5 percentage points. Thus, the causal estimates are about ten times higher than the estimates derived from the logit model. However, the estimates are only significant in the most recent time period from 2007 – 2014.

At first sight, this is surprising. However, there are several possible explanations for this finding. First, the changes in distance were larger from 2007 and onwards (as shown in Table 5). Thus, the effect can only be estimated if actual variation exists. Another possible explanation could be that the composition of participants in general upper secondary education changed significantly during the 30 years studied (1982 – 2019), so an increasing share of young people from non-academic homes are enrolling in general upper secondary education. In the 1980s, about 27 percent of children from family backgrounds with no qualifications beyond compulsory schooling enrolled in general upper secondary education (see Table A5 in Appendix). About 30 years later, (in 2012) the share was 43 percent. As we know that the effect of distance from home to the nearest general upper secondary education institution is strongest among young people at the margin of participation in upper secondary education, this could explain why distance might have a larger impact on enrolment today than 30 years ago.

Table 6.
Association between changes in distance from home to the nearest general upper secondary education institution and enrolment, within siblings. By time periods. Table reports the average marginal effects from a sibling fixed effects model.

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Over all mean 1982 - 2014	Time 1 1982 - 1987	Time 2 1988 - 1994	Time 3 1995 - 2001	Time 4 2002 - 2006	Time 5 2007 - 2014
Distance to nearest general upper secondary education	-0.0009	0.0286	0.0039	0.0103	0.0094	-0.0499***
	(0.0011)	(0.0164)	(0.0079)	(0.0196)	(0.0055)	(0.0141)
Gender dummy (male = 1)	-0.1763***	-0.1771***	-0.1856***	-0.1895***	-0.1716***	-0.1498***
	(0.0009)	(0.0029)	(0.0027)	(0.0031)	(0.0039)	(0.0021)
Average high school participation rate	0.58	0.46	0.53	0.59	0.61	0.69
Controls included						
Father's highest education	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.0502	0.0481	0.0526	0.0573	0.0495	0.0413
Number of sibling pairs	553,125	144,511	215,641	182,288	151,526	179,379
Observations	1,254,738	217,413	299,419	244,800	189,807	303,299

Notes: Sibling fixed effect models. Table reports average marginal effects. Standard errors in parentheses. Significant estimates are indicated by: *** p<0.001, ** p<0.01, * p<0.05. To keep residential areas fixed, siblings who did not live at the same address when they were 15 years old were excluded from the analysis.

7. Summary and Concluding Remarks

This paper has provided empirical evidence on the effect of distance from home to the nearest general upper secondary education institution on enrolment in and graduation from general upper secondary education programmes in Denmark. The results suggest that distance from home to the nearest general upper secondary education institution has a small overall effect on enrolment within three years after compulsory education. When distance from home to the nearest general upper secondary education institution increases by one kilometre, the likelihood of enrolling in general upper secondary education is reduced by 0.3 percentage points on average. These results hold for a number of different specifications of distance to school. The effect of distance increases over the period examined in this study. It is negligible from 1982 – 1987 and more substantial from 2007 – 2014.

This small effect is perhaps unsurprising given the density of general upper secondary education institutions in Denmark and the fact that 72 percent of young people have less than 10 kilometres from home to their nearest general upper secondary education institution when they leave primary school. For many young people, distance from their home to upper secondary school is not an issue, as they live in walking or cycling distance of a general upper secondary education institution. In addition, many young people have a clear idea of their future academic plans, and for those individuals, distance from home to the nearest general upper secondary education institution is not a barrier to pursuing future education. However, for those individuals on the margin between enrolment or non-enrolment in general upper secondary education, distance from their home to their nearest general upper secondary education institution could potentially play an important role, as shown by previous studies (Dickerson & McIntosh, 2013; Falch et al., 2013). This paper confirms the heterogenous effects of distance from home to the nearest general upper secondary education institution. I find that the influence of this distance is most prevalent among young people from non-academic family backgrounds, and the influence diminishes if one of the parents holds a university degree.

This paper contributes to the literature by providing new causal evidence on the effect of distance from home to the nearest general upper secondary education institution. Using an extensive dataset of 33 birth cohorts of Danish young people, I use historical changes in location of general upper secondary education institutions to estimate the effects of changes in distance from home

to the nearest general upper secondary education institution on enrolment in and graduation from upper secondary education. The results show a large effect of distance from home to the nearest general upper secondary education institution for those young people who experienced that their nearest general upper secondary education institution moved further away (because of a school closure) or moved closer (due to a school opening). I find that a one-kilometre increase in distance to school reduces participation by approximately 5 percentage points. Thus, the causal estimates are about ten times higher than the estimates derived from the logit model. However, the results are only significant in the most recent time period from 2007 – 2014. These results confirm the finding that distance from home to the nearest general upper secondary education institution has greater effect on enrolment in and graduation from general upper secondary education programmes today than it had 30 years ago.

References

Andersen, J. G., & Etzerodt, S. F. (2018). Placering af offentlige institutioner (Location of public institutions). *Økonomi & Politik*, *91*(4), 34–49.

Austin, B., Glaeser, E., & Summers, L. (2018). Jobs for the Heartland: Place-Based Policies in 21st-Century America. *Brookings Papers on Economic Activity*, *49*(1 (Spring)), 151–255.

Bek, K. D. (2021, March 26). Kaare Dybvad: Storbyerne skal ikke være rigmandsenklaver, som lærere besøger i arbejdstiden. *Altinget*.

Card, D. (1995). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. *Aspects of Labour Market Behavior: Essays in Honour of John Vanderkamp*, 201–222.

Corcoran, J., & Faggian, A. (2017). 1. Graduate migration and regional development: An international perspective. In *Graduate Migration and Regional Development*. Edward Elgar Publishing.

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.

Danske Gymnasier. (2019). *Pejlemærker for fremtidens ungdomsuddannelser—(Point of orientations for upper secondary educations of the future)* (p. 12). Copenhagen.

Danske Regioner. (2019). *Fremtidssikrede ungdomsuddannelser—(Future-proof upper secondary educations)*.

DEA. (2019). *Sammenhængen mellem afstand og unges valg og gennemførelse af ungdomsuddannelser*. Copenhagen.

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).

Dickerson, A., & McIntosh, S. (2013). The Impact of Distance to Nearest Education Institution on the Post-compulsory Education Participation Decision. *Urban Studies*, *50*(4), 742–758.

Eliasson, K., Haapanen, M., & Westerlund, O. (2020). Regional concentration of university graduates: The role of high school grades and parental background. *European Urban and Regional Studies*, *27*(4), 398–414.

Faggian, A., & Franklin, R. S. (2014). Human Capital Redistribution in the USA: The Migration of the College-bound. *Spatial Economic Analysis*, *9*(4), 376–395.

Faggian, A., & McCann, P. (2009). Human capital and regional development. In *Handbook of Regional Growth Development Theories*. Edward Elgar.

Falch, T., Lujala, P., & Strøm, B. (2013). Geographical constraints and educational attainment. *Regional Science and Urban Economics*, *43*(1), 164–176.

Florida, R. (2002). *The Rise of the Creative Class: And How It's Transforming Work, Leisure, Community and Everyday Life*. Basic Books.

Frederiksen, M. (2021, January 1). Mette Frederiksens nytårstale 1. Januar 2021 ("Mette Frederiksens New Year's Speech January 1st. 2021"). In *New Year's Speech 2021*. The Prime Minister's Office.

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

Gibbons, S., & Vignoles, A. (2012). Geography, choice and participation in higher education in England. *Regional Science and Urban Economics*, *42*(1–2), 98–113.

Glaeser, E. (2011). *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier /*. New York: Penguin Press.

Heinesen, E., & Sørensen, E. S. (2021). Vocational versus general upper secondary education and labor market outcomes. *The Rockwool Foundation Research*, (Study paper 156), 65.

Iammarino, S., Rodriguez-Pose, A., & Storper, M. (2019). Regional inequality in Europe: Evidence, theory and policy implications. *Journal of Economic Geography*, *19*(2), 273–298. doi: 10.1093/jeg/lby021

Jakobsen, R. G. (2020). Unge på erhvervsuddannelser har dobbelt så langt til skole som gymnasieelever—(Young people enrolled in vocational upper secondary educations have twice as long distance to school as pupils in general upper secondary educations). *Momentum*, *12. årgang*(15. udgave). /momentum-nr-15-15-september-2020.pdf

Jensen, M. T. (2019). *Choices of higher education*. Aarhus University, Department of Economi and Business Economics.

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

Larsen, J. N., Andersen, H. T., Haldrup, K., Hansen, A. R., Jacobsen, M. H., & Jensen, J. O. (2014). *Boligmarkedet uden for de store byer* (SBI rapport No. 2014:05; p. 142). Aalborg Universitet.

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

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. doi: 10.1080/17421770601009791

Sørensen, E. S. (2021a, April). *Gaps in Higher Education Choice and Participation: Understanding the Role of Distance to Higher Education*. Presented at the Working Paper, Copenhagen.

Sørensen, E. S. (2021b). *Human capital, migration and regional inequality: How the location of educational institutions influences our society* (PhD thesis). University of Copenhagen, Copenhagen.

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

The Ministry of Higher Education and Science. (2016). *The Danish Education System* (p. 24).

Toft, L. V. (2020). *De unges afstand til ungdomsuddannelse—(Distance to upper secondary education)*. Copenhagen: KL - analysenotat.

White, P. M., & Lee, D. M. (2020). Geographic Inequalities and Access to Higher Education: Is the Proximity to Higher Education Institution Associated with the Probability of Attendance in England? *Research in Higher Education*, *61*(7), 825–848.

Appendix

Table A1.
Number of openings and closures of general upper secondary education institutions over time.

Year	Opening	Closure
1982	1	0
1983	0	0
1984	0	1
1985	0	0
1986	1	0
1987	0	2
1988	1	0
1989	0	0
1990	0	3
1991	0	3
1992	0	2
1993	0	0
1994	0	0
1995	14	1
1996	1	1
1997	0	2
1998	1	1
1999	0	0
2000	0	0
2001	2	0
2002	0	3
2003	0	0
2004	0	2
2005	1	1
2006	1	3
2007	4	0
2008	1	1
2009	2	4
2010	5	2
2011	0	0
2012	1	2
2013	1	1
Total	37	35

Table A2.
Summary statistics on distance to nearest general upper secondary education institution.

Year groups	Median	Mean
1982-1987	4.49	7.21
1988-1994	4.81	7.37
1995-2001	4.43	7.17
2002-2006	4.25	7.09
2007-2014	4.07	6.88

Table A3.
Distance from home to general upper secondary education institution, summary statistics by parental highest education.

Parents' highest education	Mean distance	Std. Dev distance	N
1: Mother and father no education	7.74	(7.2)	424,679
2: Mother or father vocational edu.	7.54	(6.8)	909,172
3: Mother or father tertiary edu.	6.79	(6.3)	550,791
4: Mother or father holds university degree	4.82	(5.2)	189,435

Note: table reports means, standard deviations and number of observations. Distances are in km.

Table A4.
Distance distribution in 5-kilometre intervals.

Distance	Freq.	Pct.	Cum. Pct.
0 - 5 km	1,101,001	53.08	53.08
5 - 10 km	388,263	18.72	71.80
10 - 15 km	291,979	14.08	85.88
15 - 20 km	187,004	9.02	94.90
20 - 25 km	72,001	3.47	98.37
25 - 30 km	21,993	1.06	99.43
30 - 35 km	4,400	0.21	99.64
35 - 40 km	4,338	0.21	99.85
40 - 45 km	2,973	0.14	99.99
45 - 50 km	125	0.01	100.00
Total	2,074,077	100.00	

Table A5.
Enrolment rate in general upper secondary education and parental background.

Socio-Economic Status	1982-1987	1988-1994	1995-2001	2002-2006	2007-2014
1: Mother and father: no education	0.27	0.31	0.34	0.34	0.43
2: Mother or father: vocational edu.	0.40	0.44	0.49	0.48	0.56
3: Mother or father: tertiary edu.	0.65	0.68	0.72	0.72	0.78
4: Mother or father: university degree	0.86	0.88	0.89	0.88	0.91

Table A6.
Share of sibling population who experienced a change in distance to their nearest general upper secondary education institution.

	All changes		Changes above 1 km	
	Frequency	Percentage	Frequency	Percentage
No change	658,488	93.85	688,273	98.10
Change in distance	43,125	6.15	13,340	1.90
Total	701,613	100.00	701,613	100.00

Figures

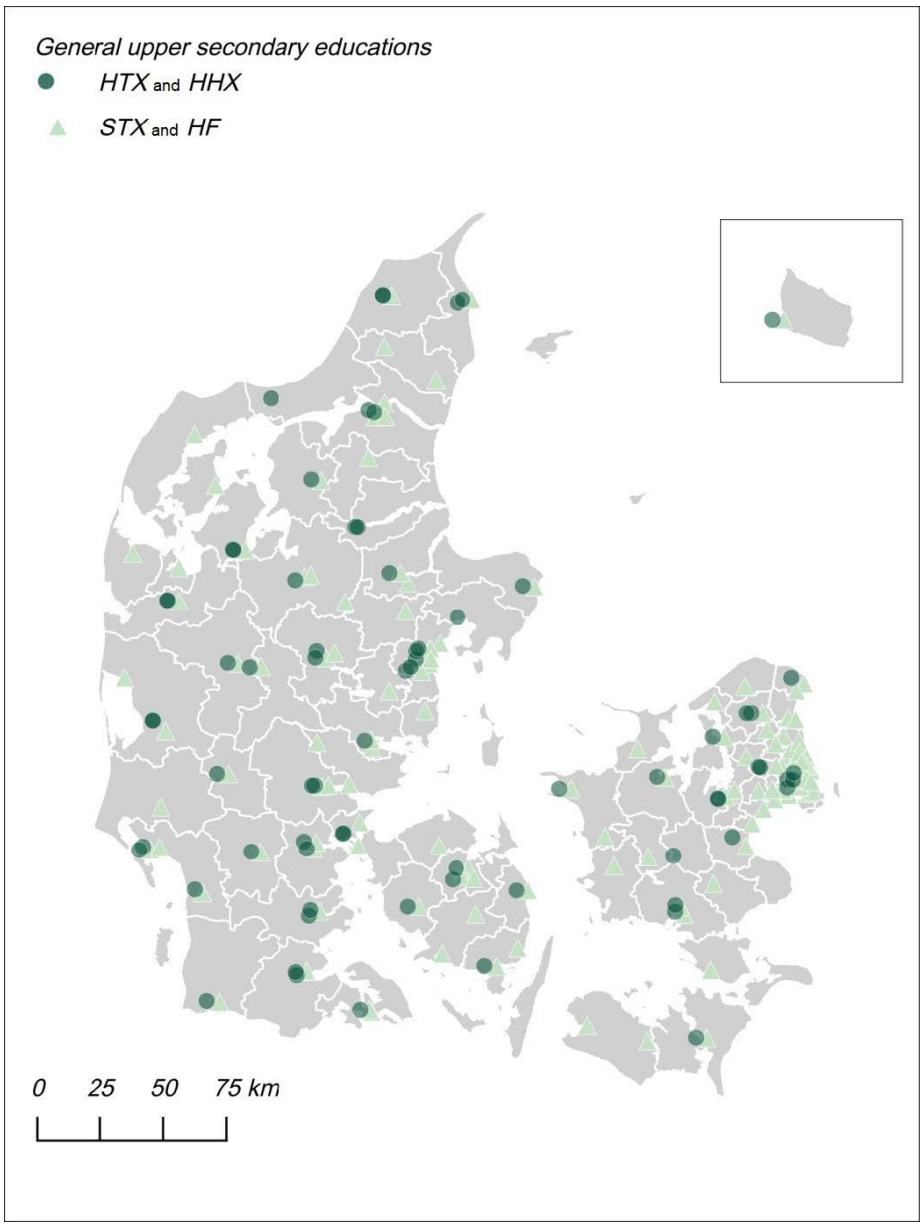

Figure A1. Location of general upper secondary education institutions in Denmark (2014).

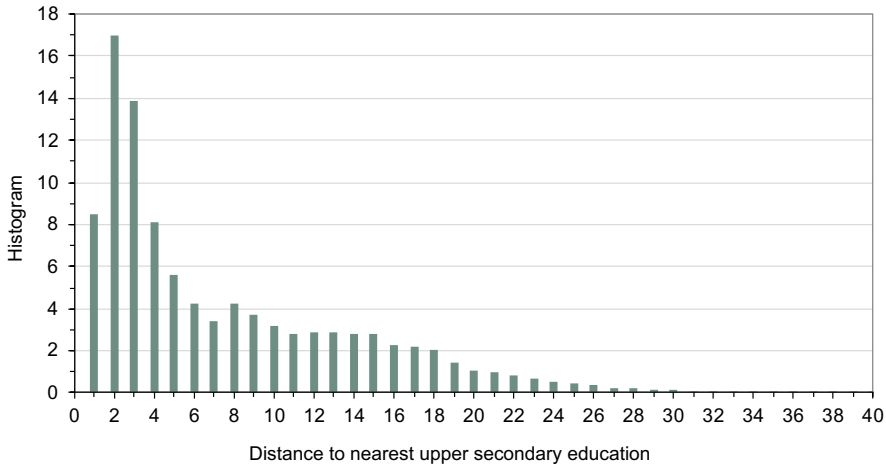

Figure A2. Distance to nearest upper secondary education institution, 1982 - 2014

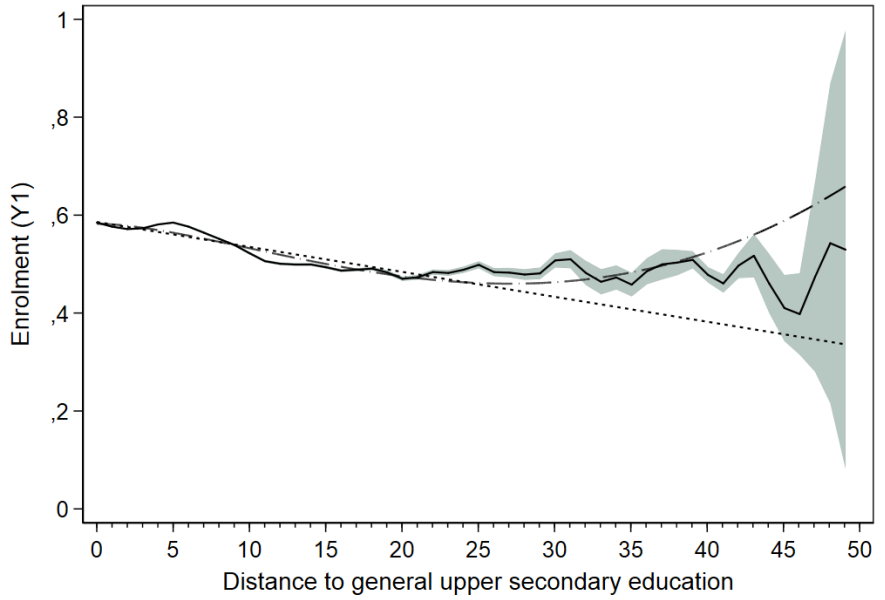

Figure A3. Association between distance to nearest general upper secondary education institution and enrolment, 0 – 50 km.

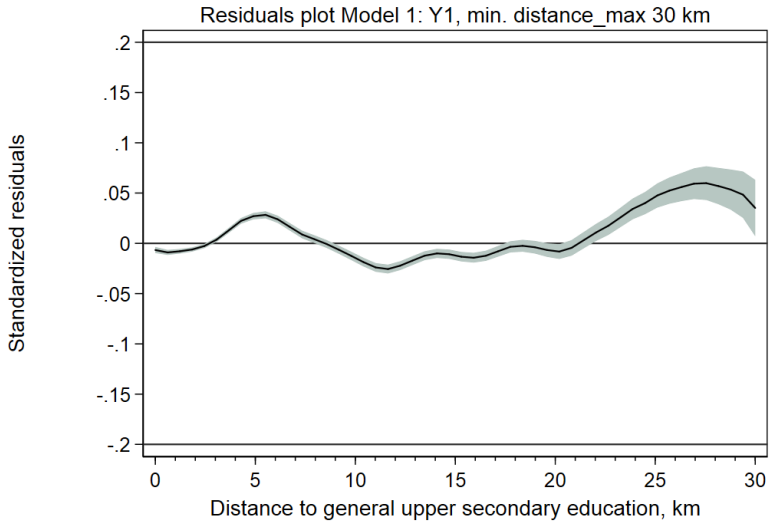

Figure A4. Association between distance to nearest general upper secondary education institution and enrolment, standardized residual plot, 0 – 30 km.

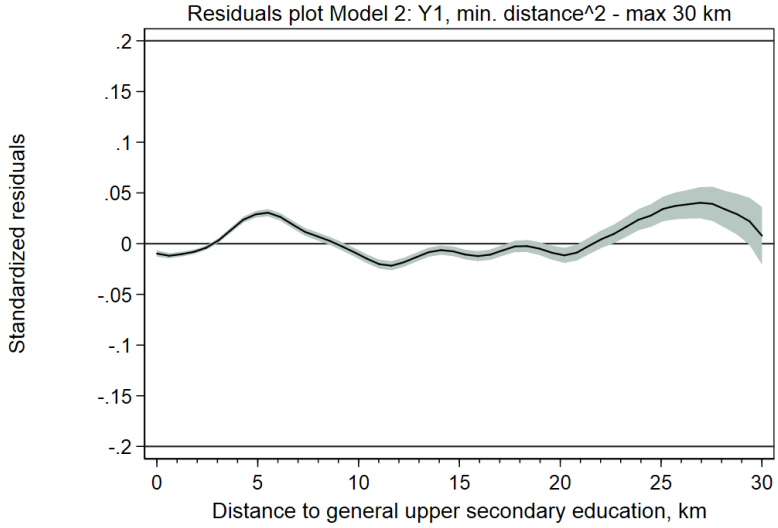

Figure A5. Association between distance (squared) to nearest general upper secondary education institution and enrolment, standardized residual plot, 0 – 30 km.

Paper 2

PAPER 2

Moving out of Rural Areas: Estimating the Causal Effect of General Upper Secondary Education on Youth Migration*

Elise Stenholt Sørensen[1] & Anders Holm[2]

Abstract

This paper provides novel evidence on the causal effect of general upper secondary education on young adults' migration choices. Using data on five birth cohorts who grew up in rural areas across Denmark, we find large effects of completing general upper secondary on decisions to move out of rural areas. We use distance to general upper secondary schools as an instrumental variable to account for the endogeneity in choice of post-compulsory education. We find that the causal effect of completing general upper secondary education varies across parental backgrounds. The effect is largest for young adults from low socioeconomic backgrounds, and is negligible if one of the parents holds a university degree. The evidence points to the importance of general upper secondary education as a driver of youth migration, but also suggests that parental background is a central factor in shaping young adults' migration patterns.

Keywords: interregional migration, general upper secondary education, human capital, graduate migration, distance to education

JEL: I20, J24, J61, R23, R58

---

* The authors thank Bo Honoré, Robert Andersen, Alessandra Faggian, Thomas Crossley and Charles Levine for reading previous versions of the paper and giving their helpful suggestions and comments. We also thank Steven Durlauf, Don Davis, Nikolai Kuminoff, Rasmus Landersø, Ismir Mulalic and Lars Pico Geerdsen for their constructive comments.

[1] Ph.D. student at Department of Geosciences and Natural Resource Management, University of Copenhagen and Research Fellow at Department of the Built Environment, Aalborg University

[2] Research Leader and Professor, The Rockwool Foundation

1. Introduction

This paper examines youth migration from rural to urban areas and the role of youth education in this process. The educational disparities between rural and urban areas have increased significantly during the last decades, both in the European Union (EU) and in the US[3]. The education gap between rural and urban areas has always existed, but over the last decades, the general rise in educational level has increased much faster in the large urban areas, which has resulted in increasing regional inequality (Berry & Glaeser, 2005; Eurostat, European Union, 2016; E. Glaeser, 2013; Parker et al., 2018). This development is a potential problem because educational disparities causes regional divergence in income, unemployment, economic prosperity (Austin, Glaeser & Summers, 2018). In a knowledge economy[4], high levels of human capital[5] are crucial for innovation and economic growth, both at a national and a regional level (Florida, 2003; Gennaioli, La Porta, Lopez-de-Silanes & Shleifer, 2013; Lucas, 1988; Moretti, 2004a; Rauch, 1993).

Empirical evidence documents that the regional distribution of human capital is an important determinant of regional development, in countries all over the world (Gennaioli et al., 2013). Rural areas that are losing population might therefore be concerned with the fact that young highly educated individuals are more likely to move than their less-educated peers (Faggian, Corcocan & Partridge, 2015; Faggian, Rajbhandari & Dotzel, 2017). If those migrating have more valuable skills than those staying behind, the net out-migration may leave the local labour market deprived of valuable human capital. In addition, a large literature argues that the aggregate level of human capital has a positive effect on productivity over and above its individual effect

[3] In the US, a growing share of residents aged 25 and older graduate from college. However, this growth has been much larger in urban areas. In 2000, 28% of urban residents had a bachelor's degree, compared with 15% in rural areas. In 2016, the share of urban residents with a bachelor's degree had increased to 35% in urban areas and 19% in rural areas (Parker et al., 2018). The same picture is seen in the EU, where 37% of the urban population aged 25-64 hold a tertiary education, compared with only 21% of those living in rural areas (Eurostat, European Union, 2016).

[4] Notion of the "post-industrial" service economy.

[5] Human capital refers to "embodied knowledge and skills" of a person and is often measured empirically as educational level (Becker, 1964; Faggian & McCann, 2009). In this paper, we use the terms "education" and "human capital" interchangeably.

(E. L. Glaeser, 1999; Moretti, 2004a, 2004b, 2004c; Rosenthal & Strange, 2008). The effect may arise if the presence of educated workers makes less educated workers more productive through knowledge sharing more (Moretti, 2004c). Moretti provides empirical evidence of human capital spillovers at the plant-level and also within cities in the US (Moretti, 2004a, 2004c). Furthermore, aggregate level of human capital is also found to have other social benefits in addition to its effect on productivity and earnings. For example, education reduces the probability of engaging in crime and rises political behaviour and voting (Moretti, 2004b). Thus, the local level of human capital has implications for both individual and social welfare.

The question of the significance of human capital for regional growth has spurred a large literature on regional development that occupies itself with how to attain and attract highly skilled individuals (Faggian et al., 2015, 2017). The literature has focused primarily on the migration rates of college or university graduates and their subsequent migration in relation to job search (Faggian & Franklin, 2014; Venhorst & Cörvers, 2018). However, this focus on the movement of graduates from higher education into the labour market and the inferred sorting from rural to urban areas could potentially underestimate the extent of such sorting (Ahlin, Andersson & Thulin, 2018). We know that young adults between 18 – 24 years are the most geographically mobile population group (Andersen & Nørgaard, 2018; Plane, 1993; Sørensen, 2021a, Chapter 4). In addition, young adults are in a period of their lives where educational choices are relevant. Often, they move in relation to enrolment into tertiary education after completion of general upper secondary education. Because young adults who migrate from rural regions to larger cities typically do not return to their rural home, the location decisions of young people have long-term implications for spatial disparities in human capital (Ahlin et al., 2018; Haapanen & Tervo, 2012). Hence, knowledge about the determinants of youth migration is important.

The focus of this paper is the role of *human capital* in individual migration decisions among youths. There is vast empirical evidence which shows that high-human-capital individuals are more migratory than others (Faggian & McCann, 2009a; Faggian et al., 2017). However, we still lack causal evidence about how much of this strong association is in fact due to education itself, and how much is due to unobserved characteristics of the people who engage in higher education. Malamud & Wozniak's (2012) study of American college graduates is an exception. They use exogenous variation in college attainment

and conclude that college attainment increases the probability of long-distance moves. The aim of the present study is to contribute to the field of graduate migration by providing causal evidence of the effect of general upper secondary education on the sorting of young adults into urban areas.

In addition, we contribute to the literature with knowledge about determinants of youth migration and the initial decision of individuals to move away from home. We examine the effect of general upper secondary education on individuals decision to move to an urban area. The majority of the empirical contributions on human capital and migration have been concerned with individuals who completed a college or a university degree (Faggian & Franklin, 2014). Only a few empirical studies have examined the migration patterns after general upper secondary school. Exception to this are the contributions by Cooke & Boyle (2011) and Faggian & Franklin (2014), which provide descriptive evidence of the migration patterns of high school leavers in the US.

However, our analysis departs from Cooke & Boyle (2011) and Faggian & Franklin (2014) in three aspects. First, we propose a different methodology that provides causal estimation results. Second, we are interested not only in the overall migration flows of individuals who graduated from a general upper secondary education but also in their migration patterns compared to their peers who did not graduate from a general upper secondary education[6]. By sampling 15-year-old individuals who are still living at home, we avoid sample selection problems regarding migration history. Thirdly, we examine the rural-urban migration flows with administrative population data at the individual level, instead of postsecondary education data analysed at the state level in the US. We are able to conduct an empirical analysis at the city level, because we can geo-locate each residential address in Denmark within (or outside) of each town in Denmark. The definitions of towns are based on the geographical delineation of towns provided by Statistics Denmark. It is defined according to the spatial distance between dwellings and is independent of administrative boundaries[7].

[6] In Denmark, about 40 percent of each birth cohort does not graduate from a general upper secondary education before turning 25 years. About 20 percent complete a vocational upper secondary education and the remaining 20 percent do not complete any post-compulsory education.

[7] We use the geographical delineation of towns and cities provided by Statistics Denmark. Statistics Denmark defines a "town" as an area with an agglomeration of 200 or more dwellings with less than 200 meters between each house.

The remainder of the paper is organized as follows. Section 2 reviews the previous empirical evidence about human capital and interregional migration. Section 3 presents the data and variables used in the empirical analysis. Section 4 provides descriptive evidence of the historical rural to urban migration trends in Denmark and the association with graduation from general upper secondary education. In addition, the section describes the Danish education system. Section 5 presents a theoretical model for young people's decision to enter general upper secondary education and to migrate out of a rural area. Section 6 describes the empirical models and presents the estimation results. Finally, section 7 offers some concluding remarks.

2. Human capital and interregional migration

Empirical studies agree that individual human capital is an important determinant of migration behaviour (Faggian et al., 2015). Numerous studies confirm that people with higher levels of education are more geographically mobile. The association between higher education and interregional migration has been shown empirically since the 1970s. For instance, Bartel (1979) reported that an individual's years of education were positively correlated with the probability of migrating within the US in the 1970s. These results have been confirmed in several studies, among others Schwartz (1976) and Levy & Wadycki (1974). Levy and Wadycki (1974) also found, that the better educated are less deterred by increased distance when deciding to migrate.

More recently, Silvia Dixon (2003) looked at the role of education on geographical mobility. Using panel survey data from Britain, she showed that higher-skilled individuals moved more often both for jobs and other reasons. In addition, she reported that longer-distance moves between regions are more likely to be motivated by a job change or enrolment in an education. Malamud & Wozniak (2012) were the first to examine the causal impact of a college education on the likelihood of migration. They exploited changes in the risk of induction into Armed Forces during the Vietnam conflict as a source of exogenous variation in college attainment in the US. Their paper concluded that college attainment significantly increased the probability of long-distance moves (Malamud & Wozniak, 2012).

2.1 Migration of highly educated graduates

More recently, empirical contributions have focused on the highly skilled population, typically university graduates, and how to attract and retain high-skilled labour in a region. These studies cover a range of European countries, the US, and Australia. Faggian, McCann and Sheppard (2006, 2007) and Faggian & McCann (2009) studied the migration behaviour of recent university graduates in the UK. Venhorst, Dijk & Wissen (2010, 2011) studied graduate migration in the Netherlands, while Ahlin et al., (2018) examined how long university graduates remained in urban regions in Sweden. Similarly, Haapanen & Tervo (2012) focused on the timing of migration after graduation in Finland. All of these studies reported that the majority of university graduates stayed in the city where they completed their university education.

Using US panel data, (Kodrzycki, 2001) found that 70 percent of graduates lived in the city where they graduated five years after graduation. Both (Busch & Weigert, 2010 and Haapanen & Tervo, 2012) applied duration models to describe the probability of migration during education and up until 10 years after graduation. Using German data, Busch and Weigert (2010) confirmed the results from Kodrzycki (2001) and reported that 70 percent of university graduates lived in their university region 10 years after graduation. In addition, they reported that a third of the out-migration occurred within the first year after graduation, and after five years the probability of moving out of the urban area was almost non-existent.

Overall, the empirical evidence suggests relatively limited regional mobility after university graduation. However, more recent studies have presented remarkable regional differences between peripheral regions and capital regions concerning the share of graduates who remain in their university region after graduation (Kotavaara, Kotavaara, Rusanen, & Muilu, 2018).

3. The Danish education system and rural-urban migration

In this section, we describe the overall changes in rural to urban migration in Denmark between 1989 – 2017. We also show how this development is associated with changes in general upper secondary education attainment among youths. We begin by describing the institutional context, namely the Danish education system, in order to clarify the educational choices that young adults face after completing compulsory education in Denmark.

3.1 The Danish education system

The education system in Denmark is universal. There are no tuition fees in upper secondary education or tertiary education. Compulsory education in Denmark consists of 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 education system, continue in 10th grade, or enter upper secondary education. Individuals that do not directly enrol in upper secondary education may enter later, with no loss of rights or opportunities for enrolment.

Danish upper secondary education comprises two main tracks of education: general upper secondary education and vocational upper secondary education. Most individuals (about 90 percent of a birth cohort) eventually

choose one of the two tracks. General upper secondary education consists of academic tracks, such as mathematics, technical studies, and linguistics. Vocational upper secondary education consists of occupation-specific tracks such as carpentry, bricklaying, mechanics, and hairdressing. The vocational system is a dual-education system, such that students shift between school-based learning and practical apprentice training. The two types of upper secondary education are placed in different institutions and are usually not located in the same geographical location.

Tertiary education in Denmark is on three levels: lower tertiary education, intermediate tertiary education, and higher tertiary education. A formal requirement for admission to all tertiary educations is usually a general upper secondary education. However, some lower and intermediate tertiary tracks may sometimes allow individuals to use selected vocational upper secondary educations to meet the admission requirements. The system is illustrated in Figure 1 below. Figure 1 shows that vocational upper secondary education is a 'dead end' in terms of the option to pursue further education, while general upper secondary education offers the option of enrolment in a tertiary education.

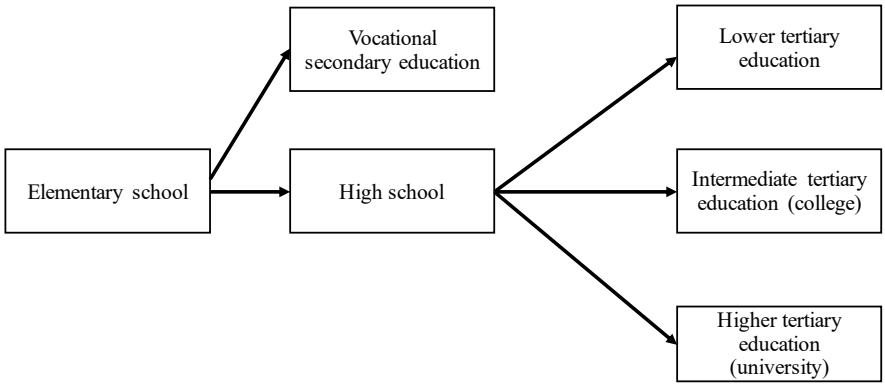

Figure 1. Typical pathways in the Danish education system
Note: A few lower and intermediate tertiary tracks may sometimes allow individuals to use selected vocational secondary qualifications to meet the admission requirements of the lower and intermediate tertiary tracks.

3.2 Changes in educational attainment among youth

In Denmark today, entry into general upper secondary education is the natural choice for many young people after compulsory education. However, it has not been like this historically. During the past four decades the trend has shifted towards an increasing share of elementary school leavers choosing to continue on to more general academic educations rather than vocational education programmes (Sørensen, 2021b , Chapter 4). Back in 1982, about 23 percent completed a general upper secondary education before the age of 24 years . Today about 57 percent of the population had graduated from a general upper secondary education at the age of 24 years. The historical development in educational attainment is described in more details in Chapter 4 in the present thesis.

3.3 Rural to urban migration patterns

Along with the changes in young adults' educational attainments, we have seen significant changes in rural to urban migration patterns among young people. Figure 2 illustrates the net migration rates from rural to urban areas by age groups in Denmark. It is evident that the negative net migration in rural areas is almost solely driven by the age group 18-24 years old. For older age groups, net migration is also negative, but much smaller and constant over time. The negative net migration of young adults in rural areas has almost doubled over the past three decades. In 1989 the net migration among 18 to 24-year-olds was about minus 8 percent. In 2017, the negative net migration rate was about 15 percent in the same age group. This suggests that young adults play a key role in the rural to urban migration development.

During the same period, the share of young adults completing general upper secondary education has increased significantly, both in rural and urban areas. The development in rural and urban areas is similar, however urban youths are more likely to complete general upper secondary education and rural youths are more likely to complete vocational upper secondary educations, see Figure A1 in the appendix.

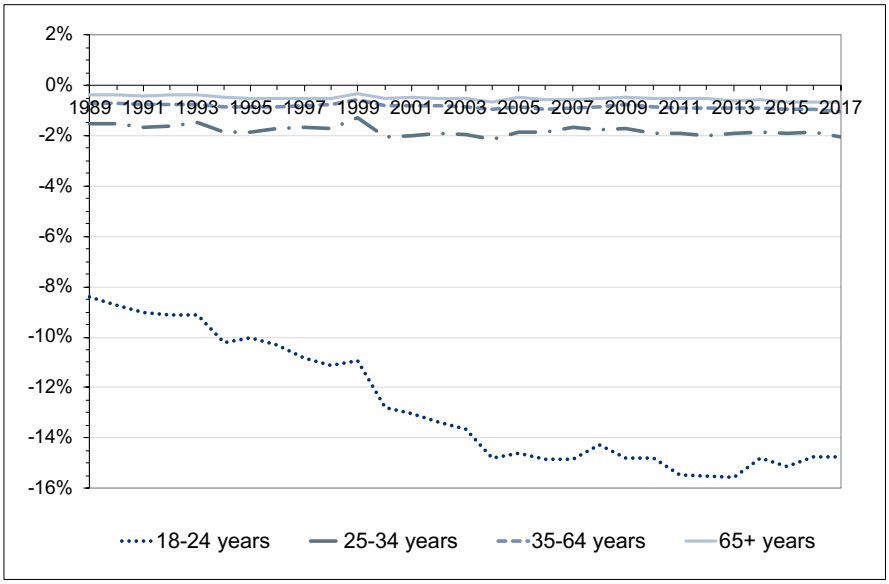

Figure 2. Net migration rates in rural areas by age group, 1989 – 2017

Note: Net migration is calculated using a full population sample of people living in rural areas and information regarding moving decisions each year. Net migration rate in rural areas is calculated using the formula below, for each age group: $N = \frac{I-E}{M} * 100$
Where $\boldsymbol{N}$ denotes Net migration rate in percentage points. $\boldsymbol{I}$ denotes the number of immigrants (people moving in) and E denotes the number of emigrants (people moving out). $\mathbf{M}$ denotes mid-year population, that is: $M = \frac{Population\ at\ start\ of\ year + Population\ at\ end\ of\ year}{2}$
Sources: Administrative registers from Statistics Denmark (BEF, BOL, BYSTR).

Figure 3 illustrates how the growth in youth migration towards urban areas largely follows the growth in young people who holds a general upper secondary education in rural areas. The dashed line in Figure 3 shows that the share of rural young adults completing a general upper secondary education has increased from 33 percent in 1989 to 57 percent in 2017 (right axis). The solid line illustrates increasing out-migration from rural areas among 18 to 24-year-old adults, (left axis).

This association suggests that growth in general upper secondary education among young adults could be a driver of the increasing youth migration out of rural areas. If the association between general upper secondary education and migration among youths is a causal relationship, it implies that an unintended consequence of increasing the share that obtains a general upper

secondary education among rural youths is that a larger share of young people will move to urban areas.

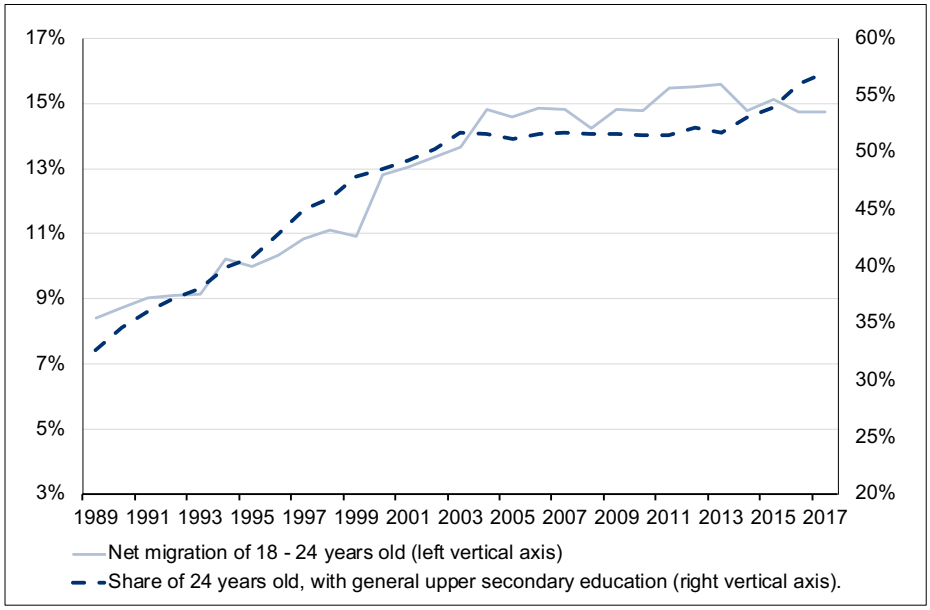

Figure 3. Growth in out-migration from rural areas and educational attainment, 1989 – 2017
Note: For the period 1989 – 2017, the figure shows the net migration out of rural areas and the share of 24-year-olds in rural areas who hold a general upper secondary education. In cases where a person completes two different youth educations before turning 24, the education completed first is used. The population is divided into a rural and an urban population on the basis of a person's residential address when he or she was 17 years old. Persons without residential address in Denmark when they were 17 and 24 years old are omitted.
Source: Own calculations based on administrative data from Statistics Denmark (KOTRE, 2019).

3.3 Mechanisms

Which mechanisms might explain the relationship between general upper secondary education and migration? One central mechanism is that enrolment into tertiary educations is an important pull factor when young adults decide to leave their (rural) residence of origin (Faggian & Franklin, 2014; Gulløv & Gulløv, 2020). The majority of higher education institutions (about 80 percent) are located in the four largest cities in Denmark. Especially universities are located in the largest urban areas, whereas other types of tertiary educations (university colleges and business academies) are more geographically dispersed

(see Chapter 4 in the PhD introduction). Many graduates from general upper secondary education enrol in a tertiary education in the same year, or the year before or after they move to an urban area (66 percent of the movers enrol, as shown in Figure A2 in the appendix). This shows that two-thirds of the general upper secondary education graduates who leave their rural residence decide to move in relation to enrolment in higher education. The last third of the movers enrol later, or do not enrol in higher educations at all.

The numbers indicate that school-to-school migration plays a central role, but there are also other factors that influence rural to urban migration among young adults. Other factors could be the academic values of peers and teachers in general upper secondary school, the urban amenities, job possibilities, population diversity, and so forth (Gulløv & Gulløv, 2020). However, it is beyond the scope of this analysis to document all the mechanisms. In the following section, we present a theoretical model of the influence of educational choice on migration.

4. Theoretical model

In this section, we present a theoretical model of individual choice of completing general upper secondary school (HS) and the subsequent choice of migrating from rural to large urban areas (Move). General upper secondary education will be denoted "high school" in the present section. This section first illustrates a potential mechanism behind the causal effect of general upper secondary education on the decision to leave a rural area. Second, our theoretical model illustrates the endogenous aspect of choosing high school.

In the model the individual and his or her family are faced with two consecutive choices: first, whether to complete high school, and second, whether to move out of the rural area. We note that tertiary educations in Denmark are typically located in urban areas. The majority of high school students, therefore, have to move to urban areas to enrol in tertiary education.

For each choice there is instantaneous utility associated with the choice. For the completion of high school this is denoted u_{hs}. Utility for not taking high school (including enrolling and dropping out) is normalized to zero. The instantaneous utility of completing high school may consist of the utility of adhering to family norms and the cost of studying. Therefore, u_{hs} may be either positive or negative.

If the individual decides to move after completing high school, the instantaneous utility contains two components, one which is known when completing high school u_{hs}^{m}, and a component that is first realized after moving e_{hs}^{m}. If the individual decided not to complete high school and decides to move, utility likewise contains two components, one which is known when deciding not to enrol u_{hs}^{m}, and a component that is first realized if moving e_{hs}^{m}. The instantaneous utility if the individual decided not to move is normalized to zero. Special attention should be given to the two unknown terms realised after moving u_{hs}^{m} and e_{hs}^{m}. At the time before entering high school these are option values. We consider the option values of high school as the utility of working in an urban labour market and the option value of entering tertiary education. Entering tertiary education would involve long commuting time if staying in a rural area. The situation is illustrated in Figure 4 below.

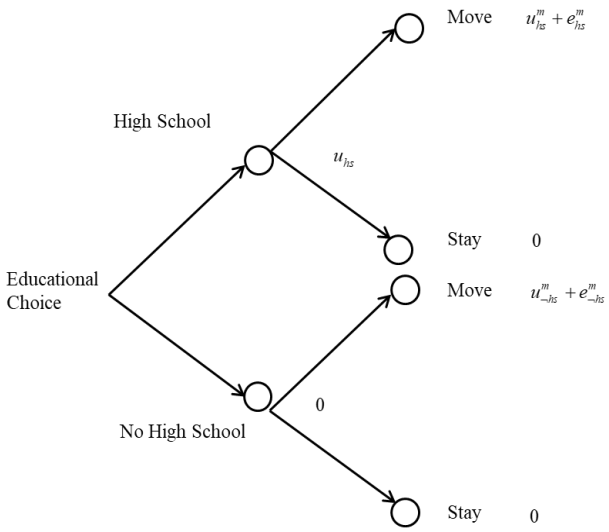

Figure 4. Choice behaviour

When the individual decides whether to enter high school, he or she has to take into account not only the instantaneous utility from either entering or not entering high school, but also the expected maximum utility from either moving or not moving, given the choice of going to high school. Upon assuming that the *e*'s are standard normally distributed, the expected maximum utility from entering and completing high school is

$$V_{HS} = \underbrace{u_e}_{\substack{\text{Instantanious Value of completing HS}}} + \underbrace{E\max\left[u^m_{hs} + e^m_{hs}, 0\right]}_{\substack{\text{Expected value of being able to make the optimal choice between moving versus staying with HS}}} = u_e + \underbrace{u^m_{hs}\Phi\left(u^m_{hs}\right)}_{\substack{\text{Expected value of moving}}} + \underbrace{\varphi\left(u^m_{hs}\right)}_{\substack{\text{Option value of being able to make the optimal choice between moving and staying}}}$$

where $\Phi(.)$ and $\varphi(.)$ are the standard normal cdf and pdf respectively. Similarly, we find the expected value of not completing high school:

$$V_{\neg HS} = E\max\left[u^m_{\neg hs} + e^m_{\neg hs}, 0\right] = u^m_{\neg hs}\Phi\left(u^m_{\neg hs}\right) + \varphi\left(u^m_{\neg hs}\right).$$

Now the choice of completing is thus:

$$V_{HS} > V_{\ HS}$$

$$\Leftrightarrow$$

$$u_e > \underbrace{u^m{}_{hs}\Phi\left(u^m{}_{hs}\right) - u^m_{hs}\Phi\left(u^m_{hs}\right)}_{\text{Differences in expeteed value of moving}} + \underbrace{\varphi\left(u^m{}_{hs}\right) - \varphi\left(u^m_{hs}\right)}_{\text{Difference in option value of moving}}.$$

From this we see that to choose high school, either the excepted value and/or the option value of moving with high school must outweigh the similar values without high school and the difference must justify the cost of completing high school. Thus, if u_{hs} is negative, the net difference of moving with HS relative to non-high school must be positive.

Once high school is completed, moving with HS is: $e^m_{hs} > u^m_{hs}$ and similarly for moving without HS. From the perspective of the individual, the choice of high school is deterministic, whereas the choice to move is probabilistic at the time of choosing high school. However, both choices may have to be modelled as probabilistic if components of the choices are unobserved (both components known to the individual and components not known). Further, in this case, the choice of high school may, from the perspective of the econometrician, depend on unobserved components determining the decision to move, and hence the two components must be estimated simultaneously.

We note that moving with high school entails the opportunity to enter tertiary education, hence a relatively 'large' option value. Moving without a high school degree entails only the opportunity to enter skilled or unskilled occupations, which gives a relatively 'low' option value because the labour market conditions for unskilled and skilled workers may not differ much from rural to urban areas. Hence, the option value of high school will in many cases be larger than option values without high school. Option value may, therefore, be an important driver of the choice to enter high school, and a major value of high school is the option of entering tertiary education in the future.

5. Data and variables

5.1 Sample restrictions

We use panel data from the population registers of Statistics Denmark in combination with detailed distance measures between place of residence and nearest general upper secondary education, and distance to the nearest urban area. The registers contain rich information about individual characteristics, such as gender, immigration status, age, educational attainment, grades in elementary school, geography of place of residence, and parental education and annual income. Our sample encompasses five birth cohorts: all individuals born from 1986 – 1990 who at age 15 were resident in a rural area in Denmark. A rural area is defined as non-urban areas, villages and towns with 20,000 inhabitants or less in 2016. Our definition of 'village' follows that used by Statistics Denmark, which defines a village (or town) as an area with interconnected houses, the distance between which does not exceed 200 metres (Danmarks Statistik, 2020; Sørensen, 2021a; Sørensen & Holm, 2019). Figure 5 below shows the location of rural and urban areas in Denmark. We also restrict our sample to young individuals who lived in a rural area when they were 15 to 18 years old. This is done to ensure that the individuals in our sample do not move before graduating from general upper secondary school.

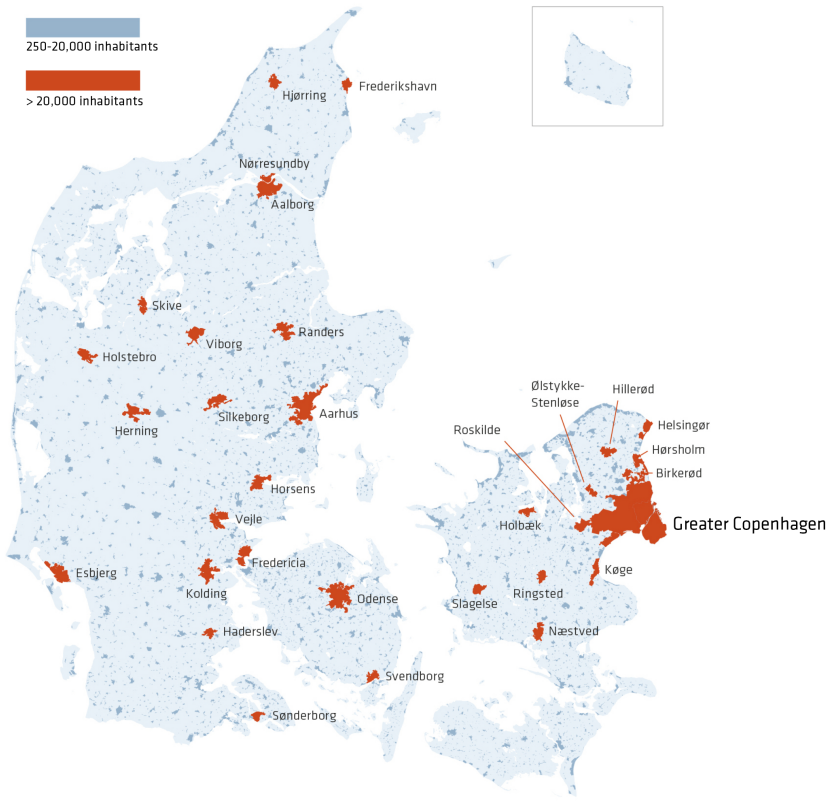

Figure 5. Map of rural and urban areas in Denmark in 2016
Note: We defined a rural area as non-urban areas and towns with 20,000 inhabitants or less in 2016. Residential addresses situated in rural areas in 2016 keep their rural status throughout the entire study irrespective of whether their populations rose above 20.000 inhabitants after 2016. they may have a different status in previous years. The 33 largest towns in Denmark are defined as urban areas. Our sample shows that 60 percent of 15-year-olds resided in rural areas in the period 1986-90.
Source: Own calculations based on administrative data from Statistics Denmark.

5.2 Definition of interregional migration

We define interregional migration as a change of residential address from a rural area to an urban area. The dependent variable is a dummy variable, which is equal to one if an individual has moved to an urban area before turning 25. Only moves away from the childhood home are counted as moves. This means that individuals who move together with their parents are not defined as movers. We account for repeated moves from the age of 18 years until the age of 25 years. If just one of the moves goes towards an urban area, we categorise it as a move.

5.3 Variables and summary statistics

The explanatory variables measure individual characteristics, family characteristics, municipality of residence, and distance to the nearest urban area. All variables are measured for individuals in the sample at age 15. The only exception is our primary explanatory variable that indicates if an individual graduates from general upper secondary school before turning 25. The definitions of all variables are described in Table A1 in the appendix. "General upper secondary education" is a dummy variable, which is equal to one if the individual has graduated from general upper secondary school before turning 25 years. If the individual has completed a vocational upper secondary education or has no post-compulsory education, the variable is equal to zero. About 12 percent of the sample completed both a vocational- and a general upper secondary education before turning 25. In these cases, we include the first completed upper secondary education in our analysis.

We present descriptive statistics of all the variables, in Table 1 below. About 71 percent of the sample moved to an urban area before turning 25 years and about 54 percent completed general upper secondary school. The relatively low completion rate for general upper secondary school, from an international perspective, reflects the structure of upper secondary education in Denmark, where a large number of individuals complete vocational upper secondary education rather than general upper secondary school. Among general upper secondary school graduates who grew up in rural areas, about 90 percent move to an urban area before the age of 25 years. This only applies to about 50 percent of the young adults who attained a vocational upper secondary education or who did not complete any post-compulsory education. About 47 percent of the sample are females and 96 percent are non-immigrants. We have information about each individual's exam grade in 9th grade in written Danish and math. The grade point average is 5.5 in written Danish and 5.3 in Maths, which is lower than the national average of 6.35 and 6.2 respectively. Information about the national GPA average comes from the Danish Ministry of Education from 2012 (The Ministry of Education, 2020).

In our sample, the mean distance to the nearest general upper secondary school is 10.7 kilometres and the mean distance to the nearest urban area is about 26 kilometres. We also include municipality fixed effects in the models, to take local economic measures (such as youth unemployment) and local amenities into account. We also control for parental education and household income. The distribution of mother's and father's highest education

is also shown in Table 1. About one-fourth of the mothers and one-fourth of the fathers have no education past elementary school. A total of 39 percent of the mothers and 46 percent of the fathers completed a vocational upper secondary education. This means that only about 30 percent of the mothers and 20 percent of the fathers have a further education at the tertiary level. The education distribution reflects both the age cohort of the parents (typically born in the 1950s and 1960s) and that the parents live in rural areas.

Table 1
Descriptive statistics.

Variables	Variable names	Mean	Std. Dev.	Min.	Max
Dependent variable	Move	0.71	0.45	0	1
Individual characteristics					
Youth education	General upper secondary education	0.54	0.50	0	1
Gender	Female	0.47	0.50	0	1
Immigration status	Danish	0.96	0.20	0	1
	First-generation immigrant	0.03	0.17	0	1
	Second-generation immigrant	0.01	0.10	0	1
Year of birth	1986	0.19	0.39	0	1
	1987	0.19	0.39	0	1
	1988	0.20	0.40	0	1
	1989	0.20	0.40	0	1
	1990	0.22	0.41	0	1
Grades from the final examination in secondary school	Written Danish	5.46	3.04	-3	12
	Missing grade in Danish	0.08	0.27	0	1
	Written Maths	5.29	3.19	-3	12
	Missing grade in Maths	0.08	0.27	0	1
Geographical characteristics					
	Municipality Fixed Effects *				
	Distance to school	10.67	6.87	0	46
	Distance to urban area	26.05	15.37	1	88

(continued on next page)

Table 1 (continued)

Variables	Variable names	Mean	Std. Dev.	Min.	Max
Family characteristics					
Mother's education	Missing education information	0.02	0.15	0	1
	Elementary school	0.26	0.44	0	1
	General upper secondary education	0.04	0.20	0	1
	Vocational upper secondary education	0.39	0.49	0	1
	Short-cycle higher education	0.04	0.19	0	1
	Medium-cycle higher education	0.22	0.41	0	1
	Long-cycle higher education	0.03	0.18	0	1
Father's education	Missing education information	0.06	0.24	0	1
	Elementary school	0.23	0.42	0	1
	General upper secondary education	0.03	0.17	0	1
	Vocational upper secondary education	0.46	0.50	0	1
	Short-cycle higher education	0.04	0.20	0	1
	Medium-cycle higher education	0.11	0.31	0	1
	Long-cycle higher education	0.06	0.23	0	1
Household income	Household income (log)	12.79	0.38	6	17
	Poor	0.01	0.10	0	1
	Rich	0.01	0.10	0	1

Note: N = 146,375.

5.4 Distance to general upper secondary school

Our instrumental variable is the travel distance from place of residence, at age 15, to the nearest general upper secondary school. To lower the number of calculations, we locate each address in a 100 x 100-meter geographically defined quadrant. We measure the distance from the bottom-left corner of each quadrant to all of the general upper secondary school institutions, through the road network. Young adults living on small islands, without a bridge to the mainland, are excluded from the analysis. In total, 5 percent of the sample is excluded on this basis.

The average distance to general upper secondary school is 10 kilometres, with a minimum distance of 0 km and a maximum distance of 46 kilometres to the nearest general upper secondary school. About 50 percent of the population have less than ten kilometres to the nearest general upper secondary school and about 23 percent have less than 5 kilometres (Table A1 in the appendix shows the cumulative distance distribution in 5-kilometre intervals). Keeping in mind that our population of young adults live in smaller towns or rural areas, this seems reasonable. The families have to arrange and pay for their transportation to general upper secondary schools, as there are no school buses provided. However, the public transportation system is relatively well-developed in Denmark and the students' transportation costs are subsidized by the state. In addition, there are no tuition fees or additional costs for attending general upper secondary school.

6. Empirical models and results

6.1 Empirical strategy

The starting point of our empirical analysis is a simple linear probability model (LPM):

$$P(y_i = 1|H_i, \boldsymbol{x}_i) = \alpha + \delta H_i + \boldsymbol{\beta x}_i + \varepsilon_i \qquad (1),$$

where subscript *i* denotes individuals and where $\alpha, \delta, \boldsymbol{\beta}$ are parameters to be estimated. y is a binary endogenous variable equal to one if the individual moves into an urban area before the age of 25 and equal to zero otherwise. A vector of control variables is denoted by **x,** and *H* is a variable indicating if the

individual has completed general upper secondary school, which takes the value of one if general upper secondary school is completed and zero otherwise. All variables are described in Table 2. Further, ε is an error term that captures the effect of omitted variables.

It is reasonable to assume that the decision to move (y) and the decision to complete general upper secondary school (H) share the same unobserved characteristics. If parental aspirations affect both the decision to move and the decision of post-compulsory education, the decision to complete general upper secondary education is an endogenous variable. To circumvent the endogeneity of H we employ several empirical approaches: a family fixed effects (FFE) approach, and an IV approach. The two models will be described in the following sections.

Family Fixed Effects model

We employ a family fixed effect estimation to take unobserved family characteristics into account. We use the subsample of siblings on our original sample. This amounts to 41,172 siblings and 20,586 families in total. The error term in equation (1) ε is decomposed into two terms. The first is a_f , a family fixed effect that captures family aspirations and possible genetic endowments that may be correlated with completing general upper secondary school. The second term u_{fs} is an individual component that is assumed to be independent of completing general upper secondary school. The latter assumption may be questionable, and we test it by adding important individual controls, such as school grades. The family fixed effect model is:

$$\mathrm{P}\left(y_{fs} = 1 \middle| H_{fs}, \boldsymbol{x}_{fs}\right) = \delta H_{fs} + \beta' \boldsymbol{x}_{\mathrm{fs}} + a_f + u_{fs} \qquad (2),$$

Instrumental variable model

The family fixed effect model purges unobserved heterogeneity between families. However, it cannot take unobserved heterogeneity between siblings into account. Also, it restricts the sample to siblings. To take individual unobserved characteristics into account, we employ an instrumental variable approach. We introduce an auxiliary regression equation stating that completing general upper secondary school depends on the instrumental variable distance to nearest general upper secondary school. The assumption is that the instrumental variable only affects completing general upper secondary

school, but not whether one moves to an urban area. The first-stage equation is:

$$\mathrm{P}(H_i = 1|\boldsymbol{x}_i, z_i) = \tau + \boldsymbol{\theta}\boldsymbol{x}_i + \gamma z_i + \epsilon_i \qquad (3),$$

where z is the instrumental variable (distance to general upper secondary school). We expect ϵ and ε to be correlated due to common omitted variables. In the case where the treatment effect δ is heterogeneous, the interpretation of the 2SLS or Wald estimate of δ is the local average treatment effect (LATE), see (Imbens & Angrist, 1994). This means that we estimate the treatment effect for the compliers. In our case, the compliers are those who are affected by distance when choosing a post-compulsory education. This group will choose a general upper secondary education if a general upper secondary school is located close to their home and will not choose a general upper secondary education if the distance is too far from their home.

Many researchers have used a similar instrumental variable, distance to college, in studying the return to college education e.g. (Cameron & Taber, 2004; Card, 1995; Carneiro et al., 2011; Currie & Moretti, 2003).

The identification strategy relies on the assumption that, conditional on $\mathbf{x}$, distance is independent of ϵ and ε. If some characteristics are unobserved but relevant in both equations in the system (1) and (3), then distance being independent of ϵ and ε amounts to assuming that unobserved characteristics are independent of distance to general upper secondary school. This may not be a tenable assumption, for example, Cameron & Taber (2004) show that distance to college is correlated with cognitive ability. We add cognitive ability by using selected school grades and apply a broad range of control variables.

6.2 Main results

The main results of the three models are presented in Table 2 below. We compare the results from the linear probability model (model 1) with results from the family fixed effects model (model 2) and the IV model (model 3). We find a substantial effect of general upper secondary education attainment on the probability of moving to an urban area. On average, the probability of moving increases by 26 percentage points if an individual has a general upper secondary

education, according to the estimates from the LPM model[8]. The family fixed effect estimates in model 2 purges all family fixed effects. We find that the family fixed effects estimate is about 0.22, thus about 4 percentage points lower than the LPM estimates. Hence, we would expect some of the estimated effects in the LPM estimate to cover some omitted variable bias alleviated by our family fixed effects design.

Table 2
Main results.

Dependent variable: Moving to an urban area	Model 1	Model 2	Model 3
	LPM	FFE	IV
General upper secondary education	0.2641***	0.2219***	0.6587***
	(0.0027)	(0.0070)	(0.0750)
Individual controls	Yes	Yes	Yes
Family controls	Yes	Yes	Yes
Municipality FE	Yes	Yes	Yes
IV	No	No	Yes
Number of observations	146,375	41,172	146,375
R2	0.2341	0.0930	
F-test (IV)			110.73

Notes: *** p<0.001, ** p<0.01, * p<0.05. Standard errors in parentheses. Table estimates report the average marginal effects of general upper secondary education and move to an urban area. Control variables are gender, immigrant status, year dummies, GPA in Maths and Danish in secondary school, distance to urban area, father's highest education, mother's highest education, household income and municipality dummies.

Table 2 also shows our second-stage results from the IV regressions in model 3. The estimate suggests a causal impact of general upper secondary education on migration that is larger in magnitude than the estimates in model 1 and 2. The probability of moving increases by 66 percentage points if an individual has a general upper secondary education compared to if they do not have a general

[8] Table A2 in the appendix compares estimates derived from the LPM model with Average Marginal Effect (AME) from a logit model. The results show that the AME from a logit model is about 4 percentage points lower than the estimate derived from OLS.

upper secondary education. The first-stage results are shown in Table A4 in the appendix and show that distance is a strong instrument for completion of general upper secondary education.

In sum, all three models indicate large effects of general upper secondary education attainment on the probability of moving to an urban area. On average we find that outmigration rates go up by approximately 22 - 26 percentage points when an individual obtains a general upper secondary education (model 1 and 2). We find even larger effects for compliers using distance to general upper secondary schools as an instrument for choice of education. The estimates indicate a local average treatment effect of 66 percentage points (model 3).

The large difference between the estimates of the models may be rooted in the difference between the effect for the compliers (a local average treatment effect) and the average effect in the total population (LPM and FFE). Those affected by distance in their choice of general upper secondary education (compliers) may have a low baseline probability of moving without a general upper secondary education, and their decision to move is very much affected by completing a general upper secondary education education. In contrast, non-compliers move to urban areas regardless of having a general upper secondary education. This explanation is supported by our findings for heterogeneous effects, described in section 6.4.

6.3 Robustness concerning our definition of an urban area

When we perform robustness checks on our definition of an urban area, we see that our main results are robust in relation to other definitions of urban areas. In the main analysis, an urban area is defined as towns above 20,000 inhabitants (y1). To perform the robustness check, we run three alternative models. One model defines an urban area as one of the four largest cities in Denmark (Copenhagen, Aarhus, Aalborg and Odense) (y2). In another model we define an urban area as a town with at least one university (y3), and lastly, we run the analysis where only moves to the Capital area, Copenhagen, are measured (y4). Table 3 and 4 present the results from the LPM and the IV model, respectively.

In the LPM regressions, the two alternative definitions y2 and y3 do not change the results substantially (see Table 3). In the third definition (y4), the parameter estimate of general upper secondary education is substantially smaller (14 percentage points compared to 26 percentage points). This is not surprising, as y4 only accounts for moves to the capital area. Looking at the IV

estimates in Table 4, we see no statistically significant difference in the estimates across the four different y-variables.

Table 3
Robustness concerning the definition of an urban area: LPM regressions.

	y1 Move to large town	y2 Move to one of the four largest cities	y3 Move to university town	y4 Move to Copenhagen city
General upper secondary education	0.2641***	0.3163***	0.3090***	0.1439***
	(0.0027)	(0.0028)	(0.0028)	(0.0025)
Intercept	-0.1603***	-0.6727***	-0.5285***	-0.0804*
	(0.0512)	(0.0528)	(0.0533)	(0.0461)
Individual controls	Yes	Yes	Yes	Yes
Family controls	Yes	Yes	Yes	Yes
Municipality FE	Yes	Yes	Yes	Yes
IV	No	No	No	No
Observations	146,375	146,375	146,375	146,375
R2	0.2341	0.3243	0.2889	0.2699

Notes: *** p<0.001, ** p<0.01, * p<0.05. Standard errors in parentheses. Table estimates report the average marginal effects of general upper secondary education on move to an urban area. Control variables are gender, immigrant status, year dummies, GPA in Maths and Danish in secondary school, distance to urban area, father's highest education, mother's highest education, household income and municipality dummies.

Table 4
Robustness concerning the definition of an urban area: IV regressions.

	y1 Move to large town	y2 Move to one of the four largest cities	y3 Move to university town	y4 Move to Copenhagen city
General upper secondary education	0.6587***	0.8784***	0.7054***	0.7085***
	(0.0750)	(0.0817)	(0.0777)	(0.0737)
Intercept	0.3132***	0.0019	-0.0527	0.5971***
	(0.1052)	(0.1146)	(0.1091)	(0.1035)
Individual controls	Yes	Yes	Yes	Yes
Family controls	Yes	Yes	Yes	Yes
Municipality FE	Yes	Yes	Yes	Yes
IV	Yes	Yes	Yes	Yes
Observations	146,375	146,375	146,375	146,375

Notes: *** p<0.001, ** p<0.01, * p<0.05. Standard errors in parentheses. Table estimates report the average marginal effects of general upper secondary education on move to an urban area. Control variables are gender, immigrant status, year dummies, GPA in Maths and Danish in secondary school, distance to urban area, father's highest education, mother's highest education, household income and municipality dummies.

6.4 Heterogeneous effects of family background

To see if there are heterogenous effects of general upper secondary education in our study, we split our sample according to family background. Parental background is an important driver for educational choice (S. R. Lucas, 2001). A study by Ahlin et al. (2018) reports significant spatial sorting by general upper secondary education grades and parental education, among university graduates. Studying heterogeneous effects across family background may shed light on the mechanisms behind our large complier effect in Table 3. We expect compliers to vary across parental background, because previous studies have shown that individuals from lower-income families are mostly affected by distance to upper secondary educations (Dickerson & McIntosh, 2013; Falch, Lujala, & Strøm, 2013; Sørensen, 2021b).

To investigate the heterogenous effects, we split the sample into four groups based on the parents' highest education. The four groups are presented in Table 6 alongside summary statistics on distance to general upper secondary

school. Parents with no secondary education comprise SES group 1, which comprises 14 percent of the sample. The largest group, containing almost half of the sample, is SES group 2, which represents parents with a vocational upper secondary education. A further 30 percent of the sample consists of SES group 3, which represents individuals who have at least one of the parents who holds a short- or medium-cycle higher education. Finally, parents with a long-cycle higher education, SES group 4, make up seven percent of the sample.

We divide the sample according to the highest education of the two parents. This means that at least one of the parents holds a vocational upper secondary education in SES group 2, and that at least one of the parents holds a university degree in SES group 4 and so forth. The skewed educational distribution towards relatively short-term educations of the parents reflects both the age group of the parents (typically born in the 1950s and 1960s), and the fact that the parents constitute a rural population.

Table 5
Distance to general upper secondary education by parental background.

		Parental education		Distance to nearest education			
Groups		Freq.	Pct.	Mean	Std. Dev.	Min	Max
SES 1:	Both parents have no education beyond elementary school	21,084	14%	11.43	7.29	0.00	45.15
SES 2:	One or both parents hold a vocational upper secondary education	71,141	49%	11.11	6.96	0.00	46.12
SES 3:	One or both parents hold a shorter- or medium-cycle higher education	43,302	30%	10.10	6.55	0.01	45.23
SES 4:	One or both parents hold a long-cycle higher education	10,848	7%	8.67	6.18	0.06	45.04

Note: 146,375 observations

We show LPM and second-stage results split by family background in Table 6 below. The results suggest a declining effect of completing general upper secondary school across family background in both the LPM and the IV model. The LPM results are listed in the first column. It shows that, on average, the effect of general upper secondary education is identical for SES group 1 – 3,

about 25-26 percentage points, and lower for SES group 4, about 21 percentage points. The baseline results indicate the probability of moving (y=1) in the control group (H=0) holding all other covariates constant at their means. The baseline estimates show that individuals with well-educated parents are more likely to move, all other things held equal.

The IV results, reported in the second column in Table 6, support the heterogenous finding from the LPM model. For SES group 1 (unskilled parents), we find an effect size of almost 80 percentage points. The effect size drops to 66 percentage points for SES group 2, and 43 percentage points for SES group 3. The effect is statistically insignificant for SES group 4 because there is no significant first stage for this group. This is because the fraction of compliers in SES group 4 is very small.

We conjecture that individuals with very well-educated parents are, to a large extent, 'always takers' and hence enrol in general upper secondary education irrespective of the distance to general upper secondary schools. In contrast, the lower SES groups are more inclined to enrol in general upper secondary education if the costs of doing so, including transportation costs, are low. This is supported by the first-stage estimates split by SES group, as seen in Table A5 in the appendix. We find that the effect of distance declines across SES groups, so that the association between distance and general upper secondary education attainment is strongest if none of the parents holds an education above secondary school (SES group 1) and the association is absent if one of the parents hold a university degree (SES group 4).

In sum, we find large, significant complier effects of completing general upper secondary education on the probability of moving to an urban area. However, if one of the parents holds a university degree, the effect of general upper secondary education is negligible.

Table 6
LPM and IV estimations split by parental education.

Groups	Parental education		LPM	IV
SES 1	Both parents have no education beyond elementary school	General upper secondary education	0.2609***	0.7936***
			(0.0084)	(0.2066)
		P(Y=1) baseline	0.4609***	
			(0.0041)	
		Observations	21,084	21,084
SES 2	One or both parents hold a vocational upper secondary education	General upper secondary education	0.2664***	0.6647***
			(0.0040)	(0.1023)
		P(Y=1) baseline	0.5294***	
			(0.0025)	
		Observations	71,141	71,141
SES 3	One or both parents hold a shorter- or medium-cycle higher education	General upper secondary education	0.2535***	0.4285***
			(0.0044)	(0.1260)
		P(Y=1) baseline	0.6435***	
			(0.0035)	
		Observations	43,302	43,302
SES 4	One or both parents hold a long-cycle higher education	General upper secondary education	0.2066***	0.9376
			(0.0081)	(0.7253)
		P(Y=1) baseline	0.7432***	
			(0.0074)	
		Observations	10,848	10,848

Note: *** p<0.001, ** p<0.01, * p<0.05. Standard errors in parentheses. Table estimates show the marginal effects of completing general upper secondary education on moving to an urban area. P(Y=1) baseline indicates the conditional probability of moving (y=1) holding all other covariates constant at their means.

7. Conclusion

This paper is motivated by an increasing rural-urban divide, witnessed in Denmark, in much of Europe, and in North America. We have seen that the educational disparities between rural and urban areas have increased significantly during the last decades. As human capital is an important determinant of regional development, the increase in educational disparities has consequences for the rural population and potentially the social cohesion of society. We show that an essential factor that contributes to this rural-urban divide is an increasing net-migration of young adults from rural to urban areas. We also show that choice of post-compulsory education is crucial to young adults' migration decisions. It is well known that highly educated individuals are more geographically mobile than others. It is also known that individuals with a university degree tend to stay in the urban area where they have attained their degrees. Thus, knowledge about young adults' initial migration decisions is important, as it has long-term implications for the regional distribution of human capital.

Using Danish register data, we are the first to examine the causal effect of obtaining a general upper secondary education on the probability of moving to an urban area. We find a substantial and significant effect of general upper secondary education on migration. Using a linear probability model, controlling for relevant characteristics, we find an average effect in the range of 22 – 26 percentage points. As the choice of upper secondary education might be endogenous, we also conduct an IV analysis. We use distance to general upper secondary school as instrumental variable and find very large local average treatment effects (LATE) at 66 percentage points. These effects are especially large for individuals with unskilled parents and are negligible among individual with well-educated parents.

The reason for the large complier effects is rooted in two reasons. One is that compliers – individuals that choose a general upper secondary education because of proximity – have a very low probability of moving without a general upper secondary education. A second reason is that compliers see very high optional value of having a degree. These individuals appear to have their optimal choice set changed when obtaining a general upper secondary education. Obtaining a general upper secondary education opens access to tertiary educational opportunities, and these opportunities are mostly located in urban areas.

Our study contributes to the literature on the determinants of migration. We add new knowledge about the role of general upper secondary education on an initial migration decision. Previous studies of human capital and migration have mainly been concerned with highly educated individuals who have completed a tertiary education. However, our analysis suggests that a large part of the sorting of highly educated individuals into urban areas occurs at a very early stage of the life cycle – when young adults move away from home to attend further education. Thus, studies examining rural to urban migration based on movements of highly educated workers could potentially underestimate the extent of sorting from rural to urban areas.

References

Ahlin, L., Andersson, M., & Thulin, P. (2018). Human capital sorting: The "when" and "who" of the sorting of educated workers to urban regions. *Journal of Regional Science*, *58*(3), 581–610. doi: 10.1111/jors.12366

Andersen, H. T., & Nørgaard, H. (2018). Bosætning og flyttemønstre i Danmark—En oversigt. *Økonomi & Politik*, *91*(4), 7–19.

Austin, B. A., Glaeser, E. L., & Summers, L. H. (2018). Jobs for the Heartland: Place-Based Policies in 21st Century America. *NBER Working Paper Series*, *Working paper 24548*.

Bartel, A. P. (1979). The Migration Decision: What Role Does Job Mobility Play? *The American Economic Review*, *69*(5), 775–786.

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.

Berry, C. R., & Glaeser, E. L. (2005). The divergence of human capital levels across cities*. *Papers in Regional Science*, *84*(3), 407–444.

Busch, O., & Weigert, B. (2010). Where have all the graduates gone? Internal cross-state migration of graduates in Germany 1984-2004. *Annal of Regional Science*, *44*, 559–572.

Cameron, S. V., & Taber, C. (2004). Estimation of Educational Borrowing Constraints Using Returns to Schooling. *Journal of Political Economy*, *112*(1), 132–182.

Card, D. (1995). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. *Aspects of Labour Market Behavior: Essays in Honour of John Vanderkamp*, 201–222.

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

Cooke, T. J., & Boyle, P. (2011). The Migration of High School Graduates to College. *Educational Evaluation and Policy Analysis*, *33*(2), 202–213.

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.

Danmarks Statistik. (2020). *Statistikdokumentation for Byopgørelsen 2020*.

Dickerson, A., & McIntosh, S. (2013). The Impact of Distance to Nearest Education Institution on the Post-compulsory Education Participation Decision. *Urban Studies*, *50*(4), 742–758.

Dixon, S. (2003). Migration within Britain for job reasons. *Labour Market Trends; London*, *111*(4), 191.

Eurostat, European Union. (2016). *Urban Europe—Statistics on cities, towns and suburbs*. Luxembourg.

Faggian, A., Corcocan, J., & Partridge, M. (2015). 22 Interregional migration analysis. In C. Karlsson, M. Andersson, & T. Norman (Eds.), *Handbook of Research Methods and Applications in Economic Geography*. Cheltenham, UK: Edward Elgar Publishing Limited.

Faggian, A., & Franklin, R. S. (2014). Human Capital Redistribution in the USA: The Migration of the College-bound. *Spatial Economic Analysis*, *9*(4), 376–395.

Faggian, A., & McCann, P. (2009a). Human capital and regional development. In *Handbook of Regional Growth Development Theories*. Edward Elgar.

Faggian, A., & McCann, P. (2009b). Human capital, graduate migration and innovation in British regions. *Cambridge Journal of Economics*, *33*(2), 317–333.

Faggian, A., McCann, P., & Sheppard, S. (2006). An analysis of ethnic differences in UK graduate migration behaviour. *The Annals of Regional Science*, *40*(2), 461–471.

Faggian, A., McCann, P., & Sheppard, S. (2007). Some Evidence That Women Are More Mobile Than Men: Gender Differences in U.k. Graduate Migration Behavior. *Journal of Regional Science*, *47*(3), 517–539.

Faggian, A., Rajbhandari, I., & Dotzel, K. R. (2017). The interregional migration of human capital and its regional consequences: A review. *Regional Studies*, *51*(1), 128–143. doi: 10.1080/00343404.2016.1263388

Falch, T., Lujala, P., & Strøm, B. (2013). Geographical constraints and educational attainment. *Regional Science and Urban Economics*, *43*(1), 164–176.

Florida, R. (2003). Cities and the Creative Class. *City & Community*, *2*(1), 3–19.

Gennaioli, N., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2013a). Human Capital and Regional Development. *The Quarterly Journal of Economics*, *128*(1), 105–164.

Gennaioli, N., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2013b). Human Capital and Regional Development. *The Quarterly Journal of Economics*, *128*(1), 105–164.

Glaeser, E. (2013). A Review of Enrico Moretti's The New Geography of Jobs. *Journal of Economic Literature*, *51*(3), 825–837.

Glaeser, E. L. (1999). Learning in Cities. *Journal of Urban Economics*, *46*(2), 254–277.

Gulløv, E., & Gulløv, J. (2020). *Opvækst i provinsen. Om dem, der bliver, og dem, der rejser.* Aarhus: Aarhus Universitetsforlag.

Haapanen, M., & Tervo, H. (2012). Migration of the Highly Educated: Evidence from Residence Spells of University Graduates. *Journal of Regional Science*, *52*(4), 587–605.

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

Kodrzycki, Y. K. (2001). Migration of recent college graduates: Evidence from the National Longitudinal Survey of Youth. *New England Economic Review*, 13

Kotavaara, N., Kotavaara, O., Rusanen, J., & Muilu, T. (2018). University graduate migration in Finland. *Geoforum*, *96*, 97–107.

Levy, M. B., & Wadycki, W. J. (1974). Education and the Decision to Migrate: An Econometric Analysis of Migration in Venezuela. *Econometrica*, *42*(2), 377–388.

Lucas, R. E. (1988). On the Mechanics of Economic Development. *Journal of Monetary Economics*, *22*(1), 3–42.

Lucas, S. R. (2001). Effectively maintained inequality: Education transitions, track mobility, and social background effects. *American Journal of Sociology*, *106*(6), 1642–1690.

Malamud, O., & Wozniak, A. (2012). The Impact of College on Migration Evidence from the Vietnam Generation. *Journal of Human Resources*, *47*(4), 913–950.

Moretti, E. (2004a). Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, *121*, 175–212.

Moretti, E. (2004b). Human Capital Externalities in Cities. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics. Volume 4. Cities and Geography* (First edition). Amsterdam: Elsevier B.V.

Moretti, E. (2004c). Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions. *The American Economic Review*, *94*(3), 656–690.

Parker, K., Horowitz, J., Brown, A., Fry, R., Cohn, D., & Igielnik, R. (2018). *What Unites and Divides Urban, Suburban, and Rural Communities*. Washington DC: Pew Research Center. Retrieved from Pew Research Center website: https://www.pewresearch.org/social-trends/2018/05/22/what-unites-and-divides-urban-suburban-and-rural-communities/

Plane, D. A. (1993). Demographic Influences on Migration. *Regional Studies*, *27*(4), 375–383.

Rauch, J. E. (1993). Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities. *Journal of Urban Economics*, *34*(3), 380–400.

Rosenthal, S. S., & Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, *64*(2), 373–389.

Schwartz, A. (1976). Migration, Age, and Education. *Journal of Political Economy*, *84*(4), 701–719.

Sørensen, E. S. (2021a). *Human capital, migration and regional inequality: How the location of educational institutions influences our society* (PhD thesis). University of Copenhagen, Copenhagen.

Sørensen, E. S. (2021b, April). *Youth Participation in Education: The Impact of Distance to School and Parental Education*. Presented at the PhD Thesis, Copenhagen.

Sørensen, E. S., & Holm, A. (2019). *Unge flytter væk fra Danmarks mindste byer—Har uddannelse betydning?* København: Kraks Fond Byforskning.

The Ministry of Education. (2020, March 31). Grades in 9th grade. Conspectus of national average. Retrieved March 31, 2020, from Grades in 9th grade. Conspectus of national average. website: https://uddannelsesstatistik.dk/Pages/Reports/1809.aspx

Venhorst, V. A., & Cörvers, F. (2018). Entry into working life: Internal migration and the job match quality of higher-educated graduates. *Journal of Regional Science*, *58*(1), 116–140.

Venhorst, V., Dijk, J. V., & Wissen, L. V. (2010). Do the Best Graduates Leave the Peripheral Areas of the Netherlands? *Tijdschrift Voor Economische En Sociale Geografie*, *101*(5), 521–537.

Venhorst, V., Dijk, J. V., & Wissen, L. V. (2011). An Analysis of Trends in Spatial Mobility of Dutch Graduates. *Spatial Economic Analysis*, *6*(1), 57–82.

Appendix tables

Table A1
Definitions of variables.

Variables	Variable names	Description
Dependent variable	Move	Dummy variable equal to 1 if moved to an urban area, before age 25
Individual characteristics		
Youth education	General upper secondary education	Dummy variable, equal to 1 if completed general upper secondary education before age 25
Gender	Female	Dummy variable, equal to 1 if female
Immigration status	Danish	Reference category
	First-generation immigrant	Dummy variable, equal to 1 if first-generation immigrant
	Second-generation immigrant	Dummy variable, equal to 1 if second-generation immigrant
Year of birth	1986	Reference category
	1987	Dummy variable, equal to 1 if born in 1987
	1988	Dummy variable, equal to 1 if born in 1988
	1989	Dummy variable, equal to 1 if born in 1989
	1990	Dummy variable, equal to 1 if born in 1990
Grades from the final examination in secondary school	Written Danish	Exam grade in written Danish in 9th grade
	Missing grade in Danish	Dummy variable, equal to 1 if missing exam grade
	Written Maths	Exam grade in written Maths in 9th grade
	Missing grade in Maths	Dummy variable, equal to 1 if missing exam grade
Geographical characteristics		
	Municipality	Residential municipality at age 15
	Distance to urban area	Distance between residential address and nearest urban area (km)
	Distance to school	Distance between residential address and nearest general upper secondary education (km)

(continued on next page)

Table A1 (continued)

Variables	Variable names	Description
Family characteristics		
Mother's education	Missing education information	Dummy variable, equal to 1 if education information is missing
	Elementary school	Reference category
	General upper secondary education	Dummy variable, equal to 1 if highest education is general upper secondary educaiton
	Vocational upper secondary education	Dummy variable, equal to 1 if highest education is vocational upper secondary education
	Short-cycle higher education	Dummy variable, equal to 1 if short-cycle higher education
	Medium-cycle higher education	Dummy variable, equal to 1 if medium-cycle higher education
	Long-cycle higher education	Dummy variable, equal to 1 if long-cycle higher education
Father's education	Missing education information	Dummy variable, equal to 1 if education information is missing
	Elementary school	Reference category
	General upper secondary education	Dummy variable, equal to 1 if highest education is general upper secondary educaiton
	Vocational upper secondary education	Dummy variable, equal to 1 if highest education is vocational upper secondary education
	Short-cycle higher education	Dummy variable, equal to 1 if short-cycle higher education
	Medium-cycle higher education	Dummy variable, equal to 1 if medium-cycle higher education
	Long-cycle higher education	Dummy variable, equal to 1 if long-cycle higher education
Household income	Household income (log)	Log of yearly household income (in DKR).
	Poor	Dummy variable, equal to 1 if in the poorest income percentile
	Rich	Dummy variable, equal to 1 if in the richest income percentile

Note: All attributes are measured when individuals were 15 years old, unless otherwise indicated.

Table A2
Distance to nearest general upper secondary education in 5-kilometre intervals.

Distances	Freq.	Pct.	Cum. Pct.
0 - 5 km	34,584	23.63	23.63
5 - 10 km	38,715	26.45	50.08
10 - 15 km	36,799	25.14	75.22
15 - 20 km	23,097	15.78	91.00
20 - 25 km	9,192	6.28	97.28
25 - 30 km	2,568	1.75	99.03
30 - 35 km	491	0.34	99.37
35 - 40 km	529	0.36	99.73
40 - 45 km	389	0.27	99.99
45 - 50 km	11	0.01	100.00

Note: Distances are in km. 146,375 observations

Table A3
Comparison of OLS and LOGIT estimation results.

	LPM	LOGIT (AME)
General upper secondary education	0.2641***	0.2239***
	(0.0027)	(0.0024)
Individual controls	Yes	Yes
Family controls	Yes	Yes
Municipality FE	Yes	Yes
IV	No	No
Number of observations	146,375	146,375
R2	0.2341	0.2137
F-test (IV)		

Note: *** p<0.001, ** p<0.01, * p<0.05. Standard errors in parentheses. Table estimates report the average marginal effects of completing general upper secondary education on moving to an urban area before the age of 25. All other covariates are held constant at their means.

In table A4 we present results for our first-stage results. In the first column, we present first-stage results without any additional covariates, in the second column, we present a model with a limited set of covariates (individual characteristics) and in the third column, we present a model with a full set of covariates (parental characteristics). We add covariates gradually to see if adding covariates changes the effect of the instrument. This would indicate whether the instrument is correlated with observables and hence if it is likely that it is also correlated with unobservables.

From Table A4, we first see that our instrument has a strong significant effect on our endogenous variable, the decision to complete the academic track of general upper secondary education. Second, we see that the effect of the instrument changes in models with additional covariates. Both adding individual and parental covariates, as well as municipality fixed effects, change the effect of the instrumental variable in the first-stage regression. This indicates that the instrument is correlated with observed parental characteristics, which again makes it somewhat harder to maintain the important assumption that it is uncorrelated with unobserved parental characteristics, i.e. that the exclusion restriction is viable.

Table A4
Results of 1. stage regressions.

	First stage - covarians	First stage - reduced model	First stage - full model
Variable	(1)	(2)	(3)
Distance to nearest general upper secondary education	-0.0103***	-0.0074***	-0.0051***
	(0.0005)	(0.0004)	(0.0004)
Distance to nearest general upper secondary education, squared	0.0002***	0.0002***	0.0001***
	(0.0000)	(0.0000)	(0.0000)
Intercept	0.6264***	0.0606***	-1.1550***
	(0.0032)	(0.0046)	(0.0490)
Individual covariates	No	Yes	Yes
Parental covariates	No	Yes	Yes
Municipality FE	No	Yes	Yes
Number of observations	146,375	146,375	146,375
R-squared	0.0066	0.3855	0.4136
F-test	483.71	303.62	110.73

Notes: *** p<0.001, ** p<0.01, * p<0.05. Standard errors in parentheses. Table estimates report the association between distance to the nearest general upper secondary education and participation. Distances are in km. Control variables are gender, immigrant status, year dummies, GPA in Maths and Danish in secondary school, distance to urban area, father's highest education, mother's highest education, household income and municipality dummies.

In Table A5, we show the first-stage results by SES groups. We find strong and significant first-stage results for the least educated group, SES 1. Further, we find that the effect of distance declines across SES groups, such that the first stage is strongest for SES group one and two and completely absent in SES group four. Hence, we see that the lower the educational background of the parents, the stronger the first stage effect.

Table A5
First-stage regressions split by family background.

	SES 1	SES 2	SES 3	SES 4
	(1)	(2)	(3)	(4)
Distance to nearest general upper secondary education	-0.0063***	-0.0057***	-0.0043***	-0.0017
	(0.0010)	(0.0006)	(0.0008)	(0.0014)
Distance to nearest general upper secondary education, squared	0.0002***	0.0001***	0.0001***	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0001)
Intercept	-0.7071***	-1.3444***	-1.1699***	-0.3994***
	(0.1132)	(0.0760)	(0.0948)	(0.1309)
Individual controls	Yes	Yes	Yes	Yes
Family control	Yes	Yes	Yes	Yes
Municipality FE	Yes	Yes	Yes	Yes
Number of observations	21,084	71,141	43,302	10,848
R-squared	0.3650	0.3704	0.3296	0.2430

Notes: *** p<0.001, ** p<0.01, * p<0.05. Standard errors in parentheses. Table estimates report the association between distance to the nearest general upper secondary education and participation. Distances are in km. Control variables are gender, immigrant status, year dummies, GPA in Maths and Danish in secondary school, distance to urban area, father's highest education, mother's highest education, household income and municipality dummies.

Appendix figures

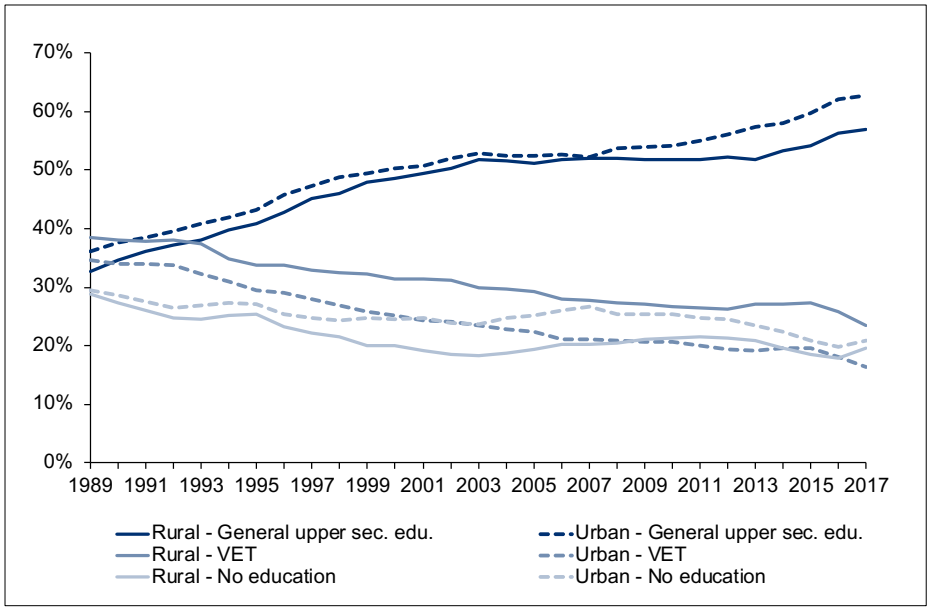

Figure A1. Educational attainment among youths in rural and urban areas, 1989 – 2017
Note: The figure shows the first completed youth education at age 24. Calculations are based on a full population of 24-year-olds each year. In cases where a person completes two different youth educations before turning 24, the first education is used. People with missing educational information's are not shown on the figure. Therefore, the numbers do not sum to 100%. The population is divided into a rural and an urban population on the basis of a person's residential address, when he or she was 17 years old. Persons without a residential address in Denmark when they were 17 and 24 are omitted.
Source: Own calculations based on administrative data from Statistics Denmark (KOTRE, 2019).

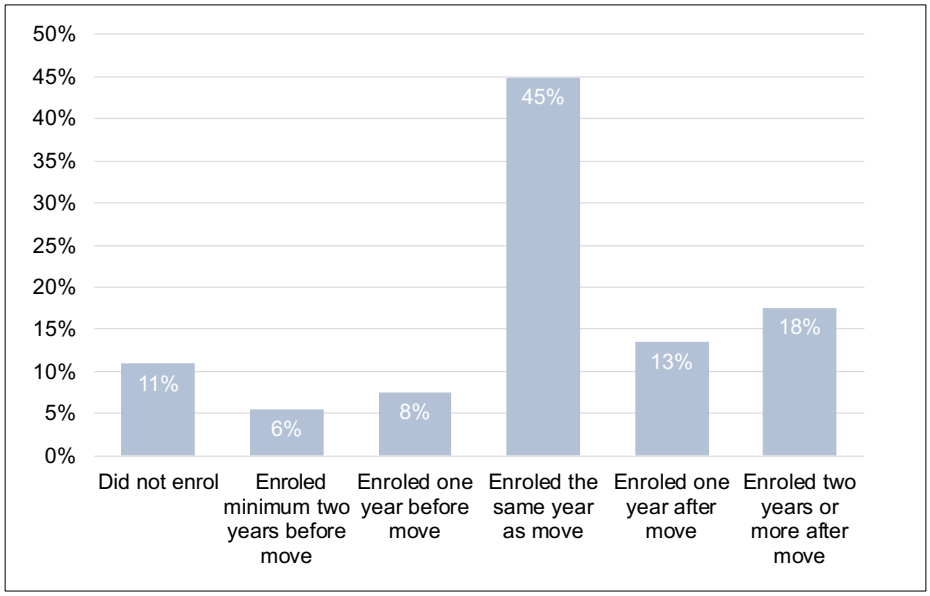

Figure A2. Timing of move and enrolment in higher education. Sub-sample of general upper secondary graduates who moves (N = 70,165).

Paper 3

PAPER 3

Gaps in Higher Education Choice and Participation: Understanding the Role of Distance to Higher Education

Elise Stenholt Sørensen

Abstract

This paper examines to what extent the proximity of an individual's home to higher education institutions influences their participation in and choice of higher education study programme. Geographical distance between home and education institution poses a potential barrier to enrolment because of commuting or re-location costs. In addition, distance could be a deciding factor when choosing between college and university. This paper uses administrative data on the population of all young individuals who completed a general upper secondary education in Denmark from 2006 – 2011. The results suggest that the impact of proximity to higher education institutions varies across family background; while it has a small impact on participation in higher education among students from less-educated families, it has no significant impact if one or both parents have a higher education. However, the distance from an individual's home to the nearest university has a significant (but small) influence on their choice between college and university, especially among students from lower-educated families. Thus, the geographical location of colleges and universities has implications for the sorting of students across institutions, and hence the regional formation of human capital.

Keywords:

Higher education, geographical accessibility, distance to school, choice of higher education

JEL: C25, I21, R10

1. Introduction

This paper investigates a potential contributory cause of the increasing regional inequality in human capital and economic prosperity that is witnessed throughout Europe (Berry & Glaeser, 2005; Eurostat, European Union, 2016; Glaeser, 2013; K. Parker et al., 2018). The investigation takes its case from the regional inequality that is found in Denmark (see Chapter 4 in the present PhD thesis), and addresses a central question in the debate about how to narrow the gap between peripheral and urban regions, namely, to what extent does the geographical location of a country's higher education (HE) institutions contribute to the increasing regional inequality in human capital and economic prosperity? In Denmark, a free and universal education system, including HE, is a cornerstone of the common narrative on the strong welfare state, where everybody is offered the same public services and opportunities. If geography poses a barrier to access to HE, then one could argue that Denmark does not live up to the expectations created by the narrative on the strong welfare state.

Distance to HE institutions may act as a barrier to participation because of commuting or re-location costs (Gibbons & Vignoles, 2012). These may include direct financial costs if the student has to leave home and/or has a long commute to an education institution. In addition, long distance can be associated with emotional costs, because individuals are rooted in their local environment and may be unwilling to leave their social network. The barriers might be higher for young individuals from lower-income homes, and for individuals from lower-educated families where the cultural incentives to pursue higher education are lower. If students from different socio-economic backgrounds differ in their sensitivity to distance, the geographical accessibility of HE could have an impact on the differences in educational attainment for different socio-economic groups, and also on differences in future earnings (Gibbons & Vignoles, 2012; White & Lee, 2020).

According to the incumbent Danish Minister of the Interior and Housing, Kaare Dybvad Bek, it is a problem if the place where you are born shapes your opportunities and future earnings (Broberg, 2021). Thus, the question about the extent to which geographic distance from home to HE institution causes some young individuals to refrain from pursuing HE is central in the present political debate on regional inequality in Denmark. During the past ten years, shifting governments across the political spectrum in Denmark have had a political strategy on decentralising HE in Denmark to "create a better balance and local

growth" in all regions (Kildegaard & Domino, 2019; Ministry of Higher Education and Science, 2020a; Vestergaard, 2015).

However, there is no large-scale, empirical evidence for Denmark that shows that proximity to higher education matters for participation in or choice amongst different types of higher education programmes. The present paper examines patterns in participation in or choice amongst different types of higher education programmes by using rich administrative data on the population of all young individuals who completed a general upper secondary education in Denmark from 2006 to 2011. These data allow me to link education choices of young individuals from different socio-economic backgrounds to a unique dataset on distance information on their home-to-HE-institution distances, whilst accounting for gender, ethnicity, schooling, residential region and other background characteristics.

The present paper provides new insight into the effect of distance from home to HE institution on participation in HE and choice of HE programme in a Danish context. The case of Denmark provides an ideal laboratory to study distance effects, as the (potentially unobserved) costs of attaining a HE degree are ruled out[1]; participation in HE is free and students receive a state-funded education grant. Thus, the costs of education, beyond distance, are fixed[2]. In addition, the geographical dispersion of HE institutions is high, and distances from home to HE institutions are relatively short. Hence, if there are any signs of distance from home to HE institution playing a role for HE participation and choice in Denmark, one would expect to find the effect in other countries as well.

In addition, the present paper supplements previous work by specifying rich models that control for individual's human capital prior to the decision to participate in HE and choice of HE programme. The majority of previous studies lack sufficient individual controls to take residential sorting into account e.g. White & Lee (2020). Methodologically this paper improves on previous work by using more accurate distance measures: distance is measured at the individual level by measuring distance from an individual's residence to all HE

[1] 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 fully subsidised by the State of Denmark.

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

institutions through the road network at the exact time of the individual's graduation from upper secondary education. I use a full sample of individuals who graduated from an upper secondary education, which is the admission requirement to enrolment in HE in Denmark. The bulk of the studies available only consider students who are enrolled in either college[3] or university. Thus, they study a very selective group and do not take the selection of students into HE into account.

The paper is organised as follows. Section 2 reviews the existing empirical literature on the relationship between distance from home to HE institution and participation in HE, and home proximity to HE and choice amongst different types of HE programmes. Section 3 presents the research design and the statistical models. Section 4 describes the institutional background, data and descriptive statistics of distance from home to HE institution and participation in HE in Denmark. Section 5 presents the regression results of the association between distance from home to HE institution and participation in HE and section 6 presents the results of the association between distance from home to HE institution and choice of HE programme. Section 8 summarises the results and concludes.

[3] In this study, in a Danish context, the term 'college' to covers both the Danish business academies and university colleges, which are HE institutions. For further explanation, see section 4.1.

2. Literature

The literature about the relationship between home proximity to HE and participation in HE address two main questions: 1) does the geographical proximity of an individual's home to HE influence their participation in HE?; and 2) does home proximity to different HE institutions have an impact on an individual's choice of HE programme? Distance from home to HE schools has also been used as an instrumental variable for education choice in several studies that estimate causal effect of education on individuals' outcomes such as wages (e.g., Card, 1993; Currie & Moretti, 2003; Kane & Rouse, 1995). The following section provides a short literature review, with emphasis on recent European studies. I focus on the European studies because Denmark's geographical size, or its relatively high population density and education system have more in common with countries such as the Netherlands and Germany than with the US and Canada.

2.1 Evidence of Distance and Enrolment in Higher Education

A large literature has addressed the question of whether distance from home to HE institution deters some individuals from pursuing HE. The literature comprises studies in a variety of countries and education systems, e.g. Sweden, Germany, the Netherlands, the United Kingdom, Ireland, Turkey, Australia, the US and Canada (Cullinan et al., 2013; Frenette, 2006; Gibbons & Vignoles, 2012; Griffith & Rothstein, 2009; Öckert, 2012; Parker et al., 2016; Polat, 2017; Sá, Florax & Rietveld, 2006; Spiess & Wrohlich, 2010; White & Lee, 2020). The question addressed by these studies is important for several reasons. First, it would be a "democratic" problem if some individuals refrain from enrolling in HE because they were born in a place with poor accessibility to HE institutions. Second, as human capital is central to economic growth, innovation and development (Moretti, 2004; Romer, 1994), it can be costly for society if the full potential the talent of young individuals is not harnessed. While there is an abundance of literature in this field, only few studies use a research design that enables a causal analysis and interpretation of the results. I return to this issue at the end of this section.

In the following, I start by briefly describing the main findings from the bulk of the literature that examines the association between distance from home to HE institutions and participation in HE. I then go on to describe studies that have applied more advanced research designs to address the potential

endogeneity of distance from home to HE institutions and participation in HE.

Overall, the existing evidence of suggests that distance has little or no impact on the decision to participate in HE. The majority of the studies find a small negative effect (e.g. Frenette, 2006; Öckert, 2012; Spiess & Wrohlich, 2010), whereas others find no significant effect (Cullinan et al., 2013; Gibbons & Vignoles, 2012; Sá et al., 2006). Among the studies that do find a small effect, findings show that the impact of distance is larger in magnitude for, or only applies to, individuals from low socio-economic status households (Denzler & Wolter C., 2011; Frenette, 2006; P. D. Parker et al., 2016).

Some studies have found distance from home to HE institution plays a role in relation to participation in HE. Based on Swedish data, Öckert (2012) finds evidence for a deterrent effect of distance on college enrolment among a full population sample of Swedish males (cohorts born in 1974–1976). Adolescents who live more than 25 km from college have an average 4 percentage points lower probability of enrolling in college than those who live nearer a college[4]. Similarly, Spiess & Wrohlich (2010) find a reduced university participation above a 12.5-km threshold, employing survey data encompassing 1,219 German upper secondary school leavers. Based on a Canadian household survey, Frenette's analyses from 2004 and 2006 demonstrate that students who live beyond commuting distance (80 km) are less likely to attend university than students living within commuting distance. Further, he finds that students from lower income families are particularly disadvantaged by distance (Frenette, 2004, 2006).

However, there is also empirical evidence suggesting that there is no deterrent effect of distance. Sá et al. (2006) employ survey data on 3,263 Dutch students who completed their final year of upper secondary school in 1998–2000. They find no deterrent effect on the general participation in HE across the geographic access to community colleges and universities[5]. Similarly, based

[4] 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 HE.

[5] 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 upper secondary school profile (Science and technology, Health, Culture and society, Economics and society). Hence, distance varies across space and individuals.

on survey data on 858 Irish upper secondary school leavers in 2007, Cullinan et al. (2013) find no effect of proximity to the nearest HE institution, encompassing both community colleges and universities. However, they do find reduced participation among students without university-educated parents. Gibbons & Vignoles (2012), employing a full population sample of British adolescents at the age of 16, find that distance from home to the nearest university had little or no deterrent effect on the probability of continuing to university.

In sum, previous work finds that distance from home to HE institution has a (small) negative influence or no influence on participation in HE, depending on the country. Studies based on data from European countries suggest a more muted impact of distance than those from Canada and the US. However, the above studies provide no evidence on how distance from home to HE institutions may affect the likelihood of whether various socio-economic groups attend different types of educations, which could still be linked to geographical inequality. I examine studies on this question in the following.

2.2 Evidence of Distance and Choice of Study

The literature provides extensive evidence that distance from home to university influences students' choice of education type and institution – a so-called diversion effect (Denzler & Wolter C., 2011; Frenette, 2009; Steve Gibbons & Vignoles, 2009; Griffith & Rothstein, 2009; Sá et al., 2006). Frenette (2009) estimated the effect on HE participation in a given geographical area when a community college or university is established in that area. He concluded that the opening of a new university increased the share of young people from the university's geographically local catchment area who enroled in a university by 6.4 percentage points. However, this occurred at the expense of existing local community colleges, as there was a five-percentage-point reduction in the probability of potential students enrolling at a community college. Therefore, the expansion of HE to a large extent led to diversion effects.

Gibbons and Vignoles (2012) have examined almost the entire population of upper secondary school leavers in England in 2002 and modelled the association between distance from home to HE institutions and students' choice of HE institution. They conclude that "home-institution distance is very strongly linked to institution choice" (Gibbons & Vignoles, 2012). Hence, that the locations of higher-quality universities in England "partly determine the skill

composition of the local human capital stock in cities, labour markets and regions".

A handful of studies have exploited college openings and expansions as a source of exogenous variation in distance to overcome the possible problem of endogeneity due to household sorting (Frenette, 2009; Holzer, 2007; Oppedisano, 2011). These studies examine the effect of a newly opened college on HE enrolments within a given radius from the college. Frenette (2009) compares students who lived within a radius of 80 km from a college with students who lived further away. In the studies by Holzer (2007) and Oppedisano (2011), students were assumed to be affected by the opening of a college if the college opened within their home municipality or region. The studies find that an expansion of the local supply of HE institutions increases participation in HE.

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

In sum, a large part of the variation in the presence and size of the effects of distance from home to HE institution may originate from country-specific differences. These differences encompass within-country variation in population density and location of HE 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 the studies' 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 HE in Denmark.

3. Empirical strategy and methods

The overall purpose of the present study is to measure the sensitivity of young people's education decisions to the distance between their home and HE institutions. More specifically, I consider two related questions: a) the association between the proximity of HE institutions to a young person's home and their decision to enrol in a HE programme; and, b) the association between distance from home to different types of HE and young person's choice of education type.

3.1 Research Design

The nature of the research question imposes constraints on the research design (Gibbons & Vignoles, 2012). The geographical location of HE institutions is mostly fixed, and the decision on whether and where to enrol in HE is for most young people a one-off decision[6]. This limits this study to a cross-sectional analysis, despite the fact that I use longitudinal data sources. Thus, the research design of this study compares the decisions of individuals who face different home-to-education distance patterns according to where they lived when they graduated from their general upper secondary education. Using this variation in distance (thus implicitly residential location) to infer education choice might be problematic because choice of residential location is to a large extent determined by household and individual factors, which may also determine educational decisions. In other words, residential location and distance to school are not completely random.

I therefore need to rely on a control-variable-based regression strategy to adjust for observable factors that determine educational decisions, and which may also influence distance from home to education institutions. This method was inspired by methods in a similar study by Gibbons and Vignoles (2012). This estimation strategy relies on the assumption that I have information about all relevant control variables that are correlated with both distance from home to HE institutions and HE participation, which might be a problematic assumption. Hence, the results of the regression models cannot necessarily be

[6] I have mapped and documented the openings and closures of all HEs in Denmark during the time period 2006 – 2014. The aim was to use the creation of new institutions as a source of exogenous variation in distance to school. However, the changes in distances due to creations or closures of institutions were too small and insignificant to be used as a source of variation. See Chapter 4 in the PhD Introduction for an overview and description of the historical openings and closures of tertiary institutions in Denmark.

interpreted as the causal effect of distance. Fortunately, rich administrative data sources provided by Statistics Denmark allow me to take a broad range of central factors, such as family background, income, geographical region and school test scores, into account. I return to a more detailed description of the control variables in section 4.5.

3.2 Participation Models

The main results presented are estimated probabilities of enrolment in HE (at any institution) with respect to distances from the individual's home to the three nearest HE institutions. The probabilities are estimated using a binary logit model on individual data. The probability of enrolment in HE (y=1) by individual *i* is expressed in terms of distance from the home of individual *i* to the three nearest tertiary institutions, given by distances D_{1i}, D_{2i} and D_{3i}, vectors of observables individual characteristics $\mathbf{x}_i$ and $\boldsymbol{z}_i$, and parameters $\alpha, \delta, \beta, \gamma$, to be estimated:

The following logit model is estimated:

$$P(y_i = 1) = \frac{\exp(V_i)}{(1+\exp(V_i))}, \qquad \text{where } V_i = \alpha + \delta D_i + \mathbf{x}_i'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{\gamma} \qquad (1)$$

where $D_{1i} - D_{3i}$ are continuous variables that measure distances in kilometres to the three nearest HE institutions. The associated parameter δ indicates the change in the enrolment rate for each kilometre increase in the distance to the HE institutions. The vector $\mathbf{x}_i$ comprises the individual characteristics gender, age, immigration status, year of graduation, type of general upper secondary education, grade point average from upper secondary school (GPA), and geographical region. The vector $\boldsymbol{z}_i$ comprises socio-economic variables such as mother's and father's highest education, the natural logarithm of family household income and family type. The explanatory variables $(D_i, \mathbf{x}_i, \mathbf{z}_i)$ are measured in the final year of upper secondary education. The dependent variable y_i measures if the individual enrols in a HE within 2 years after the completion of general upper secondary education. By assumption, the error term e is independent of $\mathbf{x}_i$, and has the standard logistic distribution. The parameters are estimated by maximum likelihood methods.

3.3 Education Choice Models

I use a multinomial logit model to model the individual choice between three different alternatives: 1) no enrolment in HE, 2) college enrolment and 3) university enrolment. The interpretation depends on a referent alternative. I choose "1) no enrolment" as the referent alternative in the following analysis. Thus, the results should be interpreted as the probability of enrolment in college and university, respectively, relative to no enrolment at all.

The following multinomial logit model is estimated:

$$P(i \text{ chooses } j) = \exp(V_{ij}) / \sum_k \exp(V_{ij}) \ , \ \text{where } V_{ij} = d_{ij}\alpha_j + x_i'\beta_j + z_i'\gamma_j \qquad (2)$$

V_{ij} is interpreted as a utility term. The parameters are estimated by maximum likelihood methods.

4. Institutional Context and Data

4.1 Institutional background

In Denmark, HE institutions fall into three broad categories: "business academies", "university colleges", and "research-based universities". The business academies offer a broad range of short-cycle higher education programmes, corresponding to 1 ½ – 2 ½ years of full-time study, targeted at the industry and commerce. The university colleges offer a range of education programmes aimed at the public welfare and health sectors, for example, nurses and midwives, childcare workers and elementary school teachers, as well as a range of technical training programmes such as electrical and mechanical engineering. The programmes are classified as medium-cycle HE programmes and the study period is 3 ½ years.

The research-based universities comprise all academic study programmes at the university level (Bachelor and Master's programmes) and range from 3 to 8 years of full-time study. Typically, the prescribed period of study at these institutions is 5 – 6 years of full-time study. In this paper, I refer to business academies and university colleges under one broad term: "colleges" and I will refer to research-based universities as "universities". Finally, there are a few remaining HE programmes that do not fall into the three described categories. These educations comprise maritime education programmes, education programmes in the Danish armed forces and education programmes within the Royal Danish Academy of Fine Arts.

Almost all HE institutions are state funded, and all education is free in Denmark[7]. Thus, there is no need consider tuition fees in the statistical analysis, as they do not vary by institution. In addition, all students receive a monthly student grant from the government (about 740 €[8]). A formal requirement for admission to all HE programmes is a general upper secondary education qualification[9]. Students are free to apply to any HE programme. In some programmes, the number of applicants exceeds the number of study places and admission is restricted. In these programmes, admissions are allocated in two quotas: Quota 1 and Quota 2. Quota 1 admissions are allocated according to a

7 The few private HEs are omitted from the analysis.
8 Equivalent to DKK 5,486 in 2011, before income taxes.
9 However, some short- and medium-cycle educations may allow students to use selected vocational secondary education qualifications to meet the admission requirement.

student's GPA from upper secondary school. Quota 2 admissions are allocated according to criteria published by the education institution(s), and these vary between institutions. In most cases, the number of study places in quota 2 is very limited[10] (Ministry of Higher Education and Science, 2020b).

4.2 Data

The analysis draws on administrative registers maintained by Statistics Denmark. Individuals in the sample population comprise all young people (18 – 21 years-old) who graduated from a general upper secondary school between 2006 – 2011. Persons living on an island without a bridge to the mainland are excluded from the analysis (about three percent of the population). Furthermore, students who enrol in either military, police, maritime or creative education programmes are omitted from the sample because I lack distance measures to those institutions. Enrolments in these education programmes account for about five percent of the population.

4.3 Distance to Higher Education Institutions in Denmark

First, I consider the spatial distribution of HE institutions and the geographical accessibility of these institutions to the students' homes. Distance is calculated from each individual's residential address located in a one-kilometre square raster, through the road network[11]. Figure 1 shows how distance to the nearest HE institution varies across Denmark. Proximity to HE is higher in urban areas, as expected. The "darkest blue" areas, where a higher education institution is located within 5 kilometres, tend to delineate the main urban areas. However, the map also reveals that very few areas in Denmark have poor geographical access to HE institutions. Only in a few peripheral areas is the nearest HE institution over 30 km away (the lightest blue areas). About 10 percent of the graduates from general upper secondary programmes have over 30 kilometres to their nearest HE when they graduate from upper secondary school (see Figure A1 in the appendix for distance histogram).

[10] More information about the admission criteria is published at the Higher Education and Science's website: ufm.dk/en/

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

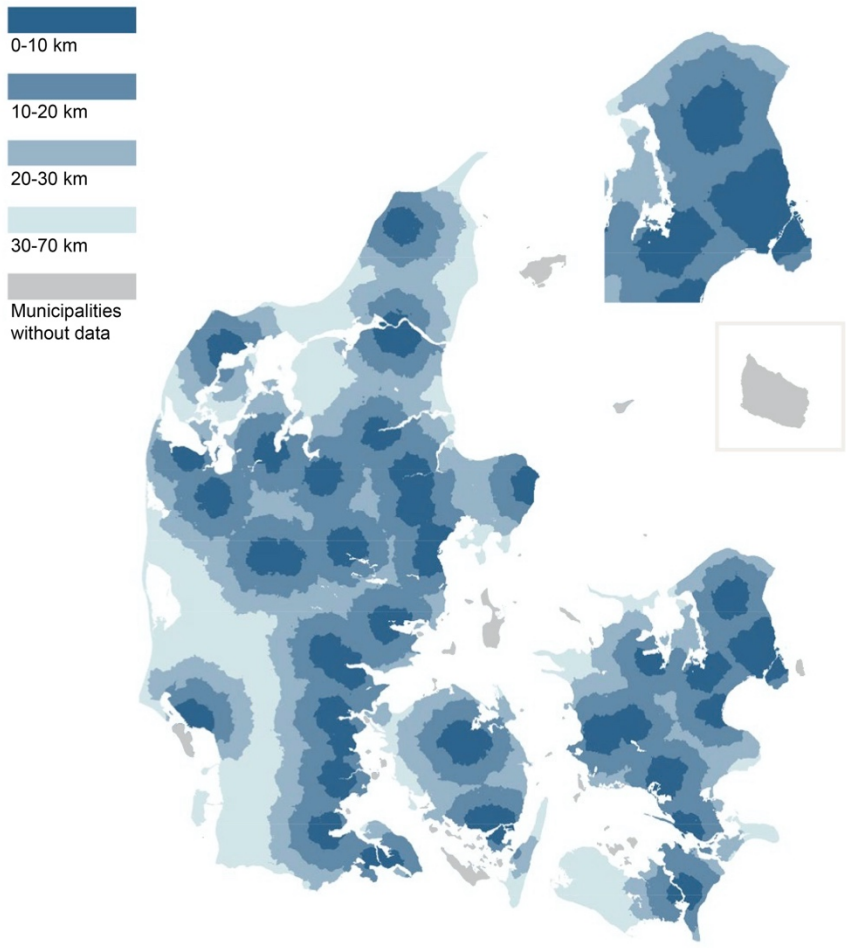

Figure 1. Distance from home to the nearest HE education in Denmark, 2011.
Note: Own calculations. Information about HE location is derived from the administrative Institution register (INST).

Figure 1 illustrates that the majority of the population had access to an HE institution within commuting distance; above 90% lived within 30 kilometres of their nearest HE institution, and 35% lived less than five kilometres away (see numbers in Figure A1 in the appendix). Overall, geographical dispersion of HE institutions is broad and very few areas have poor geographical access to HE. Only in a few peripheral areas (in the southwestern part of Jutland and

southwestern part of the island Lolland) is the nearest HE institution over 30 kilometres away. In sum, the distances from home to nearest HE institution are much shorter than in larger countries, for example Canada, England and Sweden (Frenette, 2006; Stephen Gibbons & Vignoles, 2012; Öckert, 2012).

4.4 The Geography of Higher Education Institutions and Participation

Figure 2 maps the descriptive association between geographical location of HE institutions and enrolment rates in Denmark (2006 – 2013). Colleges (marked with small red dots) are distributed widely across the country. This is in line with the Danish political strategy to ensure regional access to HE[12]. Universities (dark red pentagons) are primarily located in the largest cities. The map shows that enrolment rates are higher in municipalities with numerous HE institutions, i.e. urban areas such as the municipality of Odense (the third largest city) and the municipality of Aalborg (the fourth largest city)[13]. Thus, the map does, to some extent, suggest that there is an association between home proximity to education and the local participation rate. However, the enrolment rates in the capital city, Copenhagen, and the second largest city, Aarhus, are below the national average of 73%, which is inconsistent with the hypothesis of distance as a barrier to enrolment. Thus, it is crucial to take other central factors into account, such as parents' education and household income when analysing the importance of distance from home to HE institution for enrolment in HE.

[12] The term 'colleges' covers both business academies and university colleges.
[13] In the municipality of Odense, the enrolment rates were 76%, on average, and 79% in the municipality of Aalborg.

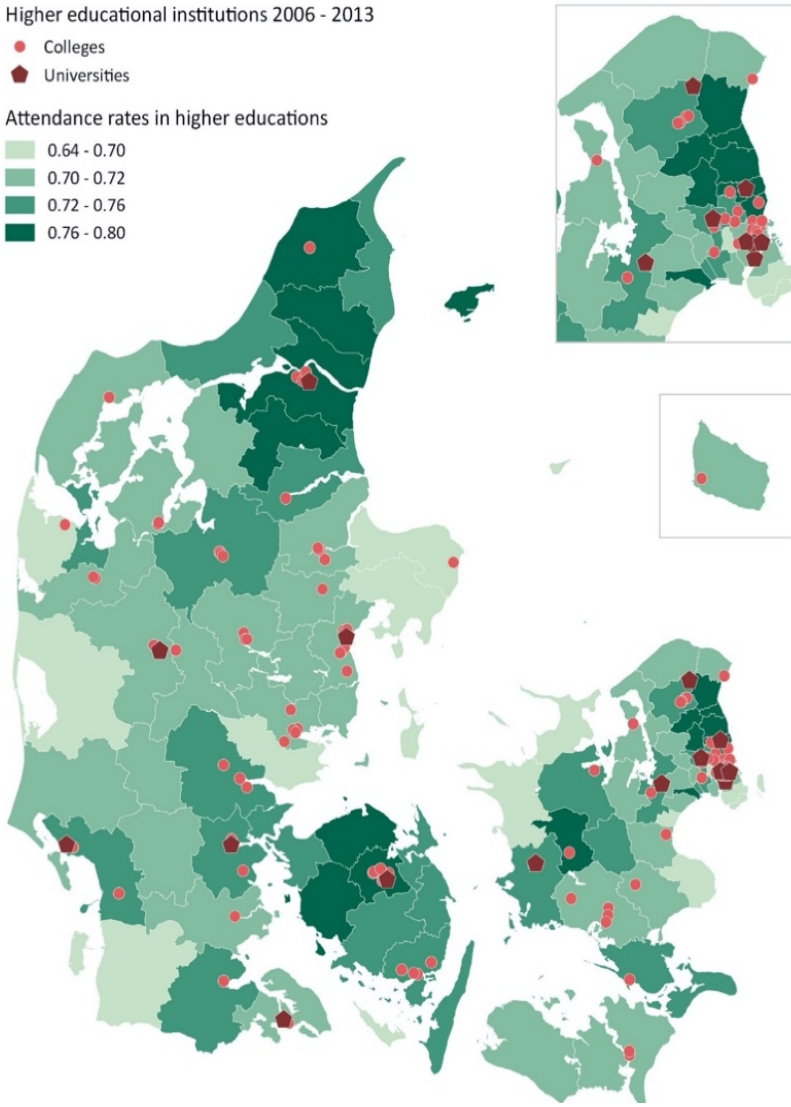

Figure 2. Association between location of HEs in Denmark and participation in HEs.
Note: For each municipality, Figure 2 presents the percentages of all upper secondary graduates who enrolled in a HE within two years after graduation from upper secondary school. The data comprise students who were between 18 and 21 years old when they graduated from general upper secondary education.

4.5 Individuals and family control variables

Table 1 presents summary statistics of the sample and variables used in the analyses below. On average, 73 percent enrolled in an HE programme within two years of graduation from upper secondary education. Of these, 28 percent enrolled in a college education programme and 45 percent enrolled in a university education programme. The average distance from home to the nearest HE institution was 11.28 km, and the maximum distance was 65 km. The majority of the sample (61%) graduated from a general 3-year upper secondary education programme ('STX'). About 93% of the sample were of native Danish origin and 57% were female. To account for selection with respect to academic abilities, I control for GPA in written math and written Danish in 9th grade; both variables have a mean of 7.2 and range from -3 to 12. In addition, I control for the student's socio-economic background in terms of mother's and father's highest education, family type and log of the household income.

Table 1
Summary statistics.

Variables	Variable outcomes	Mean	Std. Dev.	Min.	Max
Dependent variables					
HE participation (y1)	Enrolled in HE	0.73	0.44	0	1
HE choice (y2)	None (ref.)				
	College	0.28	0.45	0	1
	University	0.45	0.50	0	1
Distances					
Distance to the nearest HE	Distance to 1st nearest	11.28	11.54	0	65
	Distance to 2nd nearest	16.31	14.28	0	101
	Distance to 3rd nearest	22.05	17.58	0	110
Distance to colleges	Distance to 1st nearest college	11.59	11.59	0	66
	Distance to 2nd nearest college	18.54	14.98	0	101
	Distance to 3rd nearest college	24.85	19.40	0	110
Distance to universities	Distance to 1st nearest university	25.42	23.78	0	147
	Distance to 2nd nearest university	51.51	38.78	0	205
	Distance to 3rd nearest university	65.69	39.95	1	207
Individual characteristics					
Gender	Male	0.43	0.49	0	1
Age dummies	18 years	0.22	0.41	0	1
	19 year (ref.)				
	20 years	0.25	0.43	0	1
	21 years	0.03	0.18	0	1
Immigration status	Danish (ref.)				
	1. generation immigrant	0.03	0.18	0	1
	2. generation immigrant	0.04	0.20	0	1
Type of upper secondary education	General, 3-year (stx)	0.61	0.49	0	1
	General, 2-year (hf)	0.11	0.31	0	1
	Mercantile (hhx) (ref.)				
	Tecknical (htx)	0.08	0.27	0	1
	International (IB)	0.00	0.05	0	1
Year of graduation	2006 (ref.)				
	2007	0.15	0.36	0	1
	2008	0.16	0.36	0	1
	2009	0.17	0.38	0	1
	2010	0.18	0.38	0	1
	2011	0.19	0.40	0	1

(continued on next page)

Table 1 (continued)

Variables	Variable outcomes	Mean	Std. Dev.	Min.	Max
Grades in secondary school	Written Danish	7.20	2.32	-3.0	12.0
	Written Maths	7.20	2.46	-3.0	12.0
Geographical region					
	Northern Jutland (ref.)				
	Central Jutland	0.25	0.43	0	1
	Southern Denmark	0.23	0.42	0	1
	Capital Region	0.27	0.44	0	1
	Zealand Region	0.14	0.35	0	1
Family characteristics					
Family type	Nuclear family (ref.)				
	Divorced, with new partner	0.08	0.27	0	1
	Single parent	0.16	0.36	0	1
	Not living with parents	0.15	0.36	0	1
Family characteristics					
Mother's education	Missing education information	0.02	0.15	0	1
	Elementary school (ref.)				
	General upper secondary education	0.05	0.22	0	1
	Vocational upper secondary education	0.34	0.47	0	1
	Short-cycle higher education	0.05	0.21	0	1
	Medium-cycle higher education	0.31	0.46	0	1
	Long-cycle higher education	0.09	0.28	0	1
Father's education	Missing education information	0.06	0.24	0	1
	Elementary school (ref.)				
	General upper secondary education	0.05	0.22	0	1
	Vocational upper secondary education	0.38	0.49	0	1
	Short-cycle higher education	0.05	0.22	0	1
	Medium-cycle higher education	0.17	0.38	0	1
	Long-cycle higher education	0.13	0.34	0	1
	Household income (ln)	13.07	0.40	11.6	14.3

Note: N = 177,390 observations

5. Empirical results: Participation in higher education

The purpose of the parametric analysis is to examine a) whether distance from home to the three nearest HE institution causes students to refrain from enrolling in HE, and b) whether distance from home to the three nearest HE institution influences the choice between different types of HEs (colleges vs. universities), taking individual, geographic and socio-economic factors into account. In addition, the results are divided into four groups by parents' education in order to analyse whether there are differences between these groups in their sensitivity to distance. Section 5 presents the results regarding distance from home to the three nearest HE institution and HE participation, and section 6 presents the results regarding distance from home to the three nearest HE institution and the choice of college or university education.

5.1 Association between home proximity to HE institution and participation in HE

Figure 3 illustrates the association between distance from home to the nearest HE institution and participation in any HE[14]. The solid line plots the non-parametric association between distance and participation with 95 % prediction intervals. The small-dotted line plots the predictions from a linear regression model, and the dashed line plots the predictions from a quadratic polynomial distance specification. Figure 3 suggests a modest negative association between distance from home to the nearest HE institution and participation in HE, when individual and family characteristics are not taken into account. The highest average participation rate (74.5 percent) is seen among students who live 6 to 11 kilometres away from a HE institution. The average participation rate declines above 11 kilometres with a nearly linear slope until it reaches the lowest participation rate (71.5 percent) among (the few) individuals with 49 kilometres to a HE institution. However, the prediction intervals are very wide above 30 kilometres because of the relatively few observations, so the average participation is not statistically different from the average at 20 kilometres.

[14] About 0.5 % of the sample have above 50 kilometres to the nearest HE institution. They are left out of the figure. Figure A2 in the appendix shows the association for the total population.

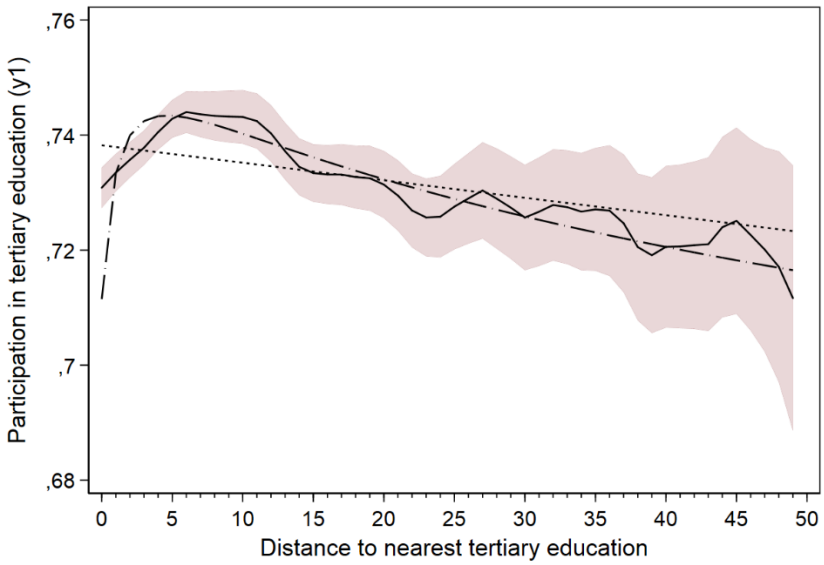

Figure 3. Association between distance from home to the nearest HE institution and participation in HE. Solid line: non-parametric association with 95% prediction intervals, dotted line: linear prediction, dashed line: quadratic polynomial prediction.

The simple association shown in Figure 3 provides no compelling evidence that geographic proximity affects the participation decision for young individuals. As yet, we have not controlled for individual or family socio-economic characteristics that may lead those individuals with high participation rates to live relatively near to HE institutions (e.g. in urban areas). In the next section I use regression analysis to control for individual, geographic and family characteristics that are likely to be correlated with HE participation.

5.2 Regression Models of Participation in Higher Education

Table 2 reports the average marginal effects (AME) of participation with respect to the three HE institutions that are nearest to an individual's home, derived from binary logit models (see description in section 3.2.). Table 2 shows the results of three model specifications. Column (1) present results from a model without controls. Column (2) presents the distance estimates, adjusted for individual and geographic covariates. Column (3) shows the distance estimates adjusted for individual, regional and family characteristics. The reported AME gives the percentage change in participation probability with respect to a one-kilometre increase in distance. Note that the models estimate the effect of each

distance variable, conditional on the others. Hence, for example, the AME with respect to distance to the nearest HE institution corresponds to a conceptual experiment which moves the nearest institution closer while keeping the other HE institutions where they are.

From the first model specification (Column (1)) we see that there is no strong association between distance and participation, even without controls. Distance to the first nearest and third nearest HE institution is insignificant, conditional on distance to the rest. Distance to the second nearest HE institution is statistically significant and has a negative sign. However, the effect size is very small in magnitude; when distance increases by 1 km, the probability of participating in HE decreases by 0.05 percentage points. The same pattern is seen in the two other model specifications where individual and socio-economic covariates are taken into account (Column (2) and (3)).

Column (3) presents the estimates from the third specification with all control variables and distances. It shows that distance to the nearest institution is negatively associated with participation in HE, distance to the second nearest institution is insignificant and distance to the third nearest institution is positively associated with participation. The estimates are still very small in magnitude. The estimates suggest that, holding all other distances and control variables fixed, a ten-kilometre increase in distance to the nearest HE intuition would decrease participation by 0.5 percentage points, on average. In my example, the estimate implies that increasing distance to the nearest HE institution by 10 kilometres would decrease the mean probability of participation in HE from 74.44% to 72.95%.

Table 2
Main results.

	(1)		(2)		(3)	
Covariates	*M.E.*	*(S.E.)*	*M.E.*	*(S.E.)*	*M.E.*	*(S.E.)*
Distance to 1st nearest	0.0001	(0.0001)	-0.0004***	(0.0001)	-0.0005***	(0.0001)
Distance to 2nd nearest	-0.0005**	(0.0002)	-0.0002	(0.0002)	-0.0001	(0.0002)
Distance to 3rd nearest	-0.0000	(0.0001)	0.0002	(0.0001)	0.0004***	(0.0001)
Male			-0.0724***	(0.0022)	-0.0847***	(0.0022)
Age						
(ref.= 19 years)						
18 years			0.0013	(0.0027)	-0.0028	(0.0027)
20 years			0.0009	(0.0025)	0.0063***	(0.0024)
21 years			0.0246***	(0.0053)	0.0430***	(0.0052)
Immigrant status						
(ref. = Danish)						
1st-generation immigrant			0.1271***	(0.0044)	0.1446***	(0.0042)
2nd-generation immigrant			0.1411***	(0.0038)	0.1622***	(0.0036)
Upper secondary edu.						
(ref. = hhx)						
General, 3-year (stx)			0.1351***	(0.0029)	0.1129***	(0.0029)
General, 2-year (hf)			0.0235***	(0.0042)	0.0256***	(0.0041)
Technical (htx)			0.1363***	(0.0043)	0.1309***	(0.0042)
International (IB)			-0.1155***	(0.0251)	-0.1425***	(0.0250)
Year						
(ref. = 2006)						
2007			0.0243***	(0.0039)	0.0221***	(0.0038)
2008			0.0632***	(0.0038)	0.0591***	(0.0038)
2009			0.0831***	(0.0037)	0.0792***	(0.0036)
2010			0.0867***	(0.0036)	0.0799***	(0.0037)
2011			0.0801***	(0.0036)	0.0740***	(0.0036)
GPA						
Written Danish			0.0078***	(0.0005)	0.0067***	(0.0005)
Written Maths			0.0255***	(0.0005)	0.0228***	(0.0005)
Region						
(ref. = Northern Jutland)						
Central Jutland			-0.0401***	(0.0036)	-0.0433***	(0.0036)
Southern Denmark			-0.0106**	(0.0036)	-0.0115**	(0.0036)
Capital Region			-0.0375***	(0.0040)	-0.0503***	(0.0040)
Zealand Region			-0.0215***	(0.0039)	-0.0246***	(0.0039)

(continued on next page)

Table 2 (continued)

	(1)		(2)		(3)	
Covariates	*M.E.*	*(S.E.)*	*M.E.*	*(S.E.)*	*M.E.*	*(S.E.)*
Family type (ref. = nuclear family)						
Divorced, with new partner					-0.0288***	(0.0039)
Single parent					-0.0315***	(0.0030)
Not living with parents					-0.0564***	(0.0032)
Mother's education (ref. = elementary school)						
Missing education information					-0.0024	(0.0106)
General upper secondary education					0.0117*	(0.0052)
Vocational upper secondary education					0.0042	(0.0032)
Short-cycle higher education					0.0272***	(0.0054)
Medium-cycle higher education					0.0221***	(0.0034)
Long-cycle higher education					0.0614***	(0.0050)
Father's education (ref. = elementary school						
Missing education information					-0.0004	(0.0084)
General upper secondary education					0.0116*	(0.0053)
Vocational upper secondary education					0.0025	(0.0031)
Short-cycle higher education					0.0295***	(0.0052)
Medium-cycle higher education					0.0269***	(0.0038)
Long-cycle higher education					0.0656***	(0.0043)
ln (household income)					0.0445***	(0.0035)
Observations	177,390		177,390		177,390	

Note: *** p<0.001, ** p<0.01, * p<0.05. The table presents average marginal effects on the probability of participation in HE, from a binary logit model. Standard errors in parentheses. Distances are in km.

5.3 Parental Education and Distance Effects

Previous work has shown that the association between distance from home to the nearest HE institution and participation in HE is stronger among young individuals on the margin of participating, according to their prior attainment and family background (Cullinan et al., 2013; Frenette, 2006; White & Lee, 2020). I examine the magnitude of this well-established association in the context of the Danish welfare system. Table 3 reports the third model specification (with all control variables) for different socio-economic groups: Group 1 (SES 1) comprises students where neither the mother nor father had an education above the level of elementary school. Group 2 (SES 2) comprises students for whom the highest education of both parents was a vocational upper secondary education qualification. Group 3 (SES 3) comprises students where the highest education of both parents was a short- or medium-cycle higher education. Finally, group 4 (SES 4) comprises students where at least one of the parents held a university degree.

Column (2) reports that students from SES group 2 are sensitive to distance to the nearest HE institution. The estimate is significant and has a negative sign. The effects of distance to the nearest HE institution is statistically insignificant for all the other groups. On the other hand, distance to the third nearest HE institution is significant and has a positive sign for SES groups 2, 3 and 4, conditional on distance to the remaining nearest HE institutions. The positive association between distance to the third nearest HE institution and participation in HE suggests that young individuals who live in areas that have lower accessibility to HE institutions (e.g. not in city centres) are more likely to participate in HE. Hence, there must be residential sorting on unobserved factors that the model does not take into account.

Table 3
Association between distance and participation in HE, by parental education.

	(1)	(2)	(3)	(4)
	SES 1	SES 2	SES 3	SES 4
Distance to 1st nearest	-0.0005	-0.0008***	0.0000	-0.0003
	(0.0005)	(0.0002)	(0.0002)	(0.0004)
Distance to 2nd nearest	0.0004	-0.0001	-0.0004*	-0.0001
	(0.0006)	(0.0003)	(0.0003)	(0.0004)
Distance to 3rd nearest	-0.0004	0.0004**	0.0005***	0.0006**
	(0.0004)	(0.0002)	(0.0002)	(0.0003)
Individual controls	Yes	Yes	Yes	Yes
Family controls	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes
Participation	0.698	0.684	0.747	0.829
Observations	14,827	63,758	68,112	30,693

Note: *** p<0.001, ** p<0.01, * p<0.05. The table presents average marginal effects on the probability of participation in HE, from a binary logit model. Standard errors in parentheses. Distances are in km. Control variables are gender, age, GPA, immigrant status, year dummies, upper secondary education, family type, father's highest education, mother's highest education, household income and region fixed effects.

Figure 4 illustrates the conditional probabilities for participation in HE for each of the four groups. The figure shows that the AME of distance at -0.0008 in SES group 2 is equivalent to a reduction in the average participation rate from 69.4 percent (among those living 1 km away from the nearest HE institution) to 65.3 percent (among those living 50 km away). Thus, a reduction of approximately four percentage points, from one to 50 kilometres. This is not completely trivial, but should be judged against other factors that influence HE participation.

Previous work has found a strong relationship between school grades and socio-economic background on HE participation (Cullinan et al., 2013; Gibbons & Vignoles, 2012). Figure 4 also illustrates the strong association between parental education and HE participation rate in Denmark; if at least one of the parents held a university degree (SES group 4), about 83 % of these students enrolled on a HE programme within two years of graduating from upper secondary school. If the parents held a vocational education, the equivalent enrolment rate was 68 %. This is a difference of 15 percentage

points. In this context, the small geographical associations in Table 2 and Table 3 seem inconsequential.

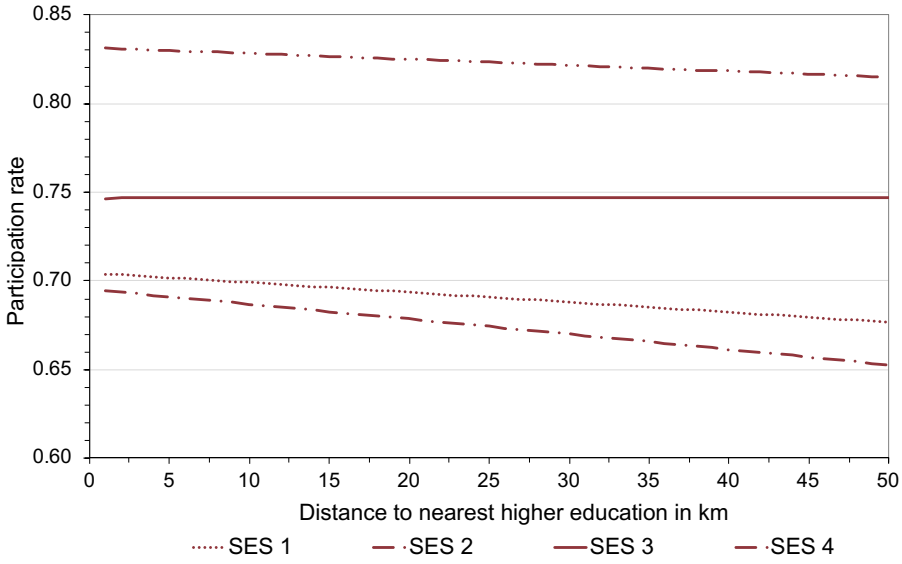

Figure 4. Association between home proximity to the nearest HE institution and participation in HE by parental education. The conditional probabilities are derived from a binary logit model. Distances are in km. Control variables are: distance to second and third nearest HE institution, gender, age, GPA, immigrant status, year dummies, upper secondary education, family type, father's highest education, mother's highest education, household income and region fixed effects. All control variables are held constant at their means.

6. Empirical results: Association between proximity and choice of education

In this section, I consider the role of distance from home to different types of HE on individuals' choice of HE education programme. To recap, I estimate a multinomial logit model to estimate the association between distance from home to the three nearest colleges and the three nearest universities, on three choice alternatives: no enrolment in HE, enrolment in college education programme or enrolment in a university education programme. The choices (and the probabilities) are conditional on graduation from general upper secondary education.

6.1 Regression models of choice of higher education

Table 4 presents the average marginal probability of participation in college or university relative to no participation in HE (the referent group), conditional on distances to the three nearest colleges, the three nearest universities, and individual and family controls. The fourth row shows that the further students live from the nearest university, the less likely they are to enrol in a university education. A 10-kilometre increase in distance is associated with a reduction of 0.6 percentage point in the probability of university enrolment, relative to no HE enrolment. By contrast, the further students live from a university, the more likely they are to enrol in a college education, relative to no education. Hence, the results suggest that at the margin, the absence of a local university will persuade a small number of individuals to switch to a college education.

Table 4
Association between distance and choice of education.

Covariates	College	University
Distance to 1st nearest college	-0.0001	-0.0000
	(0.0001)	(0.0002)
Distance to 2nd nearest college	-0.0001	-0.0001
	(0.0002)	(0.0002)
Distance to 3rd nearest college	0.0008***	-0.0009***
	(0.0001)	(0.0002)
Distance to 1st nearest university	0.0005***	-0.0006***
	(0.0001)	-0.0001
Distance to 2nd nearest university	0.0003***	0.0002*
	(0.0001)	(0.0001)
Distance to 3rd nearest university	-0.0002***	0.0004***
	(0.0001)	(0.0001)
Individual controls	Yes	Yes
Family controls	Yes	Yes
Region fixed effects	Yes	Yes
Participation	0.28	0.45
Observations	173,703	173,703

Note: *** p<0.001, ** p<0.01, * p<0.05. The table presents average marginal effects on the probability of participation in type of HE, from a multinomial logit model. Standard errors in parentheses. Distances are in km. Control variables are gender, age, GPA, immigrant status, year dummies, upper secondary education, family type, father's highest education, mother's highest education, household income and region fixed effects.

6.2 Association between distance and study choice depends on family background

Table 5 splits the results from Table 4 by parental education in order to analyse whether there are differences between groups in their sensitivity to distance to colleges and universities respectively. For students in SES group 1, SES group 2, and SES group 3, we see a negative significant association between distance to nearest university and university participation. For students in SES group 4, all estimates are insignificant. The magnitude of the estimates is largest for SES group 1: a 10-kilometre increase in distance to university decreases the probability of university participation by 1 percentage point. Table 5 also reveals large differences between groups in their study choices: in SES group 1, the choice between college and university is fifty-fifty (35 % vs 35 %). In SES group 4, about 14 % go to college and 70 % to university.

Figure 5 illustrates the conditional probabilities for SES group 1 for participation in college, university or no participation in HE. The solid curve shows that the average university participation rate falls as distance to the nearest university increases: average university participation decreases from 37 percent (among those living 1 km away) to 29 percent (among those living 100 km away). Hence, we see a decrease of eight percentage points when distance increases from one to 100 kilometres. The results suggest that individuals from different family backgrounds differ in their sensitivity to distance when they choose education; students with lower-educated parents are more sensitive to distance than their peers with highly educated parents.

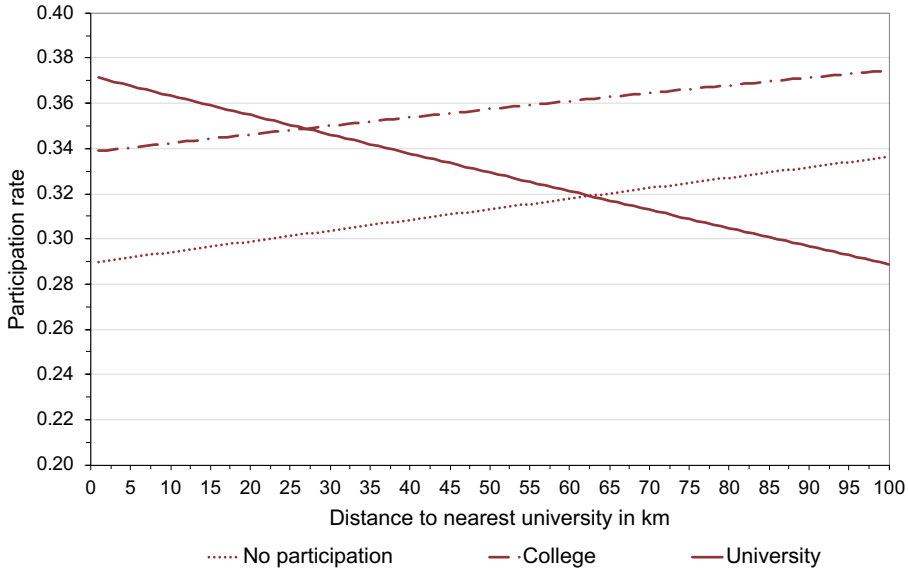

Figure 5. Association between proximity to nearest university and choice of education in SES group 1.
Note: The figure illustrates conditional probabilities of participation in college, university or no HE participation, conditional on distance to nearest university. The conditional probabilities are derived from a multinomial logit model. Distances are in km. Control variables are: distance to the three nearest colleges, distance to second and third nearest universities, gender, age, GPA, immigrant status, year dummies, upper secondary education programme, family type, father's highest education, mother's highest education, household income and region fixed effects. All control variables are held constant at their means.

Table 5
Association between distance from home to HE institution and choice of higher education, by parental education.

	SES 1		SES 2		SES 3		SES 4	
Covariates	College	University	College	University	College	University	College	University
Distance to 1st nearest college	0.0003	-0.0003	-0.0003	-0.0002	-0.0001	0.0003	-0.0000	0.0001
	(0.0006)	(0.0006)	(0.0003)	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0004)
Distance to 2nd nearest college	-0.0002	-0.0002	-0.0004	0.0005	0.0001	-0.0006**	0.0002	-0.0004
	(0.0006)	(0.0006)	(0.0003)	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0005)
Distance to 3rd nearest college	0.0008*	-0.0011**	0.0009***	-0.0012***	0.0008***	-0.0007***	0.0003	-0.0002
	(0.0005)	(0.0005)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0004)
Distance to 1st nearest university	0.0004	-0.0010***	0.0007***	-0.0007***	0.0004***	-0.0003**	-0.0000	-0.0001
	(0.0003)	(0.0003)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Distance to 2nd nearest university	0.0004	0.0002	0.0002	0.0002*	0.0005***	-0.0000	0.0002	0.0001
	(0.0003)	(0.0003)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Distance to 3rd nearest university	-0.0002	0.0006*	-0.0001	0.0003**	-0.0004***	0.0004***	-0.0000	0.0003
	(0.0003)	(0.0003)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0002)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Participation	0.347	0.353	0.346	0.338	0.264	0.483	0.139	0.690
Observations	14,517		62,016		66,777		30,393	

Note: *** p<0.001, ** p<0.01, * p<0.05. The table presents average marginal effects on the probability of participation, from four multinomial logit models. Standard errors in parentheses. Distances are in km. Control variables are gender, age, GPA, immigrant status, year dummies, upper secondary education, family type, father's highest education, mother's highest education, household income and region fixed effects.

7. Summary and concluding remarks

This paper has examined the role of distance from home to HE institutions on participation in and choice of HE programmes in Denmark. The current HE policy in Denmark emphasizes spatial decentralization of higher educations as a tool to increase enrolment in HE and narrow the socio-economic gap between rural and urban regions (Broberg, 2021; Ministry of Higher Education and Science, 2020a). The assumption of this policy is that distance to HE acts as deterrent to enrolment among young individuals in peripheral regions. However, no empirical evidence exists on the effect of distance from home to HE institutions in a Danish context. The purpose of the present study is to fill this gap.

I have used a unique data source of all home-to-HE-institution distances, calculated through the road network. This distance data was linked to rich administrative data sources of the full population of young individuals who graduated from a general upper secondary education between 2006 – 2011. Using these unique data sources, this paper provided new descriptive knowledge about the geography of HE in Denmark and the association between location of HE and geographical disparities in basic human capital formation though the channel of student enrolment.

The descriptive results demonstrate that the majority of students who graduated from upper secondary education had access to an HE institution within commuting distance; more than 90 percent lived within 30 kilometres of the nearest HE institution and 35 percent of the students lived within 5 kilometres. Using a binary logit model, I examined if distance from home to the three nearest HE institutions was associated with participation in HE, after controlling for a range of individual, regional and family controls.

The results showed that distance from home to HE institutions was negatively associated with enrolment in all types of HE. However, the impact is *very small in magnitude.* The largest marginal effect, around -0.0008, was found among individuals from lower SES homes. The effect implies that increasing the distance from home to the nearest HE institution by 10 kilometres would decrease young individuals from SES-group 1's probability of enrolment in HE by 0.8 percentage points. Hence, the results suggest that geographical proximity to HE *is not* an important factor for HE participation in Denmark. Factors such as socio-economic background, grades in upper secondary school and the type of upper secondary school are found to be much more important. This implies very little empirical evidence for policies that aim to widen HE participation through increased geographical accessibility to HE in Denmark.

I also considered the role of distance from home to different types of HE institutions on individuals' the choice of education programme (college vs. university). Even though the association between distance and enrolment at any type of HE institution was little, the geography of the colleges and universities could potentially have a large impact on the skill composition of the populations of their local areas and hence on geographical disparities in human capital. The results from a multinomial logit model showed that an increase in distance from home to the nearest university was associated with a higher probability of college enrolment and a lower probability of university enrolment, conditional on the distance to the other nearest HE institutions and a range of controls. Additional models confirmed that students from different socio-economic backgrounds differed in their sensitivity to distance on their choices; among students from high SES families, there was no association between distance from home to their nearest university and their choice of education programme. Among students from lower SES backgrounds, their probability of attending a university decreased significantly when they lived farther away from a university.

With the present geographical dispersion of HE institutions in Denmark, it seems that HE location plays no role in creating spatial disparities in human capital though student enrolment. However, the migration decisions of HE graduates may still be an important factor in shaping regional human capital disparities (Faggian & Franklin, 2014; Faggian & McCann, 2009). In addition, it should be borne in mind that distances to HE institutions are relatively small in Denmark, which means that the present study does not consider the impact of growing up in remote places. Another shortcoming of the present study is that it is based on observational data. Several studies point to the fact that the association between home proximity to educational institutions and educational choices is likely to depend on residential sorting of households of different types into different geographical areas.

References

Berry, C. R., & Glaeser, E. L. (2005). The divergence of human capital levels across cities*. *Papers in Regional Science*, *84*(3), 407–444.

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.

Broberg, M. B. (2021, January 21). Kaare Dybvad: "Vi har centraliseret for meget." *Jyllands-Posten*.

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).

Eurostat, European Union. (2016). *Urban Europe—Statistics on cities, towns and suburbs*. Luxembourg.

Faggian, A., & Franklin, R. S. (2014). Human Capital Redistribution in the USA: The Migration of the College-bound. *Spatial Economic Analysis*, *9*(4), 376–395.

Faggian, A., & McCann, P. (2009). Human capital and regional development. In *Handbook of Regional Growth Development Theories*. Edward Elgar.

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

Gibbons, Stephen, & Vignoles, A. (2012). Geography, choice and participation in higher education in England. *Regional Science and Urban Economics*, *42*(1), 98–113.

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

Glaeser, E. (2013). A Review of Enrico Moretti's The New Geography of Jobs. *Journal of Economic Literature*, *51*(3), 825–837.

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.

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

Kildegaard, K., & Domino, S. (2019, October 8). Mette Frederiksen i første store interview som statsminister: Vi ser et mere splittet land end før". *Berlingske*.

Ministry of Higher Education and Science. (2020a, October 12). Nye muligheder for uddannelse i hele landet—Uddannelses- og Forskningsministeriet

Ministry of Higher Education and Science. (2020b, December 3). The admission system in Denmark—Uddannelses- og Forskningsministeriet [Page]. Retrieved March 10, 2021, from https://ufm.dk/en/education/admission-and-guidance/how-to-apply-for-a-higher-education-programme-in-denmark-1

Moretti, E. (2004). Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, *121*(1), 175–212.

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.

Parker, K., Horowitz, J., Brown, A., Fry, R., Cohn, D., & Igielnik, R. (2018). *What Unites and Divides Urban, Suburban, and Rural Communities*. Washington DC: Pew Research Center. Retrieved from Pew Research Center website: https://www.pewresearch.org/social-trends/2018/05/22/what-unites-and-divides-urban-suburban-and-rural-communities/

Parker, P. D., Jerrim, J., Anders, J., & Astell-Burt, T. (2016). Does Living Closer to a University Increase Educational Attainment? A Longitudinal Study of Aspirations, University Entry, and Elite

University Enrolment of Australian Youth. *Journal of Youth and Adolescence*, *45*(6), 1156–1175.

Polat, S. (2017). The Expansion of Higher Education in Turkey: Access, Equality and Regional Returns to Education. *Structural Change and Economic Dynamics*, *43*, 1–14.

Romer, P. M. (1994). The Origins of Endogenous Growth. *Journal of Economic Perspectives*, *8*(1), 3–22. doi: 10.1257/jep.8.1.3

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.

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/

White, P. M., & Lee, D. M. (2020). Geographic Inequalities and Access to Higher Education: Is the Proximity to Higher Education Institution Associated with the Probability of Attendance in England? *Research in Higher Education*, *61*(7), 825–848.

Appendix

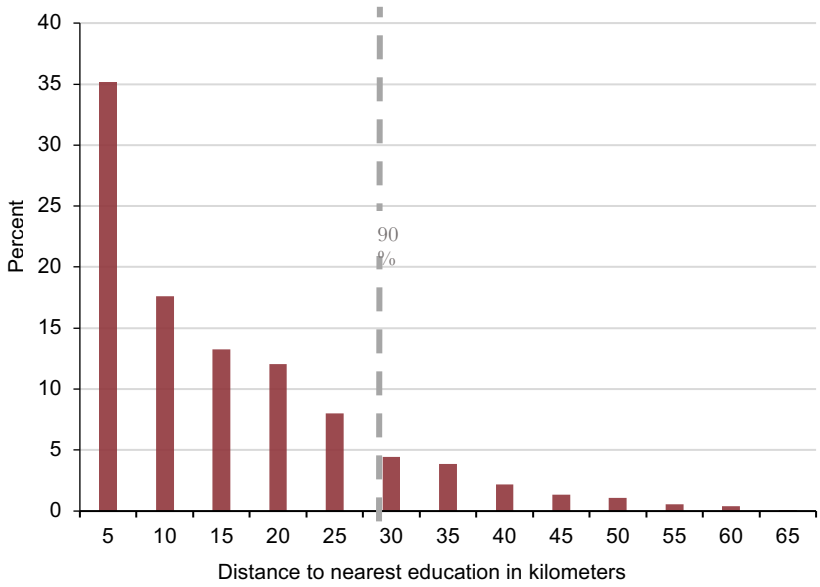

Figure A1. Histogram of distance from home to the nearest HE institution, 2006 – 2011.

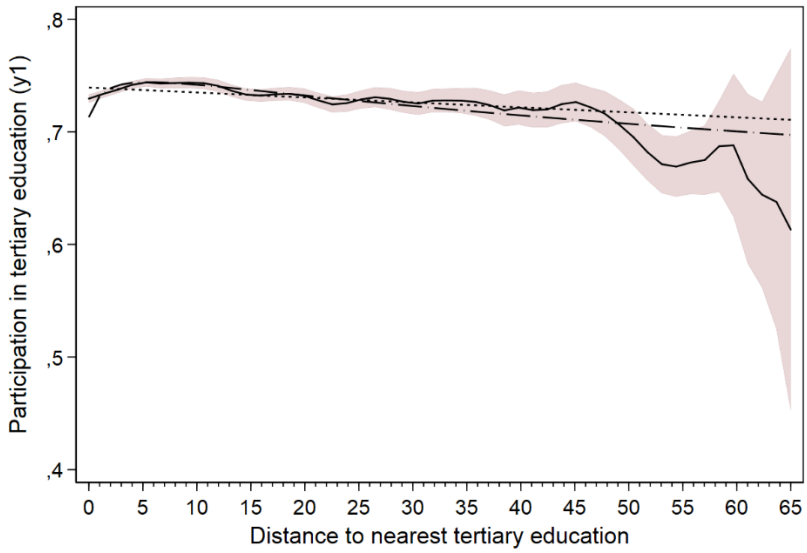

Figure A2. Association between distance from home to nearest HE institution and study choice.

Paper 4

PAPER 4

Regional Graduate Retention Trajectories in Denmark: Education, Location and City Size

Elise Stenholt Sørensen* and Lars Winther**

Abstract

This paper examines the graduate retention rates of higher education graduates in Denmark. We use data from administrative sources of information on graduate migration in the first 10 years after graduation. Our results show that new higher education graduates are highly mobile geographically, more than 50 percent having moved from their city of studies within 10 years of graduation. However, there are large spatial differences in these mobility patterns. Graduate retention is much higher among graduates in the capital city of Copenhagen than in other cities with higher education institutions. In addition, migration after graduating is substantially more likely for those who moved to study initially compared to those who studied in their home region. The findings point to the importance of city size and local labour markets for graduates' migration decisions.

Keywords

Graduate Migration, Regional Development, Regional Migration, Human Capital, Location of Higher Educations, Denmark

JEL C41, I25, R23

* Department of Geosciences and Natural Resource Management, Geography Section, University of Copenhagen and Department of the Built Environment, Aalborg University

** Department of Geosciences and Natural Resource Management, Geography Section, University of Copenhagen

1. Introduction

This paper examines to what extent cities and regions with higher education institutions (HEIs) succeed in retaining newly educated human capital within their borders. Regional growth is dependent on regional levels of human capital (Faggian, Rajbhandari, & Dotzel, 2017; Gennaioli, La Porta, Lopez-de-Silanes, & Shleifer, 2013; E. L. Glaeser, Scheinkman, & Shleifer, 1995). We analyse the residence duration of higher education graduates in Denmark from the time of graduation up until 10 years afterwards, which we call 'retention trajectories'. We are interested in how large a share of new graduates that remains in the region and for how long. It is evident that larger and more developed regional labour markets would be more likely to ensure the retention of such graduates because of better job-matching opportunities. However, in smaller regions, graduates may be more prone to acquire location-specific human capital and therefore stay in the area after graduation.

Regional stocks of human capital can be augmented by encouraging greater participation in the education system, for example, by accommodating HEIs. In policy terms, a geographically dispersed network of such institutions is seen as a way of equalizing uneven regional development (Kildegaard & Domino, 2019; Ministry of Higher Education and Science, 2020; Vestergaard, 2015). The claims made for this policy are that this helps reduce the brain drain from less developed regions and that it could also attract young people from other parts of the country (Amendola, Barra, & Zotti, 2020). It is, however, evident that highly educated individuals are a very mobile group and that human capital may therefore easily 'leak' to other regions (Faggian, Corcocan, & Partridge, 2015; Haapanen & Bockerman, 2017).

In this paper we adopt the perspective of residential duration in order to examine the geographical aspects of the urban and regional retention of graduates. We do this for two reasons. First, the significance of the gain in human capital in an area is likely to depend on how long highly educated individuals will stay within it. To our knowledge, earlier duration analyses of graduates of higher education is limited to studies by Busch and Weigert (2010), Haapanen and Tervo (2012) and Teichert, Niebuhr, Otto and Rossen (2020). Their results from Germany, Finland and Germany respectively show that the longer an individual stays where he or she studied the smaller the probability of migration in the future. They also show that a majority of graduates (about 70 percent) were still living in the region where they completed their studies 10

years after graduation. However, all three studies find that peripheral regions are faced with significantly higher rates of out-migration than central urban regions. Second, duration analysis enables us to study the importance of duration dependence relative to other factors explaining the migration decisions of graduates.

To study regional differences in the retention of recent graduates from higher education, we distinguish two groups of individuals: those studying in the 'home' region, and those studying 'away' from home. This distinction is important for two reasons. First, attachment to one's region of studies is likely to differ (Busch & Weigert, 2010; DaVanzo, 1983; Faggian, McCann, & Sheppard, 2007a; Haapanen & Tervo, 2012). Secondly, it helps determine whether a local education institution is beneficial to the region or not. We also examine the role of geography by conducting the analysis at the scale of both cities and regions (NUTS 3).

In addition, we assess the role of educational levels and types of education. Previous literature has revealed a relationship between the level of human capital and migration (Haapanen & Bockerman, 2017; Malamud & Wozniak, 2012). Studies also suggest that the discipline taken also plays a role because of the presence of labour markets (Faggian et al., 2007a; Venhorst, Dijk, & Wissen, 2011). Yet another important factor is family formation (Haapanen & Tervo, 2012; Haussen & Uebelmesser, 2018). We therefore study individuals but take marital status and children into account in the models.

In the analysis, we utilize large-scale longitudinal individual data for all Danish college and university graduates who graduated in 2010. We analyse these graduates' migration patterns during a 10-year period, from the year of graduation until 10 years afterwards, i.e. between 2010 and 2019. Denmark provides a very suitable case for such a study as HEIs are widely dispersed geographically. In addition, Denmark has a universal education system, with no tuition fees in higher education. Therefore, differences in the cost of education cannot bias the results, making it a solid case with which to study the influence of specific educational fields and education level on graduate migration rates.

The paper is organized as follows. Section 2 discusses the literature on (HEIs), local economic development and the determinants of graduate migration. Section 3 presents the statistical methods, data sources and variables used in the paper. Section 4 summarizes and discusses the results. Finally, Section 5 concludes the paper and considers some policy implications.

2. Regional Development, Human Capital and City Size

Highly skilled human capital is a major resource for economic growth in the capitalist economies. This is reflected in the Romer–Lucas models in economics (Lucas, 1988; Romer, 1986). According to endogenous growth theory, economic growth depends on the production of knowledge with a focus on increasing returns, positive externalities and knowledge spillover effects (Lucas, 1988; Romer, 1986). Thus, from an endogenous growth perspective, regional growth also depends on regional human capital and knowledge production, emphasizing regional learning and geographical proximity (Hansen & Winther, 2015; Simon & Nardinelli, 2002; Storper & Venables, 2004).

In this perspective, labour and labour skills are produced and reproduced through regional institutions, including educational institutions, regional labour markets and industrial structures. This results in a diversity of regional production and innovation systems contributing to the endogenous growth of regions based on knowledge production supported by specific capabilities, competences and skills (Storper, 1997). Human capital, in the form of the highly skilled, is concentrated in metropolitan regions (Florida, 2002; Glaeser, 2011). This emphasises the fact that urban economies are central to human capital formation because of, for instance, higher returns from investment in human capital (wages) and greater employment potential (Florida, 2002; Glaeser, 2011; Storper, 2010).

Thus, it is evident that urban size, especially in the form of large metropolitan areas, is positively associated with the intensity and productivity of human capital in regional industries (Elvery, 2010). This focus on urbanisation economies, including the variety of industries, knowledge spillovers and the potential for face-to-face interaction, has tended to overlook small and medium-sized cities (Hansen & Winther, 2018).

Higher Education Institutions and Regional Economic Development

In the past twenty years there has been a growth of interest in measuring the impact of universities on regional development and the retention, and attraction of local human capital (Amendola et al., 2020; Drucker & Goldstein, 2007). Changes in the world's developed economies from manufacturing to knowledge-intensive production and services has been a motivating force behind this growing attention both in research and among policy-makers.

Universities can have an impact on the local economy in a variety of ways, ranging from short-term effects on economic activities (e.g., employee salaries, industries providing goods and service to universities, the impact of students' expenditure) to long-term effects such as human capital creation, technological innovation and knowledge spillovers, which sustain long-term development and growth (Drucker & Goldstein, 2007). We will not review this extensive literature here, since these effects have been well described. Drucker and Goldstein (2007) and Amendola et al. (2020) provide excellent reviews.

Our analysis therefore focuses instead on a specific aspect of the role of HEIs in creating economic development, namely their influence on the local stock of human capital. HEIs can raise human capital levels locally by increasing both the supply and demand for skills. However, it is not evident that regions with HEIs will benefit from the production of highly skilled graduates, as embodied human capital in the form of education increases the probability of migration significantly (Faggian et al., 2015, 2007a; Haapanen & Bockerman, 2017; Malamud & Wozniak, 2012; Venhorst et al., 2011; Whisler, Waldorf, Mulligan, & Plane, 2008). Thus, human capital can easily 'leak' out of a region, reducing the region's multiplier effects and spillovers of HEIs. Even within countries, there are often large differences in the ability to attract and retain higher education graduates e.g. see Faggian & Mccann (2009) and Haapanen & Tervo (2012). Thus, the regional rates of return to higher education depend crucially on the migration behaviour of university graduates.

Determinants of graduate migration

In the literature, the determinants of graduate migration are often categorized into three types: individual characteristics, study-related factors and regional factors (Faggian et al., 2015; Teichert, Niebuhr, Otto, & Rossen, 2020). The main determinants of the *individual* characteristics are gender (Faggian, McCann, & Sheppard, 2007b; Krabel & Flöther, 2014), age (Faggian & Mccann, 2009; Schwartz, 1976), ethnicity (Faggian, McCann, & Sheppard, 2006), family ties (Busch and Weigert, 2010; Haussen and Uebelmesser, 2018), work experience (Haapanen and Karhunen, 2017; Teichert et al., 2020) and migration history (Faggian et al., 2007b; Haapanen and Tervo, 2012; Rehák and Eriksson, 2020). Despite country-specific differences in the regional categorizations and geographical locations of universities, some empirical regularities have emerged.

Young graduates are more mobile than older ones (Faggian et al., 2007b), and non-White ethnicities have been found to be less mobile than Whites in the UK (Faggian et al., 2006). Regarding gender differences, the evidence is mixed. Faggian et al. (2007b) find that UK female graduates are more likely to migrate after graduation, while Krabel and Flöther (2014) find no significant differences among female and male graduates in Germany. Family ties, in terms of being and/or having children, reduce the probability of migration, the emotional costs of which tend to be larger, as the graduate's partner and/or children also need to adapt to the new environment (Haussen and Uebelmesser, 2018). Work experience while studying is found to reduce graduate migration (Haapanen and Karhunen, 2017; Teichert et al., 2020), while previous migration to pursue higher education increases graduate migration significantly (Faggian et al., 2007b; Haapanen and Tervo, 2012; Rehák and Eriksson, 2020).

The *study-related* factors that determine graduate migration include educational level (Haapanen and Bockerman, 2017; Malamud and Wozniak, 2012), the quality of the HEI (Rehák and Eriksson, 2020) and the field of study (Venhorst, Dijk, and Wissen, 2010; Venhorst et al., 2011). Haapanen and Böckerman (2017) and Malamud and Wozniak (2012) provide causal evidence showing that obtaining a post-secondary educational qualification increases the probability of migration significantly. The quality of education also seems to matter, but the evidence suggests that the association between educational quality and migration depends on previous migration and the level of development of the university region (Rehák and Eriksson, 2020). Local graduates of high-quality HEIs in Sweden are more likely to migrate, whereas the opposite tendency is seen among non-local graduates (Rehák and Eriksson, 2020). Hence, Rehák and Eriksson (2020) argue that 'there is a need for a differentiated measure aiming at retention of local and non-local graduates'.

Regional labour market conditions and local amenities, i.e. *regional factors*, have a significant influence on graduate migration (Hansen and Niedomysl, 2009; Rehák and Eriksson, 2020). There is an extensive literature examining how local wage levels and the match between local HEIs and jobs in the region influence migration patterns (Drucker & Goldstein, 2007; Faggian et al., 2017). In addition, the quality and availability of local amenities have been shown to play a role (Florida, 2003; Glaeser and Gottlieb, 2006).

3. Data and Methodology

3.1 Method

Within the field of inter-regional migration, duration models have been used to describe the duration from an event happening until individuals decide to move (Busch and Weigert, 2010; Haapanen and Tervo, 2012; Jans, 2005). This method was already being recommended by DaVanzo as far back as 1982 (DaVanzo, 1982), despite which there are still relatively few studies of migration that use the method. Duration models are a powerful statistical tool for modelling the duration of a given event. They are commonly used in medical research to investigate the association between the survival time of patients and medical use. Therefore, the type of analysis is often referred to as 'survival analysis' (Jenkins, 1995).

In this analysis, we first model the timing of mobility among recent graduates using several Kaplan-Meier survival curves. The Kaplan-Meier curve shows the time before the event of interest occurs, for example, moving (Flynn, 2012; Jenkins, 2004). The survival curves serve as a descriptive, analytical tool with which to acquire an overview of overall graduate retention rates. Secondly, we analyse the association between different factors and the propensity (risk) of moving by means of Cox's Proportional Hazard model (Cox, 1972), a regression model commonly used for investigating the association between the timing of an event and one or more explanatory variables. The Cox regression model extends Kaplan-Meier survival analysis by assessing the effect of several risk factors on survival time. The Cox proportional hazard model works by modelling the hazard function. The hazard rate can be interpreted as the risk of experiencing the event of interest (moving) at any given time during the follow-up period (the time after graduation) (Flynn, 2012).

In assessing the duration until the individual moves out of the city or region of his or her studies for the first time, we restrict the assessment to a single spell analysis. Therefore, we do not observe return migration in the present model. We analyse the risk of moving given a range of explanatory variables using the Cox Proportional hazard rate model.

The dependent variable is the hazard rate $\lambda(t)$ of moving out, defined as

$$\lambda\big(t|X(t)\big) = \lim_{dt\to 0} \frac{P\big(t, t+dt|T \geq t|X(t)\big)}{dt}$$

where T is the positive, continuous variable for the time taken to exit a given state (or: the time of the event of interest), t any fixed point in time under risk, and $P(t,t+dt)$ the probability that migration occurs in the time interval [t,t+dt). $X(t)$ is a vector of the explanatory variables. The effects of the explanatory variables are assumed to be constant over the observation period.

The output from the Cox model is the Hazard Ratio (HR), which is interpreted as a relative risk (Flynn, 2012). If the HR is >1, it indicates an increased risk of an event (moving) being associated with that variable. If the HR is below 1, it indicates a reduced risk, negatively associated with the length of survival. An HR close to zero indicates no association between the variable and experiencing the event.

3.2 Data

We use data from Danish administrative registers in order to analyse the migration patterns of the highly educated. The administrative registers are a rich source of panel data on individuals, retained by Statistics Denmark. We link administrative data sources that contain information about individual characteristics, such as age, gender and residence, with educational records[1] and family characteristics. Through a unique mother-id and father-id number, each individual is linked to their respective mothers and fathers. This enables us to take parental education and incomes into account in our analysis. All variables are registered daily, monthly or yearly, depending on the variable. The panel structure of the data makes the administrative registers a rich and highly suitable source of information for studies of duration.

The sample population includes all graduates who graduated from higher education in Denmark in 2010, a total of 48,536 individuals. Individuals under 20 and older than 60 at the time of graduation are omitted from the sample.[2] We also exclude individuals for whom information is lacking about their residential addresses at the time of graduation (about 3 percent). Finally, being interested in migration patterns out of the city of region of education, we restrict

[1] Via a personal identification number (PNR) we can link individuals to the educational institution, where they graduated. The institutional dataset contains central information of the educational institution, such as geographical location, educational field, educational level and so forth.

[2] We omit the few outliers above 60 years, as we are interested in the work force's mobility patterns.

the analysis to the population who were living in the same geographical area (city or region) as their HEI at the time of graduation. This leaves us with a sample of 32,782 graduates. Table A1 in the appendix provides a more detailed description of the data reduction.

3.3 Measuring Spatial Mobility

Migration is defined as a long-distance move out of the city or region of the graduate's studies. We examine two related outcome measures: 1) migration out of the city of studies, and 2) migration out of the region of studies. We make this geographical distinction because the geographical scale and spatial unit of analysis could influence the results and the conclusions inferred from them (Chatagnier & Stillwell, 2021). Decisions to move shorter distances, i.e. between or within municipalities, are more likely to be due to family enhancements or changes to the housing market. Long-distance migration allows us to study changes in the geographical distribution of human capital.

The dependent variable is a binary variable indicating whether the graduate has moved out of the region or city where he or she obtained his or her degree at time *t*. We analyse all higher education graduates and measure their movement patterns every month from when they graduate until when they move in the ten years after graduation. Information about residential addresses and moving decisions in those ten years are obtained from the 'population register' (BEF), which contains detailed information on the exact date of moving and the residential address. The BEF also contains information about gender, age, and the unique mother-ID/father-ID number.

3.4 Explanatory Variables

Tables 1 and 2 describe the key explanatory variables: origin, type of education and location.

Origin

Graduates in higher education form a very heterogenous group. Although individual- and study-related factors can explain some of the variation in migration rates, the most important distinction is between graduates who study in their home regions and graduates who move to another region to study (Rehák and Eriksson, 2020). There is a large body of research suggesting that migration subsequent to graduating is closely correlated with previous migration behaviour (Ahlin, Andersson, & Thulin, 2018; DaVanzo, 1983;

Faggian et al., 2007a; Haapanen & Tervo, 2012). Therefore, the models are estimated separately for those studying in their home city or region and those studying away from home. Here, 'home' is defined as the city or region where the student was living 10 years before enrolment in higher education. This familiar region is referred to as 'origin' in the analysis. Based on 'origin', we categorize the sample into three distinctive groups: 1) graduates 'studying at home' who attend higher education in their home region; 2) graduates 'studying away' who leave their home region to study in another region; and 3) 'foreign', the residual group for whom we do not have information about their residential addresses 10 years prior to their enrolment in higher education. The majority of the 'foreign' group are categorised as first-generation immigrants in the administrative records, hence it is very likely that they are international students. Table 1 shows that most graduates, 57 percent, attended an HEI in their home region, 34 percent left their home region to study in another region, and 9 percent were categorized as 'foreign' in the analysis because there was no information on their 'origin'.

Education

Here we include two educational measures, one measuring the educational level, in four categories, the other the educational discipline or subject. We categorize higher education into four categories: short-cycle, medium-cycle, bachelor's degree and long-cycle, as shown in Table 1. In addition, we categorize education into eight different educational disciplines related to their future workplace and labour market. The administrative 'Pupils' Register' (KOTRE) is used to identify all individuals who graduated from higher education in Denmark in 2010. The register contains detailed information about the exact date of graduation, the length and type of education, and a unique identification number for the educational institution from which the student graduated.

Table 1
Summary statistics of main explanatory variables.

		Region population (N1)		City population (N2)		T-test between populations
Covariates		Mean	Std. Dev.	Mean	Std. Dev.	t-test
Origin	Studying at home	0.57	0.50	0.53	0.50	11.81****
	Studying away	0.34	0.47	0.38	0.49	-10.83****
	Foreign	0.09	0.29	0.09	0.29	-2.34**
Education level	Short-cycle higher education	0.11	0.31	0.10	0.29	5.21****
	Medium-cycle higher education	0.30	0.46	0.26	0.44	12.12****
	Bachelor's degree	0.32	0.47	0.35	0.48	-8.72****
	Long-cycle higher education	0.27	0.45	0.29	0.46	-6.38****
Educational field	Human Sciences	0.10	0.30	0.10	0.30	-1.00
	Arts	0.03	0.18	0.04	0.19	-3.96****
	Social Sciences	0.12	0.32	0.11	0.32	2.63***
	Business Economics, Admin. and Law	0.18	0.38	0.19	0.39	-4.57****
	Science and Engineering	0.13	0.34	0.14	0.35	-3.66****
	Social and Health	0.25	0.43	0.23	0.42	6.88****
	Teaching and Learning	0.06	0.25	0.06	0.23	4.05****
	Other	0.13	0.33	0.13	0.34	-1.99**
Observations		41,495		32,782		

Location

We obtain information on the geographical location of the higher education institution from the 'Institution Register' (INST), which contains information on the exact addresses of all institutions in Denmark. We use this to group all relevant institutions into larger geographical categories, that is, cities and regions (see Table 2).

Our definition of 'a city' follows that used by Statistics Denmark, which defines a city (or town) as an area with interconnected houses, the distance between which does not exceed 200 metres (Danmarks Statistik 2018b). We categorize the cities or towns of study into five different groups, according to their population size at the time of graduation (year 2010). Table 2 shows that the majority of higher education graduates were resident in the capital city, Copenhagen, i.e. about 52 percent in 2010. A further 39 percent were living in the three largest cites above 100,000 inhabitants (Aarhus, Odense and Aalborg). (The numbers of graduates in each of the four largest cities is shown in Table A2 in the appendix). Only 0.85 percent were living in towns with 1,000–19,999 inhabitants at the time of graduation.

Table 2
Location of higher education graduates.

Location of HE graduates	Frequency	Percentage
City size		
Town 1,000-19,999 inh.	279	0.85
Town 20,000-49,999 inh.	1,662	5.07
City 50,000-99,999 inh.	1,083	3.30
City > 100,000 inh.	12,781	38.99
Capital city (> 500,000 inh.)	16,977	51.79
Region (NUTS 3 regions)		
Capital Region	20,796	50.12
Bornholm	42	0.10
South and West Zealand	985	2.37
Funen	3,466	8.35
Southern Jutland	2,033	4.90
Eastern Jutland	8,960	21.59
Western Jutland	1,144	2.76
Northern Jutland	4,069	9.81

4. Main results

In this section, we present our duration analysis of graduate migration in Denmark. The two outcome measures, a) moving out of the study region and b) moving out of the study city, are presented simultaneously. Previous international studies have primarily used larger labour-market regions for their spatial definition of mobility: thus, Haapanen and Tervo (2012) use NUTS3 regions in their study of Finland. In order to make our results comparable with international data, we conduct our analysis on the regional level using NUTS 3 regions, as well as on the city level. Our regression results show very similar patterns for regions and cities, thus to avoid repetition we will focus on the 'regional' results in the following description and only refer to the city-level results when they differ significantly. Kaplan Meier Survival curves and Cox regression results for the cities can be found in the appendix.

We initially present the overall descriptive results concerning the size and timing of graduate migration in Denmark and compare the Danish numbers with existing empirical studies of graduate retention rates in other countries. We then present the estimation results and focus on three central factors: origin of the graduate, educational field and level, and regional location of the HEI.

4.1 Graduate Retention Trajectories

In total 41,496 individuals graduated from higher education in Denmark in 2010. Out of this total population, 32,782 individuals (67.5 percent) lived in the city where they obtained their degree and 41,495 (85.5 percent) were living in the region where they obtained their degree at the time of graduation. We are interested in how large a share of the stock of human capital remained in the local labour market and how large a share 'leaked out' to other regions and labour markets within the 10 years after graduation. We are therefore only taking into account graduates who were living in their city or region of education at the time of their graduation.

The Kaplan-Meier survival curves in Figure 1 reveal that most graduates leave the city where their received their degree within 10 years of graduation. On average, 52 percent of graduates leave the city of their HEI and 38 percent leave the region (see Figures 1 and 2). However, there are large regional disparities in the ability to attract and retain highly educated graduates, a point we will return to in Section 4.2. The survival curves also reveal that most of the out-migration happens within the first two years after graduation, in line with

previous findings (Busch & Weigert, 2010; Haapanen & Tervo, 2012). Figure 1 shows that the slope of the curve is steepest within the first year after graduation and it also shows that after 34 months (almost three years) 25 percent of graduates have moved out of the city. The probability of moving becomes smaller over time, and we see a large duration dependence in the probability of moving.

The descriptive Kaplan-Meier survival curves demonstrate that graduates educated in Denmark are highly mobile geographically. Compared to other Nordic and continental European countries, our results suggest that Danish graduates are relatively more mobile. Previous studies of graduate migration from Germany (Busch & Weigert, 2010), Finland (Haapanen & Tervo, 2012) and Sweden (Ahlin et al., 2018) find that the majority of the graduates (70 to 80 percent) do not move from their region of study (NUTS 3) within 10 years. In comparison, we find that 48 percent of Danish graduates remain in their city of study and 62 percent in the study region (NUTS 3). Studies from Britain suggests that British graduates are even more geographically mobile (Faggian & McCann,2009). However, the country-specific differences in regional classifications, education systems and geography make it difficult to compare studies of graduate migration across countries. The differences we describe should be interpreted with this in mind.

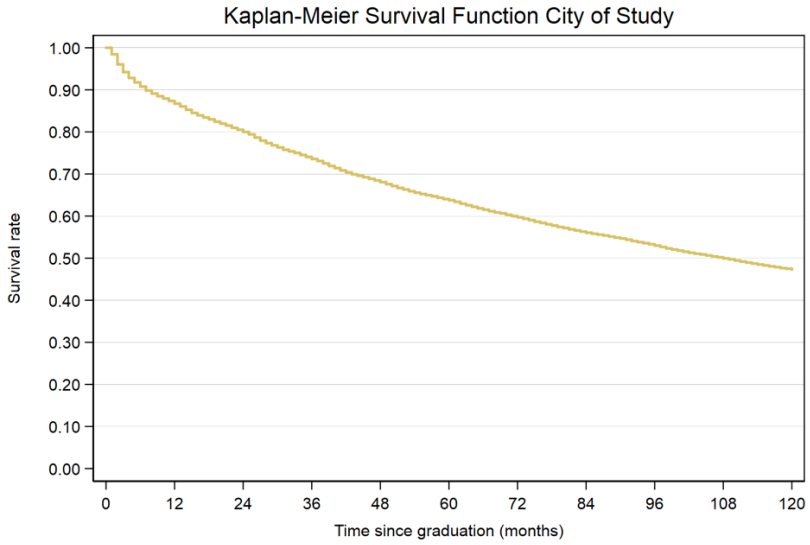

Figure 1. Kaplan-Meier Survival Functions for out-migration from city of education.

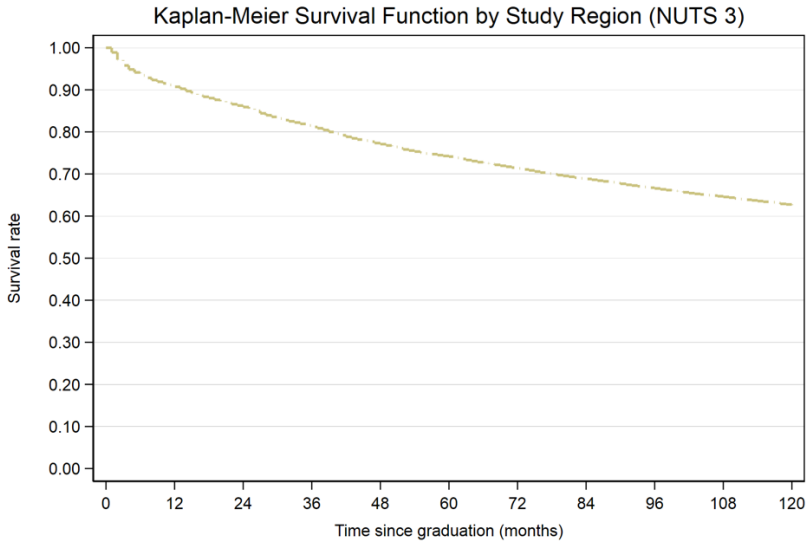

Figure 2. Kaplan-Meier Survival Functions of graduate retention in region of education (NUTS3).

4.2 Origin

In our analysis, we differentiate between three groups of graduates: those studying in their home region, those studying away from home, and those students for whom information about their home region is missing, which we label 'foreign'. This distinction is important because graduates' attachment to the study region is likely to differ according to how long a time they were resident there (DaVanzo, 1983; Faggian et al., 2007b). We therefore present four distinctive regression models: model 1 shows the results for the total population, model 2 estimates out-migration among local students, model 3 estimates out-migration for non-local students, and model 4 estimates out-migration for the foreign students.

Table 3 presents the estimates from the four Cox Proportional Hazard Rate Models of graduate migration from region of education. Overall, the regression analysis in model 1 confirms that graduates studying in their home regions have significantly higher retention rates than those studying away from home. Graduates who move to study are twice as likely to move again (HR = 2.18) compared to those who studied in their home regions. Thus, our results are consistent with previous international studies and confirm that 'movers' are much more likely to move on after graduation than 'stayers'.

Table 3
Hazard Ratios of Moving out of Study Region (NUTS 3).

		Full sample	Studying at Home	Studying Away	Foreign
Covariates		Model 1	Model 2	Model 3	Model 4
Educational level	Short-cycle higher education (ref.)				
	Medium-cycle higher education	1.0132	0.8885*	1.0629	1.1161
		(0.0376)	(0.0518)	(0.0574)	(0.1412)
	Bachelor's degree	1.1518***	1.3122***	0.9976	1.0775
		(0.0361)	(0.0612)	(0.0477)	(0.1101)
	Long-cycle higher education	1.2505***	1.3487***	1.1103*	1.4973***
		(0.0412)	(0.0667)	(0.0560)	(0.1483)
Educational field	Human sciences (ref.)				
	Arts	1.4175***	1.5778***	1.2621***	1.3824
		(0.0714)	(0.1364)	(0.0849)	(0.2802)
	Social Sciences	1.1590***	1.1941**	1.1427**	0.9581
		(0.0408)	(0.0660)	(0.0545)	(0.1603)
	Business economics, admin. and law	1.1407***	1.2235***	1.0223	1.2124
		(0.0388)	(0.0634)	(0.0493)	(0.1745)
	Science and engineering	1.1256***	1.0913	1.1209*	1.1745
		(0.0402)	(0.0614)	(0.0549)	(0.1789)
	Social and health	0.9605	1.0283	0.9547	1.0172
		(0.0362)	(0.0634)	(0.0484)	(0.1556)
	Teaching and learning	0.9787	1.0440	0.9794	0.6764
		(0.0504)	(0.0818)	(0.0704)	(0.2209)
	Other	1.2230***	1.1770**	1.2358***	1.1536
		(0.0460)	(0.0702)	(0.0638)	(0.1771)
Province	Capital province (ref.)				
	Bornholm	3.5960***	1.2409	6.3685***	7.6142***
		(0.9656)	(0.8789)	(2.1444)	(4.5261)
	South-west Zealand	2.0740***	2.0629***	2.3209***	3.0026***
		(0.1332)	(0.1728)	(0.2589)	(0.7682)
	Funen	2.8698***	2.8657***	2.7028***	3.8112***
		(0.0841)	(0.1299)	(0.1115)	(0.4269)
	Southern Jutland	2.7362***	2.5963***	3.5415***	3.1581***
		(0.1125)	(0.1439)	(0.2538)	(0.4127)
	Eastern Jutland	2.6414***	3.1231***	2.3244***	3.5549***
		(0.0558)	(0.1046)	(0.0669)	(0.2864)
	Western Jutland	3.3355***	3.3626***	3.2757***	4.5428***
		(0.1668)	(0.2380)	(0.2536)	(0.8996)
	Northern Jutland	2.9791***	2.9413***	3.1435***	2.5172***
		(0.0814)	(0.1187)	(0.1248)	(0.3010)

Continued

Table 3 (continued).

Covariates		Full sample Model 1	Studying at Home Model 2	Studying Away Model 3	Foreign Model 4
Gender	Female (ref.)				
	Male	1.0467*	1.0872**	1.0404	0.8921
		(0.0187)	(0.0301)	(0.0263)	(0.0607)
Immigration status	Danish (ref.)				
	First-generation immigrant	1.1556**	1.1573*	1.2628**	0.8362
		(0.0518)	(0.0791)	(0.0899)	(0.0878)
	Second-generation immigrant	1.1576*	1.0685	1.3862**	0.6714
		(0.0776)	(0.0898)	(0.1637)	(0.2596)
Age	Age at time of graduation	0.9569***	0.9570***	0.9556***	0.9493***
		(0.0024)	(0.0033)	(0.0043)	(0.0076)
Family type	Single (ref.)				
	Married	0.6777***	0.6848***	0.6407***	0.7455**
		(0.0218)	(0.0302)	(0.0343)	(0.0768)
	Cohabitants with children	0.5736***	0.5248***	0.6070***	0.7332
		(0.0307)	(0.0410)	(0.0480)	(0.1470)
	Cohabitants without children	0.8439***	0.7925***	0.8853***	0.8420*
Parental education	1: Neither parent holds a qualifying degree (ref.)				
	2: One or both parents hold a vocational education qualification	1.0354	1.0431	1.0209	1.2190
		(0.0342)	(0.0496)	(0.0486)	(0.2609)
	3: One or both parents hold a shorter- or medium-cycle higher education degree	1.0679*	1.1357**	0.9992	0.7832
		(0.0351)	(0.0540)	(0.0471)	(0.1702)
	4: One or both parents hold a long-cycle higher education degree	1.0672	1.0755	1.0066	1.2730
		(0.0393)	(0.0579)	(0.0531)	(0.2807)
Parental income	Ln (annual disposable parental income)	1.0105*	1.0033	1.0416***	0.9903
		(0.0045)	(0.0082)	(0.0103)	(0.0120)
	Dummy for richest 1%	0.7613**	0.8058	0.6931**	0.8575
		(0.0691)	(0.1025)	(0.0925)	(0.5082)
Origin	Studying at home (ref.)				
	Studying away	2.1792***			
		(0.0393)			
	Foreign	1.4671***			
		(0.0854)			
Number of obs.		41,495	23,712	14,084	3,699

Notes: *** p<0.001, ** p<0.01, * p<0.05. Standard errors in parentheses. Table estimates report the hazard ratios of moving out of the city where the HEI is located. Population: individuals who graduated from an HEI in 2010 and lived in the same city as their education institution at the time of graduation.

Figures 3 and 4 show the descriptive Kaplan Meier survival curves of graduate retention rates divided by origin for the Capital region and Other Regions respectively. Migration decisions are likely to be determined differently in the Copenhagen region than elsewhere in the country. The Copenhagen region is the only large metropolitan area in Denmark, and GDP per head is about 150% compared to the national average (see Chapter 4 in the present thesis). Therefore, we show Kaplan Meier Survival curves separately for sub-samples of graduates living within and outside the Copenhagen region respectively.

The differences are significant. Among local graduates who have studied in the region of Copenhagen, approximately 90 percent are still living in the region five years after graduation. Ten years after graduation about 80 percent of the graduates still remain in the Capital region. The corresponding numbers for the other regions are 70 percent after five years and 60 percent after 10 years, as illustrated in Figure 4.

The Copenhagen region is also able to keep its graduates even if they have not grown up in the region. Among the non-local graduates who studied in the Copenhagen region, about 80 percent still reside there five years after graduation and 60 percent 10 years after graduation. The corresponding figure for other regions is significantly lower; on average 30 percent remain in the region after 10 years. Thus, the contribution of local highly educated graduates to the local economy is much stronger in the Capital region compared to the other regions. The pattern is also confirmed at the city level (see Figures A2 and A3 in the appendix).

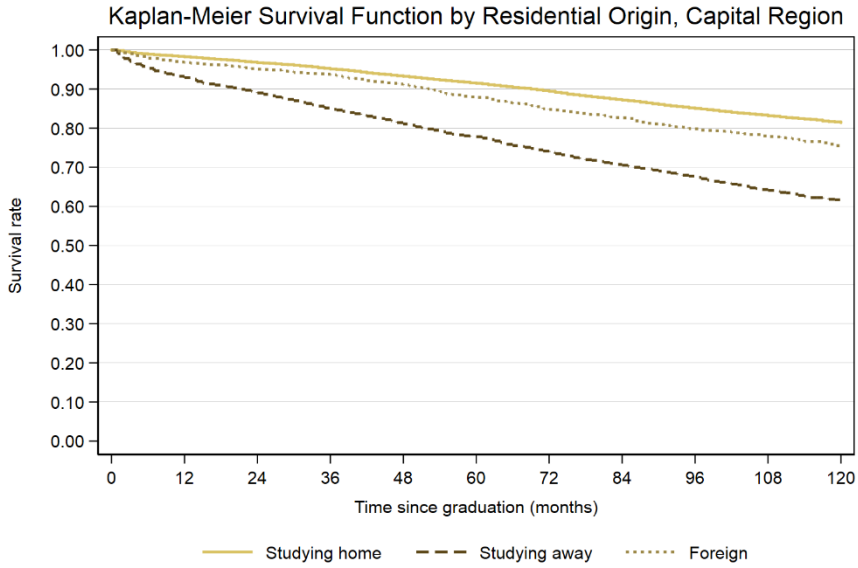

Figure 3. Kaplan-Meier Survival Functions for out-migration from the Capital region (NUTS 3). Divided by student origin.

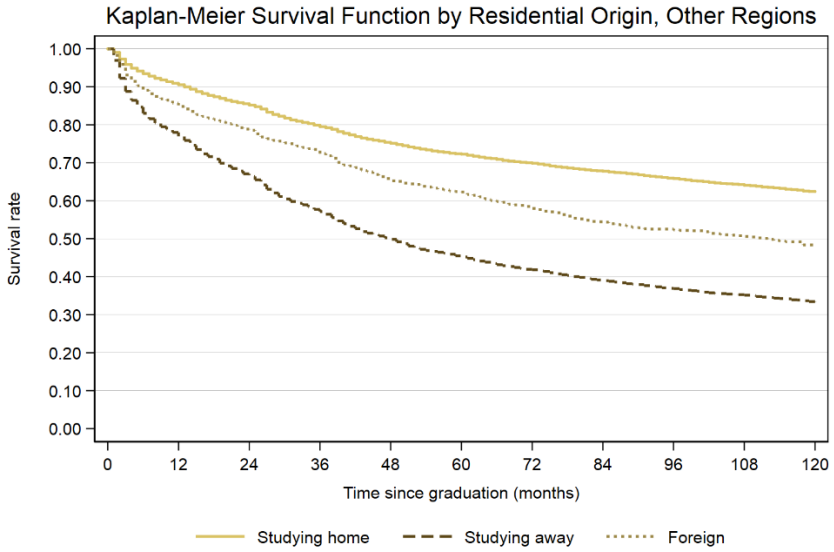

Figure 4. Kaplan-Meier Survival Functions for out-migration from other regions (NUTS 3). Divided by student origin.

4.3 Human Capital

Consistent with human capital migration theory (Sjaastad, 1962) and previous international studies (Haapanen & Bockerman, 2017; Haapanen & Karhunen, 2017), our regression results suggest that migration is associated with acquired human capital in the form of the length of the higher education. Figures 5 and 6 give graduate retention rates by educational level and educational field respectively.

Education level

Graduates from long-cycle higher education have on average a 1.25 times greater likelihood of moving out of the education region than graduates of short-cycle higher education (the reference group), as shown in Model 1 in Table 3. Cox-regression estimates also show that graduates with a bachelor's degree have a higher likelihood of moving than graduates in short-cycle education. The results are almost similar at the smaller geographical scale of the city. The results for graduate retention in education in the city are shown in Table A3 in the appendix. However, these results only apply to graduates who studied in the home region. The differences between the different education categories are relatively small compared to the differences between different city sizes, as illustrated in Figure 5. The descriptive Kaplan Meier curves in Figure 5 show that on average graduates with a long-cycle education are *less* migratory than their less educated peers. However, this descriptive association is not robust when controlling for HEI location and other covariates. This example underlines the importance of taking other explanatory variables into account when evaluating the influence of human capital on graduate migration.

Educational field

Educational field is not as strongly associated with mobility patterns as we expected. However, the regression results in Table 3 reveal some minor differences between educational disciplines that are worth highlighting. First, graduates of arts subjects have on average a greater likelihood of moving than their peers in human sciences (the reference group). Graduates in the arts, who have the greatest liability to move have a 1.42 times higher likelihood of moving than graduates in the reference group. The regression results also show that graduates in the social sciences, business economics and science and engineering have an approximately 1.14 times higher likelihood of moving compared to graduates in human sciences.

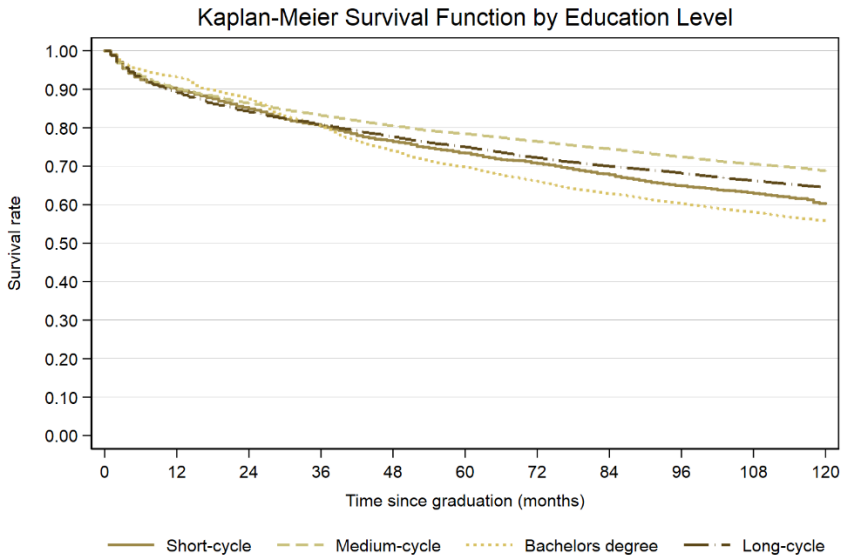

Figure 5. Kaplan-Meier Survival Functions for out-migration study region (NUTS 3). Divided by education level.

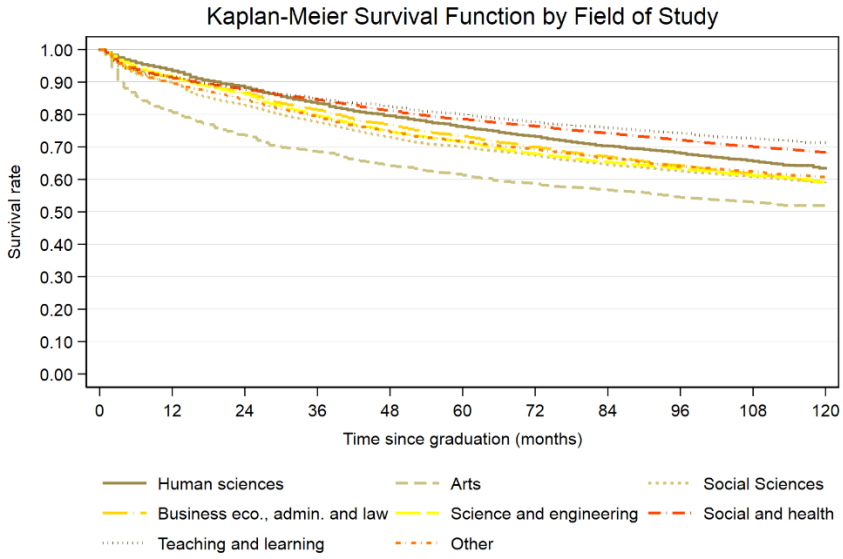

Figure 6. Kaplan-Meier Survival Functions for out-migration study region (NUTS 3). Divided by education field.

4.4 City and Region

The geographical location of the HEI is closely associated with graduate retention rates. We examine two different geographical scales of HEI locations: the local region at the NUTS 3 level (which to a large extent reflects the local labour market), and the narrower geographical scale of the local city. City size is interesting because it is associated with the local supply of amenities and agglomeration benefits (Glaeser and Gottlieb, 2006). Therefore, the numbers at the city level can help to shed new light on the association between urban amenities and graduate retention. However, if we were only looking at the cities, we would exclude the graduates who move outside the city boundary after graduation but stay in the same local labour market region and commute to work. The human capital of these graduates is therefore still available to the regional labour market.

City Size

We find significant differences in graduate migration across city sizes, as shown in Figure 7. The larger the city, the lower the probability of out-migration. The Kaplan-Meier survival curves show that the capital city, Copenhagen, differs significantly from the other cities: approximately 65 percent of the graduates are still resident in Copenhagen 10 years after graduation. In the smallest towns (1,000 to 19,000 inhabitants), about 22 percent of the graduates live in the city 10 years after graduation. The descriptive association is confirmed by the regression results from a Cox-proportional hazard model, where education level, education field, and individual and parental characteristics are taken into account. The parameter estimates are given in Table A3 in the appendix. The parameter estimates (from model 1) shows that students who reside in a town with fewer than 20,000 inhabitants are about 5.4 times more likely to move than students who lived in Copenhagen. Graduates who lived in medium-size towns with 20,000 to 49,999 inhabitants are about 3.8 times more likely to move out of the city, while graduates who lived in cities above 50,000 inhabitants are about 3.1 to 3.3 times more likely to move than graduates from Copenhagen. Thus, city size is closely associated with graduate retention rates.

Region

The regional location of the educational institution is also important for graduate migration. Figure 8 shows the Kaplan-Meier survival curves for each region and reveals that the differences between them are profound. About 75 percent of graduates in the Capital Region were still resident in the region 10 years after graduation, compared to only 45 percent in the Eastern Jutland region. The regression results from the Cox-regressions, shown in Table 3, confirms the descriptive results: the lowest likelihood of graduate out-migration is in the Capital Region (the reference group). The highest likelihood of out-migration is found on the island of Bornholm (HR = 3.5) and the regions of Western Jutland, Eastern Jutland and Funen, with hazard ratios of around 2.6 to 2.7 (see Model 1 in Table 3). The three regions with the highest levels of out-migration (except the island of Bornholm) are those where the second, third and fourth largest cities are located. These regions are characterised by thriving economies and large local labour markets. In contrast, the region of South-West Zealand is struggling with challenges regarding regional growth, labour, unemployment and low rates of youth educational achievement. Thus, it is surprising that out-migration from South-Western Zealand is lower than in the other regions. This finding might be explained by the fact that Southern and Western Zealand are located within commuting distance of the large labour market in the region surrounding Copenhagen. In sum, the results suggest that both urban amenities and access to a large labour market are important for graduate retention.

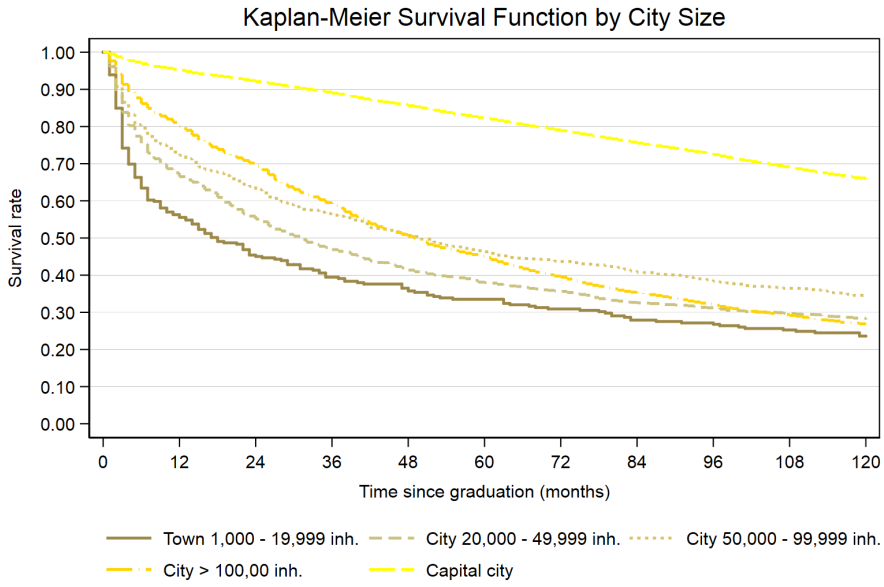

Figure 7. Kaplan-Meier Survival Functions for out-migration of education city. Divided by city size.

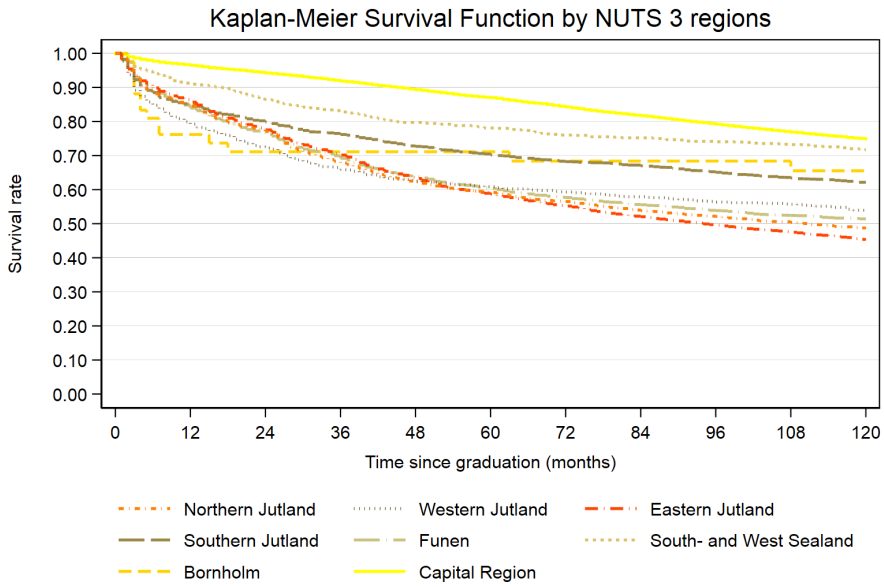

Figure 8. Kaplan-Meier Survival Functions of graduate retention by region of education (NUTS3). Divided by region.

4.5 Individual and Parental Characteristics

Finally, we consider the importance of other factors in our model, such as age, immigration status, household characteristics and family background. Most of our findings are in line with expectations and previous empirical studies. The results show that age is important for mobility patterns. The Hazard Ratio of age is below one in all models, indicating that the probability of migration decreases with age. This result is not surprising and is also found in previous empirical studies of migration (see among others (Haussen & Uebelmesser, 2018). Our result is robust to the different regional definitions of migration out of both educational city and educational region.

With regard to gender and immigration status, the results are mixed: the regional models suggest that males are more migratory than females (Table 3). However, if we measure migration at the city level instead, we find no statistically significant differences between genders (see Table A3 in the appendix). The results regarding 'immigration status' also differ according to which geographical definition we use. If we look at retention rates within the region, first- and second-generation immigrants have a slightly higher risk of moving compared to their native Danish peers. If we look at retention within the city, the numbers suggest the exact opposite association (see Table A3 in the appendix).

Household characteristics, that is, whether one has children and is married or single, are highly correlated with the likelihood of migration. It is not surprising that couples with children and married couples have a much lower likelihood of migrating compared to the singles in the reference group because the emotional costs of moving are higher if one has a spouse and children (Haussen & Uebelmesser, 2018). Finally, we controlled for parental education level and annual income in the models. The results suggest that there are minor or no significant association between family background, parental income and graduate migration rates.

4.6 Sequential Migration Behaviour

From a regional policy perspective, it is essential to know whether or not student migration towards the largest cities has negative consequences for the sending region. This depends essentially on the rate of return migration, that is, if highly qualified graduates return to their regions of origin again and contribute their newly acquired human capital to them. Previous work has shown that highly skilled graduate migrants tend to concentrate in a number of large cities

(Faggian, Corcoran, & McCann, 2013; Haapanen & Tervo, 2012). To study the sequential migration of Danish graduates, we classify them according to their migration behaviour to and from the region of higher education. The five categories are based on a study by Faggian et al. (2007b), namely *university stayers, repeat migrants, return migrants, non-migrants,* and *late migrants.*

University stayers are students who moved away from their home region to study and remain in the HEI region after graduation. *Repeat migrants* are students who moved away to enter higher education and then move to another region after graduation. *Return migrants* comprises those students who moved to study and then return to their home region after graduation. *Non-migrants* are those students who study in their home region and stay after graduation. Finally, *late migrants* are those students who remain in their home region while they study and then move away to another region after graduation. The spatial patterns of the human capital stock and flow will depend on the share of graduates in each of these five categories of sequential migration.

Table 4 presents the sequential migration behaviour of Danish graduates. About one-third (Group 1+2+3) of students moved to study, and two-thirds of the students (Group 4+5) studied in their home region. Thus, Danish students are significantly less migratory between upper secondary education and higher education compared to the British university graduates (Faggian et al., 2007b). Faggian et al. (2007b) showed that over 70 percent of UK graduates moved away from their home region to study.

From the perspective of the sending region, the share of *return migrants* is important. Table 4 shows that, among those who moved away, about 20 percent are *return migrants* and about 20 percent are *repeat migrants*. The residual 60 percent are *university stayers,* that is, those who stick to the region where they graduated 10 years after graduation. The numbers suggest that the majority of the highly educated graduates who have migrated to the region to study tend to stay and contribute to the local labour market. However, there are large regional differences, as shown in Figure A5 in the appendix. The Capital Region retains about 60 percent of their in-migrants compared to the peripheral region of Northern Jutland, which only retains about 20 percent of its in-migrants.

We find that those students who attended an HEI in their home region are significantly less migratory: only about 12 percent moved on to another region within 10 years of graduation, while the residual 88 percent stayed in the region. There are also large geographical differences in the retention rates of graduates

who studied in the region in which they are domiciled. However, Figure A4 in the appendix shows that graduate retention rates were above 50 percent in all regions.

In sum, the analysis of sequential migration underlines the fact that the Copenhagen region plays a pivotal role in sequential migration, as it has the largest number of students (above 50 percent as shown in Table 2) and the largest number of HEIs. In addition, it has the largest urban labour market in Denmark and provides work for the largest number of graduates who have migrated there to study and have stayed there on a permanent basis. Hence, it is crucial to differentiate between Copenhagen and the other regions when assessing the determinants and consequences of graduate migration.

Table 4
Sequential migration by graduates.

Group		Frequencies	Percentage within group	Percentage of total
Studying away	University stayers	7,196	60.60	22.07
	Repeat migrants	2,287	19.26	7.01
	Return migrants	2,392	20.14	7.34
	Sum	11,875	100.00	
Studying home	Non-migrants	18,317	88.36	56.18
	Late migrants	2,412	11.64	7.40
	Sum	20,729	100.00	
Total		32,604		100.00

5. Summary and conclusion

This study has examined regional graduate retention trajectories among highly educated individuals in Denmark. The importance of human capital for regional development has long been recognized in the literature (Gennaioli et al., 2013; Lucas, 1988; Romer, 1986). HEIs may play a key role in the creation and attraction of local human capital (Drucker & Goldstein, 2007). However, the local gains from HEIs depends on regional graduate retention rates. Previous literature has shown that university graduates are particularly mobile (Faggian et al., 2015, 2007b; Faggian & McCann, 2009), making it likely that they will 'leak' out of the region after graduation. Less is known about how migration propensities change over time. This is important because the longer a graduate stays in the region, the longer he or she will contribute to the local economy.

This study is the first in Denmark to analyse the length of time that graduates stay in their study area after graduation. We have analysed the graduate retention trajectories of a full cohort of Danish graduates from 2010 to 2019. The duration analysis demonstrated that the longer a graduate stays in the region of study, the lower is the propensity to move. The probability of moving is highest in the first two years after graduation, then the likelihood rate drops, suggesting strong duration dependence.

Our study also provides evidence of the importance of different determinants of migration, individual, study-related and regional. The results suggest that individual and study-related determinants have a relatively modest influence on graduate migration, while the regional location of the HEI is crucial for the ability to retain graduates in the local area. The most important individual determinant is student origin: our regression results show that students who moved to study are twice as likely to move again after graduation, compared to students who studied in their home region. Thus, our results are consistent with DaVanzo's (1983) hypothesis that subsequent migration is associated with previous migration, which is also confirmed by previous studies from Finland, Great Britain and Sweden (Faggian & McCann, 2009; Haapanen & Tervo, 2012; Rehák & Eriksson, 2020).

We have assessed two study-related factors: educational level and field of study. We compared graduate retention rates between short-cycle, medium-cycle and long-cycle higher education courses. Consistent with human capital migration theory (Sjaastad, 1962), the results suggest that migration is associated with acquired human capital: graduates of long-cycle higher

education are 1.25 times on average more likely to migrate than graduates of short-cycle higher education. The effect size is greater for students who studied in their home region than for students who studied away from home. The field of study also plays a role in migration: graduates in the arts are the most migratory, compared to graduates of human sciences, social and health, and teaching, who are the least migratory.

Although individual and study-related factors are important for graduate migration rates in Denmark, our results suggest that the regional location of the HEI is the most important factor. We find large differences in graduate retention rates depending on the size of the city and regional labour markets: in small towns with 1,000 to 19,999 inhabitants, graduates are on average 5.5 times more likely to move on than graduates in the capital city, Copenhagen, after controlling for individual and study-related covariates. We see similar results when we compare retention rates at the level of NUTS 3 regions: graduates from other regions are approximately two to three times more likely to move than graduates from the Capital region.

From a regional policy perspective, understanding the retention of highly educated graduates is important in Denmark due to plans to decentralize HEIs to more places. The expectation is that local access to an HEI in the student's home region could hinder out-migration and increase the local stock of human capital. Hence, HEIs in peripheral areas are seen as a tool contributing to more equal regional development.

Our analysis provides important new knowledge about the out-migration of college and university graduates in Denmark. The results suggest that graduates are highly mobile geographically, though migration rates differ considerably between regions and city sizes. Out-migration is much higher among graduates of more peripheral HEIs than of HEIs in the largest urban areas, Copenhagen in particular. We can see that the majority of graduates have moved from their regions or cities of studies within ten years of graduation, with the exception of Copenhagen. Hence, their acquired human capital will only have a minor effect on the city of their studies if the HEI is located in a smaller town. This suggests that the creation of new higher education institutions in peripheral areas will not change the unequal geographical distribution of human capital.

References

Ahlin, L., Andersson, M., & Thulin, P. (2018). Human capital sorting: The "when" and "who" of the sorting of educated workers to urban regions. *Journal of Regional Science*, *58*(3), 581–610.

Amendola, A., Barra, C., & Zotti, R. (2020). Does graduate human capital production increase local economic development? An instrumental variable approach. *Journal of Regional Science*, *60*(5), 959–994.

Busch, O., & Weigert, B. (2010). Where have all the graduates gone? Internal cross-state migration of graduates in Germany 1984-2004. *Annal of Regional Science*, *44*, 559–572.

Chatagnier, S., & Stillwell, J. (2021). Scale and zonation effects on internal migration indicators in the United Kingdom. *Population, Space and Place*, *n/a*(n/a), e2455. doi: https://doi.org/10.1002/psp.2455

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*(2), 187–220.

DaVanzo, J. (1982). Techniques for analysis of migration-history data. *RAND Note N-1824-AID/NICHD*. Retrieved from http://www.popline.org/node/388135

DaVanzo, J. (1983). Repeat Migration in the United States: Who Moves Back and Who Moves On? *Review of Economics & Statistics*, *65*(4), 552.

Drucker, J., & Goldstein, H. (2007). Assessing the Regional Economic Development Impacts of Universities: A Review of Current Approaches. *International Regional Science Review*, *30*(1), 20–46.

Elvery, J. A. (2010). City size and skill intensity. *Regional Science and Urban Economics*, *40*(6), 367–379.

Faggian, A., Corcocan, J., & Partridge, M. (2015). 22 Interregional migration analysis. In C. Karlsson, M. Andersson, & T. Norman (Eds.), *Handbook of Research Methods and Applications in Economic Geography*. Cheltenham, UK: Edward Elgar Publishing Limited.

Faggian, A., Corcoran, J., & McCann, P. (2013). Modelling geographical graduate job search using circular statistics. *Papers in Regional Science*, *92*(2), 329–343.

Faggian, A., & Mccann, P. (2009). Universities, Agglomerations and Graduate Human Capital Mobility. *Tijdschrift Voor Economische En Sociale Geografie*, *100*(2), 210–223.

Faggian, A., McCann, P., & Sheppard, S. (2006). An analysis of ethnic differences in UK graduate migration behaviour. *The Annals of Regional Science*, *40*(2), 461–471.

Faggian, A., McCann, P., & Sheppard, S. (2007a). Human Capital, Higher Education and Graduate Migration: An Analysis of Scottish and Welsh Students. *Urban Studies*, *44*(13), 2511–2528.

Faggian, A., McCann, P., & Sheppard, S. (2007b). Some evidence that women are more mobile than men: Gender differences in UK graduate migration behavior. *Journal of Regional Science*, *47*(3), 517–539.

Faggian, A., Rajbhandari, I., & Dotzel, K. R. (2017). The interregional migration of human capital and its regional consequences: A review. *Regional Studies*, *51*(1), 128–143.

Florida, R. (2002). *The Rise of The Creative Class. And How It's Transforming Work, Leisure, Community and Everyday Life*. Basic Books.

Florida, R. (2003). Cities and the Creative Class. *City & Community*, *2*(1), 3–19.

Flynn, R. (2012). Survival analysis. *Journal of Clinical Nursing*, *21*(19pt20), 2789–2797.

Gennaioli, N., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2013). Human Capital and Regional Development. *The Quarterly Journal of Economics*, *128*(1), 105–164.

Glaeser, E. (2011). *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier*. New York: Penguin Press.

Glaeser, E. L., & Gottlieb, J. D. (2006). Urban Resurgence and the Consumer City. *Urban Studies*, *43*(8), 1275–1299.

Glaeser, E. L., Scheinkman, J., & Shleifer, A. (1995). Economic-growth in a Cross-section of Cities. *Journal of Monetary Economics*, *36*(1), 117–143.

Haapanen, M., & Bockerman, P. (2017). More educated, more mobile? Evidence from post-secondary education reform. *Spatial Economic Analysis*, *12*(1), 8–26.

Haapanen, M., & Karhunen, H. (2017). Working while studying: Does it lead to greater attachment to the regional labour market? In J. Corcoran & A. Faggian (Eds.), *Graduate Migration and Regional Development: An International Perspective*. Cheltenham, UK: Edward Elgar Publishing.

Haapanen, M., & Tervo, H. (2012). Migration of the Highly Educated: Evidence from Residence Spells of University Graduates. *Journal of Regional Science*, *52*(4), 587–605.

Hansen, H. K., & Niedomysl, T. (2009). Migration of the creative class: Evidence from Sweden. *Journal of Economic Geography*, *9*(2), 191–206.

Hansen, H. K., & Winther, L. (2015). Employment growth, human capital and educational levels: Uneven urban and regional development in Denmark 2002–2012. *Geografisk Tidsskrift-Danish Journal of Geography*, *115*(2), 105–118.

Hansen, K. E. H., & Winther, L. (2018). Employment growth in Danish towns and regions since the crisis: Industrial structure, city size and location, 2008–2013. *Geografiska Annaler: Series B, Human Geography*, *100*(3), 244–262.

Haussen, T., & Uebelmesser, S. (2018). No Place Like Home? Graduate Migration in Germany. *Growth and Change*, *49*(3), 442–472.

Jans, A.-C. (2005). *Family relations, children and interregional mobility, 1970 to 2000* (Arbetsrapport No. 2005:12). Institute for Futures Studies. Retrieved from Institute for Futures Studies website: http://econpapers.repec.org/paper/hhsifswps/2005_5f012.htm

Jenkins, S. P. (1995). Practitioner's corner. Easy Estimation Methods for Discrete-Time Duration Models. *Oxford Bulletin of Economics and Statistics*, *57*(1).

Jenkins, S. P. (2004). *Survival Analysis*. Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK.

Kildegaard, K., & Domino, S. (2019, October 8). Mette Frederiksen i første store interview som statsminister: Vi ser et mere splittet land end før". *Berlingske*.

Krabel, S., & Flöther, C. (2014). Here Today, Gone Tomorrow? Regional Labour Mobility of German University Graduates. *Regional Studies*, *48*(10), 1609–1627.

Lucas, R. E. (1988). On the Mechanics of Economic Development. *Journal of Monetary Economics*, *22*(1), 3–42.

Malamud, O., & Wozniak, A. (2012). The Impact of College on Migration Evidence from the Vietnam Generation. *Journal of Human Resources*, *47*(4), 913–950.

Ministry of Higher Education and Science. (2020, October 12). Nye muligheder for uddannelse i hele landet—Uddannelses- og Forskningsministeriet [Press release]. Retrieved March 10, 2021, from https://ufm.dk/aktuelt/pressemeddelelser/2020/nye-muligheder-for-uddannelse-i-hele-landet

Rehák, Š., & Eriksson, R. (2020). Migration of university graduates and structural aspects of regional higher education. *European Planning Studies*, *28*(10), 1941–1959.

Romer, P. M. (1986). Increasing Returns and Long-Run Growth. *The Journal of Political Economy*, *94*(5), 1002–1037.

Schwartz, A. (1976). Migration, Age, and Education. *Journal of Political Economy*, *84*(4), 701–719.

Simon, C. J., & Nardinelli, C. (2002). Human capital and the rise of American cities, 1900–1990. *Regional Science and Urban Economics*, *32*(1), 59–96.

Sjaastad, L. A. (1962). The Costs and Returns of Human Migration. *Journal of Political Economy*, *70*(5), 80–93.

Storper, M. (1997). *The regional world: Territorial development in a global economy*. New York ; Guilford Press.

Storper, M. (2010). Why Does a City Grow? Specialisation, Human Capital or Institutions? *Urban Studies*, *47*(10), 2027–2050.

Storper, M., & Venables, A. J. (2004). Buzz: Face-to-face contact and the urban economy. *Journal of Economic Geography*, *4*(4), 351–370.

Teichert, C., Niebuhr, A., Otto, A., & Rossen, A. (2020). Work experience and graduate migration: An event history analysis of German data. *Regional Studies*, *54*(10), 1413–1424.

Venhorst, V., Dijk, J. V., & Wissen, L. V. (2010). Do the Best Graduates Leave the Peripheral Areas of the Netherlands? *Tijdschrift Voor Economische En Sociale Geografie*, *101*(5), 521–537.

Venhorst, V., Dijk, J. V., & Wissen, L. V. (2011). An Analysis of Trends in Spatial Mobility of Dutch Graduates. *Spatial Economic Analysis*, *6*(1), 57–82.

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/

Whisler, R. L., Waldorf, B. S., Mulligan, G. F., & Plane, D. A. (2008). Quality of life and the migration of the college-educated: A life-course approach. *Growth and Change*, *39*(1), 58–94.

Appendix

Tables

Table A1
Data reduction.

Population	Observations	Reduction	Percentage of all graduates
All graduates from higher educations, 2010	48,536		100.00
Only graduates resident in Denmark	47,236	1,300	97.32
Only graduates between the age 20 - 50 years	46,873	363	96.57
Region population (N1)	41,495	5,378	85.49
City population (N2)	32,782	8,713	67.54

Table A2
Location of HE graduates in the four largest cities.

Location of HE graduates	Frequencies	Percent
City name		
Copenhagen	16,977	51.79
Aarhus	7,241	22.09
Odense	2,632	8.03
Aalborg	2,908	8.87
Other	3,024	9.22
Total	32,782	100.00

Table A3
Hazard Ratios of moving out of HEI city.

Cox-Proportional Hazard Rate Models for Moving out of Institutional City					
		Full sample	Studying at Home	Studying Away	Foreign
Covariate		**Model 1**	**Model 2**	**Model 3**	**Model 4**
Educaitonal level	Short-cycle higher education (ref.)				
	Medium-cycle higher education	1.0888*	1.0944	1.0822	1.0480
		(0.0389)	(0.0591)	(0.0579)	(0.1359)
	Bachelor's degree	0.9645	1.0145	0.9299	1.0145
		(0.0299)	(0.0461)	(0.0444)	(0.1027)
	Long-cycle higher education	1.0950**	1.1333**	1.0567	1.1846
		(0.0356)	(0.0541)	(0.0538)	(0.1185)
Edicational field	Human sciences (ref.)				
	Arts	1.2649***	1.1557	1.2315**	1.6232*
		(0.0609)	(0.0959)	(0.0792)	(0.3118)
	Social Sciences	1.0779*	1.0489	1.0966	1.1186
		(0.0371)	(0.0547)	(0.0522)	(0.1916)
	Business economics, admin. and law	1.0094	1.0260	0.9552	1.2191
		(0.0332)	(0.0497)	(0.0455)	(0.1817)
	Science and engineering	1.2002***	1.1622**	1.2107***	1.4393*
		(0.0402)	(0.0587)	(0.0572)	(0.2189)
	Social and health	0.9949	0.9938	1.0057	1.2548
		(0.0350)	(0.0544)	(0.0490)	(0.1928)
	Teaching and learning	1.1317**	1.0898	1.1799*	1.0544
		(0.0531)	(0.0743)	(0.0805)	(0.3317)
	Other	1.1967***	1.1150*	1.2772***	1.2369
		(0.0428)	(0.0607)	(0.0642)	(0.1951)
City size	Greater Copenhagen (ref.)				
	Town 1,000-19,999 inh.	5.4146***	4.1459***	6.7603***	10.8548***
		(0.3915)	(0.4460)	(0.7063)	(3.4889)
	City 20,000-49,999 inh.	3.8400***	3.4660***	3.9336***	5.6586***
		(0.1335)	(0.1718)	(0.2170)	(0.6670)
	City 50,000-99,999 inh.	3.3272***	2.9654***	4.2631***	4.6072***
		(0.1407)	(0.1624)	(0.3312)	(0.6667)
	City > 100,000 inh.	3.1422***	3.3974***	2.7997***	4.0494***
		(0.0567)	(0.0895)	(0.0733)	(0.3071)

Continued

Table A3 Continued

Cox-Proportional Hazard Rate Models for Moving out of Institutional City: Continued

		Full sample	Studying at Home	Studying Away	Foreign
Covariate		**Model 1**	**Model 2**	**Model 3**	**Model 4**
Gender	Female (ref.)				
	Male	0.9964	0.9799	1.0186	0.9506
		(0.0169)	(0.0248)	(0.0251)	(0.0636)
Immigration status	Danish (ref.)				
	First generation immigrant	0.9043*	0.8010**	1.0319	0.7859*
		(0.0412)	(0.0584)	(0.0759)	(0.0823)
	Second generation immigrant	0.8145**	0.7445***	1.1289	0.5038
		(0.0564)	(0.0637)	(0.1417)	(0.1947)
Age	Age (at the time of graduation)	0.9620***	0.9593***	0.9650***	0.9746**
		(0.0024)	(0.0033)	(0.0044)	(0.0078)
Family type	Single (ref.)				
	Married	0.8191***	0.7737***	0.9011*	0.8159
		(0.0272)	(0.0361)	(0.0476)	(0.0900)
	Cohabitants with children	0.8647**	0.7854***	0.9685	0.9114
		(0.0423)	(0.0543)	(0.0726)	(0.1693)
	Cohabitants without children	1.0207	1.0544*	1.0014	0.8912
		(0.0178)	(0.0269)	(0.0252)	(0.0703)
Parental education	1: None of the parents hold a qualifying degree (ref.)				
	2: One or both parents hold a vocational education	1.0823*	1.1221*	1.0301	1.1546
		(0.0349)	(0.0508)	(0.0486)	(0.2703)
	3: One or both parents hold a shorter- or	1.0535	1.0762	1.0072	1.0627
		(0.0337)	(0.0486)	(0.0471)	(0.2317)
	4: One or both parents hold a long-cycle higher education	0.9908	0.9545	0.9899	1.1256
		(0.0353)	(0.0482)	(0.0517)	(0.2653)
Parental income	Ln (annual disposable parental income)	1.0072	1.0106	1.0210*	0.9911
		(0.0044)	(0.0081)	(0.0095)	(0.0128)
	Dummy for richest 1 %	0.8047**	0.8620	0.7416*	0.4955
		(0.0659)	(0.0959)	(0.0915)	(0.3563)
Origin	Studying at home (ref.)				
	Studying away	1.5182***			
		(0.0254)			
	Foreign	1.2107***			
		(0.0694)			
Number of obs.		32,782	17,310	12,386	3,086

Notes: *** p<0.001, ** p<0.01, * p<0.05. Standard errors in parentheses. Table estimates report the hazard ratios of moving out of the city where the HEI is located. Population: individuals who graduated from higher education in 2010 and lived in the same city as their HEI at the time of graduation.

Figures

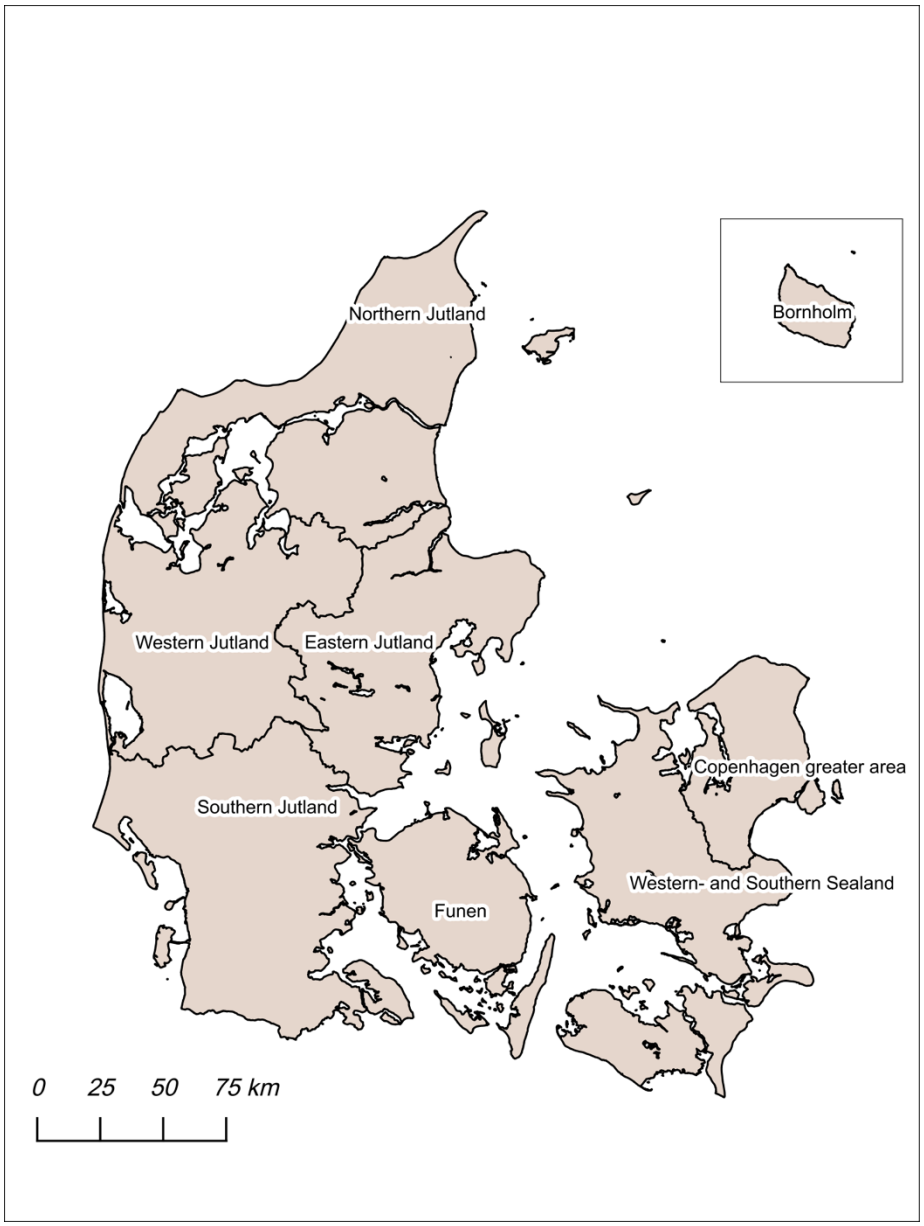

Figure A1. NUTS 3 regions in Denmark

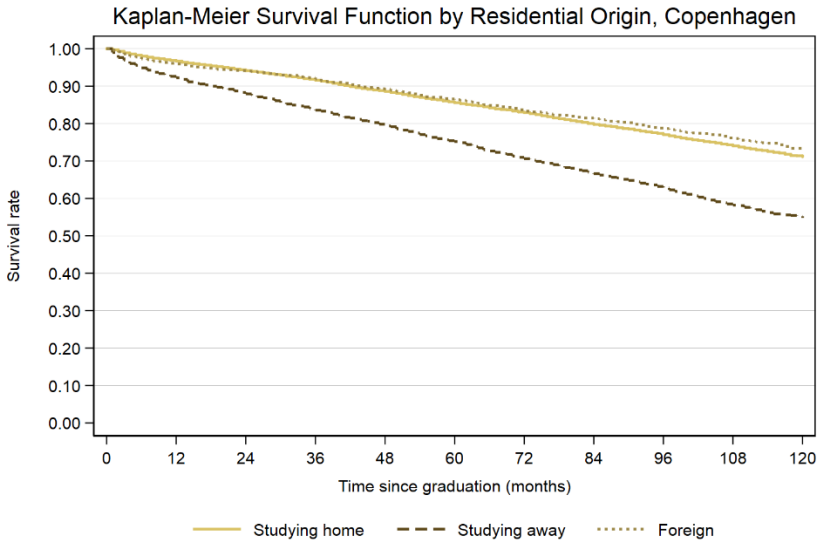

Figure A2. Kaplan-Meier Survival Functions for out-migration of Copenhagen city region. Divided by student origin.

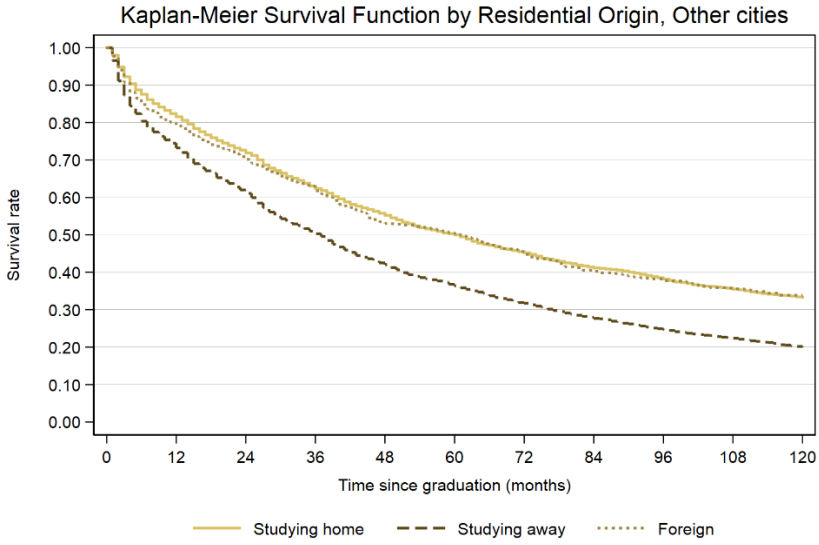

Figure A3. Kaplan-Meier Survival Functions for out-migration of cities and towns, excluding Copenhagen. Divided by student origin.

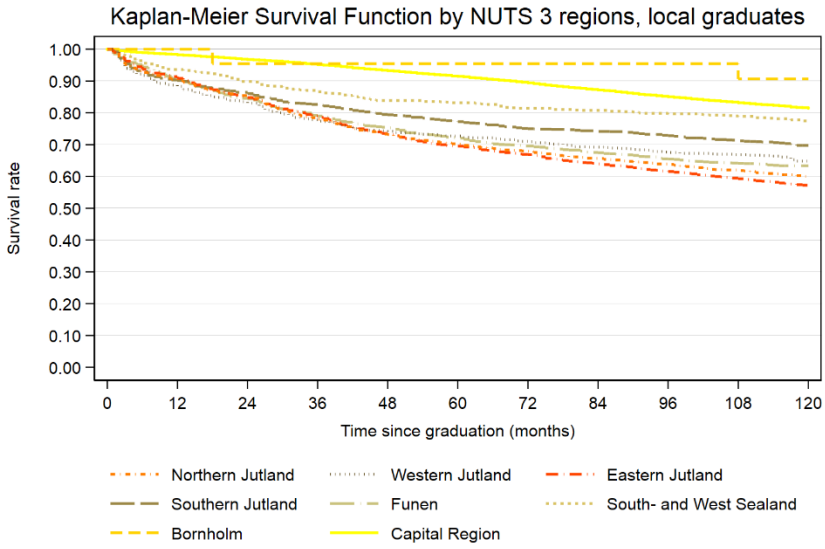

Figure A4. Kaplan-Meier Survival Functions for out-migration among graduates who studied in their home region. Divided by region.

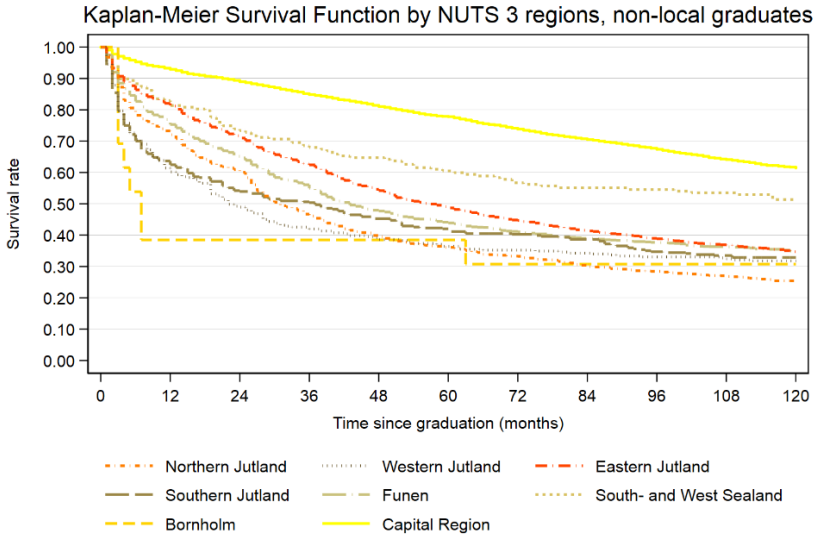

Figure A5. Kaplan-Meier Survival Functions for out-migration among graduates who moved to study in another region. Divided by region.

As regional inequality in economic prosperity has increased in Denmark in recent years, the political tools for slowing down this development are taking up space in public debates. The geographical location of public institutions, in particular educational institutions, has become a political hot topic in the recent years in Denmark. The underlying questions are to what extent the geographical localisation of educational institutions is one of the driving forces behind the regional divide, and whether the relocation of educational institutions has the potential to reduce the regional divergence in economic growth? This thesis answers these questions by providing new evidence for how the location of educational institutions in Denmark has influenced educational choices and inter-regional migration patterns in the country.

This PhD thesis has been submittet to the PhD School of
The Faculty of Science, University of Copenhagen
April 2021