JYVÄSKYLÄ STUDIES IN BUSINESS AND ECONOMICS
138

Tuomo Suhonen

Studies on Higher Education Choices and Spatial Labour Markets

JYVÄSKYLÄN YLIOPISTO

JYVÄSKYLÄ STUDIES IN BUSINESS AND ECONOMICS 138

Tuomo Suhonen

Studies on Higher Education Choices and Spatial Labour Markets

Esitetään Jyväskylän yliopiston kauppakorkeakoulun suostumuksella
julkisesti tarkastettavaksi yliopiston vanhassa juhlasalissa S212
marraskuun 22. päivänä 2013 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Jyväskylä University School of Business and Economics, in Auditorium S212,
on November 22, 2013 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2013

Studies on Higher Education Choices and Spatial Labour Markets

JYVÄSKYLÄ STUDIES IN BUSINESS AND ECONOMICS 138

Tuomo Suhonen

Studies on Higher Education Choices and Spatial Labour Markets

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2013

Editors
Tuomo Takala
Jyväskylä University School of Business and Economics
Pekka Olsbo, Sini Tuikka
Publishing Unit, University Library of Jyväskylä

URN:ISBN:978-951-39-5463-5
ISBN 978-951-39-5463-5 (PDF)

ISBN 978-951-39-5462-8 (nid.)
ISSN 1457-1986

Jyväskylä University Printing House, Jyväskylä 2013

ABSTRACT

Suhonen, Tuomo
Studies on Higher Education Choices and Spatial Labour Markets
Jyväskylä: University of Jyväskylä, 2013, 158 p.
Jyväskylä Studies in Business and Economics
ISSN 1457-1986; 138)
ISBN 978-951-39-5462-8 (nid.)
ISBN 978-951-39-5463-5 (PDF)
Finnish summary
Diss.

This thesis examines the mechanisms of individuals' higher education choices and economic returns to these choices after graduation from higher education. In the four empirical articles of this thesis, these topics are studied using administrative data on students and institutions within Finland's university system with a particular focus on a student's geographical location as a potential determinant of his or her choices and subsequent labour market outcomes.

The first two articles focus on the mechanisms of students' field-of-study choices within Finland's university system. The first article examines whether geographical distances affect these choices given that the field-of-study options vary considerably across regions in Finland. The results indicate that, on average, a longer distance from a student's region of origin to a field is negatively associated with his or her likelihood to choose that field. However, the distances do not appear to affect a student's decision to study education, arts or medicine. The second article examines to what extent the fields are transmitted from parents to children. The results show that considerable intergenerational transmission occurs; however, the strength of this transmission appears to vary in many ways. In particular, the law field is transmitted more frequently than any other field. Furthermore, the evidence suggests that men imitate their parents' field-of-study choices more often than women and that both men and women imitate their father's choice more often than the choice of their mother.

The last two articles investigate the effect of university choice on students' subsequent earnings. The third article approaches this topic from a locational perspective: one might suspect that Finnish students would economically benefit from attending a university in the Helsinki metropolitan area rather than in one of the nine smaller university regions because Helsinki is able to offer, for example, better job opportunities and more selective universities. However, the obtained results suggest, despite these benefits, an average student does not earn more during his or her early career as a result of choosing a metro area university; the earnings of humanities graduates even appear to be negatively affected by this choice. In the fourth article, the association between students' early-career earnings and different measures of university quality – related to, for example, educational resources and selectivity – are examined. The results suggest that, in general, the relationship between institution quality and earnings is rather weak in Finland. However, an increase in educational resources is, on average, found to be positively associated with the earnings of women and graduates from the humanities.

Keywords: higher education, labour markets, geographical location, field of study, intergenerational transmission, university quality

Author's address Tuomo Suhonen
School of Business and Economics
University of Jyväskylä
tuomo.j.suhonen@jyu.fi

Supervisor Hannu Tervo
School of Business and Economics
University of Jyväskylä

Reviewers Roope Uusitalo
Helsinki Center of Economic Research (HECER)
Helsinki, Finland

Anders Stenberg
Swedish Institute for Social Research (SOFI)
Stockholm University
Stockholm, Sweden

Opponent Roope Uusitalo
Helsinki Center of Economic Research (HECER)
Helsinki, Finland

ACKNOWLEDGEMENTS

I conducted most of the research reported in the thesis at the Jyväskylä University School of Business and Economics (JSBE) between 2007 and 2013, while the final revisions took place during my visit at the Umeå School of Business and Economics (USBE), Sweden, in autumn 2013. I have been lucky and privileged to have had the opportunity to work as a full-time researcher for these past six years, which has been made possible by the generous funding provided by the Academy of Finland, University of Jyväskylä and Yrjö Jahnsson Foundation. For this financial support, I am very grateful.

Although the articles in this thesis are self-authored, many people have obviously had an important influence on them. First and foremost, I would like to thank my supervisor, Professor Hannu Tervo. In autumn 2006, Hannu began to supervise my Master's thesis, proposing that I would start to examine the relationship between university location and the return to education in Finland. Thus, he is the original mastermind behind one of the main themes of the research I have conducted so far. I owe him my gratitude for his valuable advice, for the funding, and for his extreme patience.

The thesis was pre-examined by Professor Roope Uusitalo and Docent Anders Stenberg. I am grateful for their comments and constructive criticism and I am honoured to have Professor Uusitalo as my opponent.

During the final months of my PhD project, I worked as a guest researcher at the Umeå University. I want to thank Professor Olle Westerlund for arranging me this opportunity and for being an excellent host during my visit. I am also grateful to Olle for his many helpful comments and suggestions regarding the thesis chapters.

I would also like to thank all of my present and ex-colleagues at the JSBE. In particular, I want to thank Dr Mika Haapanen for being always kind to offer his help when I encountered problems in my research and Professor Kari Heimonen for being a friendly and encouraging boss. My special thanks belong to my dear friends and colleagues Dr Terhi Maczulskij, Hannu Karhunen and Juho Jokinen with whom I have shared many memorable moments both at the workplace and at leisure.

Aside from economic research, music has been a dear hobby of mine during these years in Jyväskylä. Therefore, I want to thank my past and present band mates in Art Deco for all the shared musical experiments, gigs, rehearsals, and so forth.

Finally, I want to thank my parents, Taija and Veijo, and my brother Mika, along with his family, for all the love and support over the years.

Umeå, October 2013

Tuomo Suhonen

LIST OF ORIGINAL PUBLICATIONS

Article 1. pp. 33–63
Suhonen, T. "Field-of-study choice in higher education: does distance matter?" (in referee process).

Article 2. pp. 65–90
Suhonen, T. "Intergenerational transmission of field of study: evidence from the field-of-study choice of Finnish university students" (in referee process).

Article 3. pp. 91–127
Suhonen, T. (2013). "Are there returns from university location in a state-funded university system?" *Regional Science and Urban Economics* 43(3), 465–478.

Article 4. pp. 129–154
Suhonen, T. (2013). "Quality of higher education and earnings: evidence from Finland using field-of-study-level quality measures" *International Review of Applied Economics,* forthcoming.

CONTENTS

CHAPTER 1: INTRODUCTION

1 Background

This doctoral thesis examines individuals' educational choices at the stage of moving from secondary-level education to higher education, that is, from high school to university. The transition to university education can be viewed as a complex choice situation: students have to select their area of concentration and educational institution from a large number of heterogeneous alternatives, and these choices have potentially large consequences for the students' later lives. In particular, the choice of field of study in higher education is known to be an influential determinant of students' future occupations and earnings (see Arcidiacono, 2004; Beffy et al., 2012), and some studies have also found significant returns from the choice of higher education institution (e.g., Behrman et al., 1996; Hoekstra, 2009). In many countries, supply-side restrictions add further complexity to the postsecondary educational choice. Namely, the supply of student places within a university system is usually highly regulated, which generates fierce competition in university admission. Furthermore, the supply of university education is often geographically restricted to relatively few locations, including mainly large and middle-sized cities, and partly for this reason, a considerable amount of interregional student migration is involved in the university enrolment process. In this sense, there is a significant 'spatial aspect' in the postsecondary educational choice, which distinguishes it from previous educational choices. This spatial aspect also creates challenges for policy makers who are responsible for ensuring that the university system promotes both efficient allocation of resources and equality of opportunities: if universities and different study programs are too sparsely scattered across space, costs involved in human migration could create barriers to individuals' choices, whereas an overly scattered system, consisting of small faculties in heterogeneous locations, could result in poor educational and research performance.

Since the emergence of the human capital theory, furthered by the works of Schultz (1961), Becker (1962) and Mincer (1974), economists have usually approached an individual's choice to acquire education from an investment decision perspective: individuals are viewed as rational decision makers who decide to invest in a particular educational program if the expected gain from this program is higher than that from other educational options (including the option of not studying at all). Furthermore, when defining the expected gain – or the 'return to education' – economists usually focus on the pecuniary aspects of the investment decision: the benefits of an educational program involve a potential premium in life-time earnings arising from the completion of the program, whereas the costs include both direct costs (tuition fees, study materials, transportation and living costs, etc.) and opportunity costs such as earnings that are forgone during studies (see Checchi, 2006, pp. 21–22). This kind of cost-benefit perspective is also in the core of the current thesis. More precisely, the primary focus is on questions related to the spatial aspects of the benefits and costs of higher education: Does a student's location before university affect his/her educational choices through migration costs? Do the financial benefits of university education depend on the location of the chosen university institution? These questions are studied using individual-level data from Finland on students' university and field-of-study choices within the Finnish university system. In this context, a focus on the role of location in higher education choices and their returns may be particularly relevant: given that Finland is a relatively vast and sparsely populated country, across which both educational opportunities and economic activities are unevenly distributed, one may suspect that a student's location before university, during studies and after graduation may significantly affect his or her educational and labour market outcomes. Consequently, the empirical results of this thesis may not only be of academic interest but can yield relevant information for policy makers responsible for the governance of the university system and for students making the choices.

This introductory chapter proceeds as follows. Section 2 briefly describes Finland's university system. Section 3 reviews previous literature on the mechanisms of higher education choices and economic returns to higher education and introduces a simple theoretical model to explain individuals' university and field-of-study choices in a spatial context. In addition to the introductory chapter, the thesis constitutes of four separate research articles included in Chapters 2–5; Section 4 provides a summary of these articles, including discussion on the topics, research questions, data, methods and main results.

2 University education in Finland

The empirical studies of this thesis examine young individuals' decisions regarding which university and field of study to enrol in within a specific historical and institutional context: the Finnish university system in the 1990s and early 2000s. The Finnish university system is a publicly financed and administered system responsible for the provision of academic research and education at the bachelor's, master's and PhD levels in Finland. Historically, this system is relatively young, as it was largely formed during the 20th century. Until 1908, University of Helsinki was officially Finland's only university, but over the century, the university system experienced a considerable geographical expansion, concluding with the establishment of University of Lapland in Rovaniemi in 1979.[1] In its current form, the university system comprises of 15 university institutions whose main campuses are located in 10 city regions: the Helsinki metropolitan area, Turku, Tampere, Lappeenranta, Kuopio, Joensuu, Jyväskylä, Vaasa, Oulu and Rovaniemi.[2] In addition, smaller university consortiums, serving as the universities' branch campuses, are currently located in six other cities. During the past few years, the university system has undergone major reforms aimed at increasing the universities' autonomy and competitiveness (see Tirronen & Nokkala, 2009). In the process, the number of university institutions has also been reduced from 21 to 15 through university mergers that occurred in 2010 and 2013. As the present research focuses on university graduates of the 1990s and early 2000s, the effects of these reforms are not yet visible in the obtained results.

Guaranteeing an equal access to university education for the young, regardless of, for example, family income and region of residence, has been one of the key principles of the Finnish university system since its establishment. Low economic barriers to participation have been ensured by offering all degree-oriented education free of charge and by providing students with study grants, housing allowances and state-guaranteed loans. A sufficient regional accessibility has been again ensured by extending the university network into all parts of the country. A downside of the geographical expansion is that many of the regional university campuses are relatively small and located in small and remote regions, which can, arguably, cause certain problems. First, as certain economies of scale apparently exist in knowledge production, the small average unit size may undermine the overall performance of the university system (see, e.g., Wolszczak-Derlacz & Parteka, 2011). Second, the small size of many of the universities has in practice lead Ministry of Education to restrict the number of field-of-study options that these universities offer for their students, which may have, in turn, generated regional disparities in access to certain fields (see Chapter 2). Third, as the economic activity and attractiveness of the university cities varies considerably, one may suspect that students' educational, economic and other outcomes can significantly depend on which

1 For a more detailed historical review of the Finnish university system, see Ministry of Education (2005). Chapters 2–5 of this thesis also provide additional descriptive information on the university system.

2 In the mid-1990s, the geographical expansion of higher education continued with the establishment of the polytechnic sector, which comprises of 28 higher education institutions rewarding professionally oriented degrees. The empirical analyses of this thesis focus solely on the university sector, and the polytechnic students are therefore excluded from the studied samples (with the exception of Chapter 2).

university city they graduate from (see Chapter 4). One of the key purposes of this thesis – especially Chapters 2 and 4 – is to find out whether these potential disadvantages from the geographical dispersion of the university system have in fact realised.

3 Theoretical considerations and prior research

3.1 Mechanisms of higher education choices

During the past few decades, a significant amount of research has been devoted to the study of mechanisms behind individuals' educational choices. A key motivation for this type of research has been that, as education potentially enhances both economic growth and social mobility, barriers to individuals' educational attainment, arising either from the demand or supply side, may result in an inefficient utilisation of human resources and, thus, undermine the welfare of the society (see Checchi, 2006). Most of this research has been carried out at a rather general level, focusing on the determinants of years of schooling or enrolment in post-compulsory education, whereas relatively few studies have explicitly focused on the choices of university/college and field of study at the stage of entering higher education. Broadly speaking, the existing studies on the determinants of higher education choices can be divided into three categories: 1) studies examining family background effects, 2) studies examining the effects of supply-side factors such as tuition fees and geographical distances and 3) studies examining the effects of expected future outcomes such as earnings.

The research on family background effects on higher education choices can be seen as a part of a larger strand of literature examining the *intergenerational transmission of education*. A key hypothesis in this literature is that parents may have a significant impact on a child's educational outcomes not only through a simple genetic transmission of cognitive abilities and other traits ('nature') but also through different post-birth mechanisms ('nurture'); for instance, parents may enhance their offspring's human capital by making expenditures on his or her skills, learning, motivation and other characteristics (Haveman & Wolfe, 1995).[3] Thus far, family background effects on higher education choices have been primarily studied by U.S. sociologists and educational researchers who have been concerned with whether there are differences in college and field-of-study choices across students with different socioeconomic backgrounds (e.g., Alexander et al., 1987; Davis & Guppy, 1997; Leppel et al., 2001; Goyette & Mullen, 2006). As the institutions in the U.S. higher education system are heterogeneous in many aspects – e.g., cost,

[3] See Haveman and Wolfe (1995) and Björklund and Salvanes (2011) for comprehensive reviews of the findings regarding the intergenerational transmission of educational attainment.

selectivity, social prestige and study programs – such differences could be an important factor that maintains socioeconomic inequality in the American society (e.g., Wolniak et al., 2008). A general conclusion of these studies is that a low socioeconomic status is associated with a higher probability to enrol in less prestigious community colleges and two-year colleges versus prestigious colleges and four-year colleges; however, this association appears to be, to a large extent, accounted for by differences in students' academic achievement (Davis & Guppy, 1997). Certain connections between socioeconomic background and the choice of field or major subject have also been found. For instance, Davis and Guppy (1997) find that, within highly selective colleges, students with a favourable socioeconomic background have a higher probability of enrolling in more lucrative fields. Goyette and Mullen (2006) again find that students of a low socioeconomic status are relatively likely to choose vocational majors, whereas a high socioeconomic status is associated with a high likelihood of choosing 'arts and sciences' fields. Furthermore, the study by Leppel et al. (2001) provides evidence suggesting that men and women respond differently to changes in socioeconomic background: as for women, a low socioeconomic status is associated with a lower probability of enrolling in the business field, whereas, for men, the effect of socioeconomic class is opposite. Only a handful of papers appear to have provided evidence on family background effects from outside the U.S.; the findings of Oosterbeek and Webbink (1997) from the Netherlands suggest that students from high income families are unlikely to attend a technical field, whereas, in a recent study by Bratti (2006), no significant linkage is found between social class and the choice of undergraduate degree subject within the U.K. higher education system.

The empirical research on the effects of supply-side factors on higher education choices draws its hypotheses from human capital models à la Becker (1962) and Sjaastad (1962), suggesting that, because of direct costs of education and costs of migration, the level and quality of human capital investments may differ across students possessing similar aspirations and abilities if these students differ in financial endowments or reside in different locations. In particular, in many countries, a subject of constant debate is whether tuition fees charged by higher education institutions generate differences in students' opportunities. The location of higher education might be another important supply-side factor: because of direct, informational and psychic costs of migration (see, e.g., Sjaastad, 1962; Schwartz, 1973), a student's decisions regarding whether to participate in higher education and which higher education institution to attend might be affected by the distances between the student's home and the institutions. The effects of both tuition and distance on higher education choices have been examined, to some extent, in empirical studies (e.g., Ordovensky, 1995; Avery & Hoxby, 2004; Long, 2004; Sá et al., 2004; Frenette, 2006; Alm & Winters, 2009; Jepsen & Montgomery, 2009; Cooke & Boyle, 2011; Gibbons & Vignoles, 2012). These papers systematically find that both an increase in the tuition charged by a higher education institution and the distance to the institution are significantly and negatively associated with the likelihood of choosing that institution. However, the findings of Avery & Hoxby (2004) from the U.S. suggest that the college choices of high-aptitude students are not sensitive to distances. Moreover, Long (2004) finds that the effect of distance on U.S. students' college choices has weakened over time. The evidence regarding the effect of the distance to the nearest higher education institution on enrolment in higher education has been rather mixed across studies and countries, varying from strong effects found in Canada (Frenette, 2006) to weak or insignificant effects found in Britain (Gibbons & Vignoles, 2012).

The main idea in the last of the three study categories is to test whether individuals' choices actually respond to changes in future benefits from education as predicted by the human capital theory. In particular, the general observation that earnings vary considerably across graduates from different fields/major subjects has motivated several papers to investigate whether the students' field choices react to changes in the anticipated earnings from these choices (Berger, 1988; Montmarquette et al., 2002; Arcidiacono, 2004; Bourdarbat, 2008; Bourdarbat & Montmarquette, 2008; Beffy et al., 2012). The results of these studies have been slightly mixed: for instance, whereas the early study by Berger (1988) finds that U.S. college students' demands for different major subjects are significantly elastic to the expected life-time earnings in these majors, the more recent studies by Arcidiacono (2004) and Beffy et al. (2012) – from the U.S. and France, respectively – suggest that the weight of earnings in students' major choices is rather small compared to the weight of other (non-pecuniary) characteristics of the majors. Certain heterogeneity across genders in the effect of expected earnings has also been found: Montmarquette et al. (2002) find that, in the U.S., women are less influenced by earnings than men, whereas the results of Boudarbat and Montmarquette (2008) suggest that, in Canada, men are more sensitive to initial earnings than women and that women put more weight to the growth rate of earnings. Aside from expected earnings, studies have also considered some other expected future outcomes as potential determinants of higher education choices. For instance, Freeman and Hirsch (2008) find evidence that U.S. students' major choices respond to changes in the knowledge content of jobs (e.g., the degree of 'sales and marketing' or 'engineering and technology' involved in different jobs). Furthermore, their findings indicate that women are more sensitive than men to the knowledge content of jobs when choosing a major, whereas men are more influenced by the wage returns to the knowledge content. Studies have also acknowledged that institution quality may be another important factor behind individuals' choices; the findings of Long (2004) and Cooke and Boyle (2011) from the U.S. have indeed supported the view that, all else equal, students are more likely to choose a college or a college state when institution quality in that alternative increases.

3.2 Economic returns to higher education

Since the introduction of the Mincerian earnings equation (Mincer, 1974), an empirical model that, in its basic form, explains an individual's earnings by his or her years of schooling and work experience, the economic returns to education have been one of the most extensively studied research subjects in labour economics. In this area of research, a common finding – obtained by using various techniques to control for individual ability differences – is that acquiring more education has a positive causal effect on one's life-time earnings (e.g., Angrist & Krueger, 1991; Ashenfelter & Krueger, 1994; Ashenfelter & Rouse, 1998; Card, 2001). Over the past twenty years or so, researchers have also become increasingly interested in investigating heterogeneity in the return to higher education across students who graduated from different schools and fields of study. Theoretically, differences in the economic payoffs for different higher education choices may arise from various sources (see, e.g., Lang & Siniver, 2011). First, different schools and study programs may provide a student with a different amount and type of human capital, and the resulting productivity differences across students may be reflected in their labour market outcomes. Second, the signalling value of higher education – that is, the degree to which employers interpret one's education as a signal of high productivity in the labour market – may depend on which school and field the obtained degree is from. Third, the ability of a student to obtain beneficial social networks may also depend on the school and field. In addition to these 'traditional' explanations, one might think that part of the heterogeneity in labour market outcomes arises from differences in geographical location. For example, when schools and field-of-study options are located in economically heterogeneous regions, the monetary and non-monetary costs from relocation after graduation could generate earnings differences between students from different schools and fields (see Chapter 4).

A well-documented finding from many countries is that considerable earnings differences exist across fields of study. For instance, completing a university degree in business, law or engineering generally results in higher average earnings than a degree in education or the humanities (e.g., James et al., 1989; Loury & Garman, 1995; Arcidiacono, 2004; Boudarbat & Montmarquette, 2008; Wolniak et al., 2008; Beffy et al., 2012). Furthermore, these earnings differences appear not to be solely explained by ability differences across fields, suggesting that the choice of field indeed affects earnings through the causal mechanisms discussed above (Arcidiacono, 2004). Generally speaking, it is less obvious, on the basis of both theory and empirics, that the choice of higher education institution affects earnings (e.g., James et al., 1989; Arcidiacono, 2004). One of the most widely discussed reasons for suspecting that graduation from a highly selective or prestigious institution is economically beneficial arises from the hypothesis of peer effects; that is, being in the company of strong students during higher education could have a positive effect on one's human capital accumulation. There is indeed evidence, e.g., from the analyses of randomly assigned college roommates (Sacerdore, 2001; Zimmerman, 2003), suggesting that peer effects exist in higher education. Alternatively, one may suspect that the high-quality instruction offered in prestigious institutions could result in returns to institution choice (e.g., Lang & Siniver, 2011; Triventi & Trivellato, 2012). In addition, the reputation of the attended higher education institution is – at least in some countries such as the U.S. – known to be accounted for by employers when recruiting newly graduated individuals, because of which returns from the choice of institution could arise through the signalling channel (Hershbein, 2011; Lang & Siniver, 2011). The empirical findings of the effect of university/college choice on students' subsequent earnings have been, both in the U.S. and Europe, somewhat conflicting across studies. For instance, whereas Behrman et al. (1996) find significant returns from attending a selective college in the U.S. by using data on identical twin pairs, the results of Dale and Krueger (2002; 2011) that rely on a different estimation strategy (the use of controls for students' applications and admission records) suggest that college selectivity is not, in general, a significant earnings determinant.[4] However, overall, the evidence has been more supportive of the view that the choice of institution matters, at least in the U.S. labour market.

3.3 Theoretical model of higher education choices in a spatial context

One of the main purposes of this thesis is to examine higher education choices and their returns from a locational perspective. A key feature in this setting is the assumption that the interregional mobility of university students before and after university is more or less imperfect. A traditional explanation for immobility states that direct, informational and psychic costs involved in migration may prevent a person from moving to a new location even in the presence of economic gains from migration (e.g., Sjaastad, 1962; Schwartz, 1973). Furthermore, a common belief is that migration costs are an increasing function of the migration distance, because of which an inverse relationship is usually observed between the distance of two locations and the level of migration flows between these locations (Schwartz, 1973).

In this subsection, a simple theoretical model is derived to illustrate the impact of migration costs on the demands for postsecondary educational alternatives, when the alternatives are unevenly distributed across space. The model borrows features from both the human capital theory (Becker, 1962; Sjaastad, 1962) and Hotelling-type models of demand for differentiated products (see, e.g., Hotelling, 1929; Klemperer, 1987; Shy & Stenbacka, 2003). The model involves a simple higher education system, in which there are only two universities and two fields of study. The universities, labelled as A and B,

4 Other U.S. studies within this topic include, e.g., those by James et al., (1989), Brewer et al. (1999), Monks (2000), Black and Smith (2004; 2006), Long (2008) and Hoekstra (2009). In Europe, returns to the choice of higher education institution have been previously studied at least in the U.K. (Chevalier & Conlon, 2003), Italy (Brunello & Cappellari, 2008; Triventi & Trivellato, 2012) and Sweden (Lindahl & Regnér, 2005; Eliasson, 2006; Holmlund, 2009)

are assumed to be similar in all ways except for the fact that they are at different geographical locations. The distance between the universities is set to 1. The fields of study, labelled as C and D, are again assumed to be 'highly differentiated products', i.e., they are allowed to differ with respect to both expected life-time earnings and different non-pecuniary characteristics (the nature and content of studies, expected non-pecuniary job characteristics, etc).[5] In the higher education system, there are n students making decisions regarding which university and field to enrol in. The students are assumed to be similar in all but two aspects. First, the students live in different locations along the unit interval between A and B – that is, along a 'Hotelling's street' – and the geographical distance from a student's home to university i is $x_i \in [0,1]$. Second, the students are allowed to differ with respect to their preferences for non-pecuniary field characteristics, which is captured by the 'preference distance' $y_j \in [0,1]$ from a student to field j. For simplicity, it is assumed that the students' field preferences are evenly distributed across space; that is, x_i and y_j are assumed to have a bivariate uniform distribution.[6] The students choose their university-field alternative (i, j) by maximising utility U_{ij} across $i = A, B$ and $j = C, D$; this utility is assumed to be linear in distances x_i and y_j as follows:

$$U_{ij} = u(e_j) - \alpha x_i - \beta y_j, \qquad (1)$$

where e_j depicts the life-time earnings resulting from the choice of field j, and function $u(e)$ is increasing in e; α captures the migration cost per a unit of distance involved in moving to university i; and β depicts the marginal disutility involved in moving further away from the 'optimal' field.

The supply side of the model is kept very simplistic. Namely, the number of student places is not restricted for any of the available university or field alternatives, and every student is admitted to study in the place of his/her own choosing. This 'unlimited-admission' assumption can be thought as a natural consequence of the assumption that the students are similar in almost all aspects: defining admission criteria based on any observed attributes, such as test scores, is not possible in this higher education system. However, because the purpose of the model is to illustrate students' choices under an uneven spatial distribution of alternatives, a specific supply constraint is imposed:

---

5 The basic features of the model roughly correspond to the empirical findings discussed above. Namely, fields are generally known to be very heterogeneous with respect to both expected pecuniary and nonpecuniary future outcomes, and this heterogeneity has been observed to affect students' choices (e.g., Arcidiacono, 2004; Beffy et al., 2011). Then again, it is less obvious that the choice of university has any effect on future outcomes – especially in a relatively homogeneous, state-funded university system such as that in Finland – and therefore, location may be the key feature that a student considers when deciding whether to attend a particular university.

6 The uniform distribution is only assumed to make the interpretation of the figures more straightforward: with the uniformly distributed locations of students, areas drawn in (x_i, y_j)-space can be directly interpreted as proportions of the total number of students.

while field D may be studied in both universities, field C may only be studied in university A. Despite its simplicity, this kind of supply setting roughly resembles that of Finland's university system: some universities are more multidisciplinary than others, and some fields have been made available in fewer locations than others. Under the supply constraint, a student's choice set is $\Theta = \{(A, C), (A, D), (B, D)\}$, and knowing that $x_B = 1 - x_A$ and that $y_D = 1 - y_C$, the utilities for the choice alternatives are given by:

$$U_{A,C} = u(e_C) - \alpha x_A - \beta y_C \tag{2}$$

$$U_{A,D} = u(e_D) - \alpha x_A - \beta(1 - y_C) \tag{3}$$

$$U_{B,D} = u(e_D) - \alpha(1 - x_A) - \beta(1 - y_C) \tag{4}$$

Given that the students are uniformly distributed across the (x_A, y_C)-space, it is now straightforward to solve the proportions of students enrolling in different universities and fields by examining the conditions for utility maximisation. Let us first examine the choice of field: for students choosing field C, the utilities must satisfy $U_{A,C} \geq U_{A,D}$ and $U_{A,C} \geq U_{B,D}$. By substituting (2), (3) and (4) into these conditions and by solving with respect y_C, the following pair of inequalities is reached:

$$y_C \leq \frac{1}{2} + \frac{u(e_C) - u(e_D)}{2\beta} \tag{5}$$

$$y_C \leq \frac{1}{2} + \frac{u(e_C) - u(e_D)}{2\beta} + \frac{\alpha}{2\beta} - \frac{\alpha}{\beta} x_A \tag{6}$$

By looking at the conditions above, it is immediately clear that, without the existence of migration costs α, the second condition (6) collapses into the first one (5): this condition simply states that the deviation of the demand share for field C from ½ depends on the ratio between the pecuniary and non-pecuniary benefits arising from the choice of that field instead of field D. However, when the migration costs are in effect, some of the students residing within a certain distance from university A substitute their demand from C to D despite having a relatively low preference distance to C; the distance at which this substitution begins to occur is determined at the cross-section of the two curves implied by (5) and (6), that is, at x_A = ½. In Figure 1, the solution is illustrated in the case where field C is not only more restrictedly supplied but also more lucrative than D in the sense that it yields higher expected life-time earnings e for a student (that is, $e_C > e_D$). From the figure, we see that, in the case of no migration costs, a clearly larger proportion of students choose field C (area III+IV+V+VI) than field D (area I+II) because C is more lucrative. In this case, the 'excess' demand for field C, arising solely from the higher expected earnings in this field, is represented by area III+IV. The impact of migration costs is again depicted by area IV+VI: the students in this area have a relatively high preference for field C but are forced to choose D because of the costs involved in migrating to the more distant university A. The magnitude of the distance effect depends on ratio α/β; Figure 1 shows that in the case of $\alpha/\beta = 4/3$, the overall demand for field C reduces by 25% because of the migration costs.

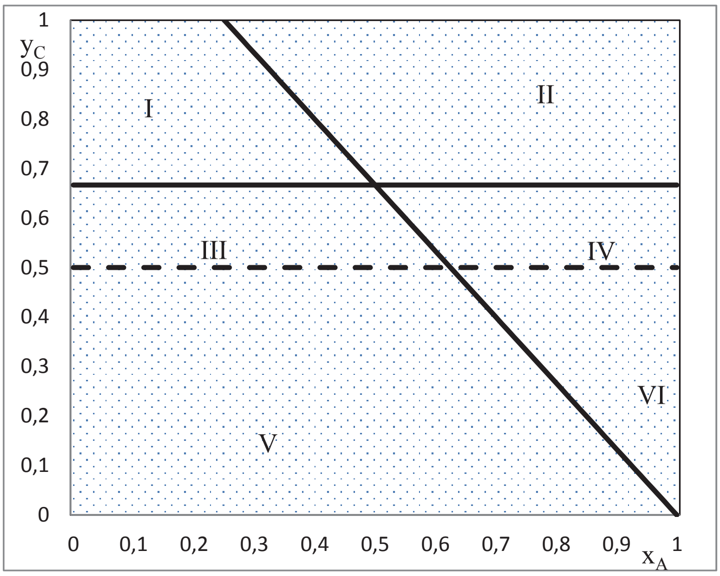


FIGURE 1 The optimal choice of field conditional on the geographical distance to university A (x_A) and the preference distance to field C (y_C). The graph was drawn using the following parameter values: $u(e_C) - u(e_D) = 0.5$, $\alpha = 2$ and $\beta = 1½$. Area III+V represents the share of students choosing field C, whereas the students in area I+II+IV+VI choose field D.

Obviously, the uneven geographical distribution of field-of-study options in the higher education system also affects the students' demands for the two university alternatives, as some of the students with relatively low preference distance to field C are willing to migrate to university A despite a relatively long geographical distance to this university. To illustrate this phenomenon, let us first note that, for a student choosing university A, the utilities must satisfy either $U_{A,C} \geq U_{B,D}$ or $U_{A,D} \geq U_{B,D}$. Thus, the set of students choosing university A in the (x_A, y_C)-space is determined by the following pair of inequalities:

$$y_C \leq \frac{1}{2} + \frac{u(e_C)-u(e_D)}{2\beta} + \frac{\alpha}{2\beta} - \frac{\alpha}{\beta}x_A \tag{7}$$

$$x_A \leq \frac{1}{2} \tag{8}$$

Of these conditions, (7) is the same as condition (6) for the optimal field choice. This condition has two notable implications. First, the more positive (negative) the ratio between the pecuniary and non-pecuniary benefits from choosing field C versus field D is, the larger (smaller) the impact of the restricted supply of field C is on the choice of university. Second, in the case of university choice, the migration costs serve as a vehicle of making the demand for the university alternatives more even: without these costs, all students satisfying $y_C \leq 1/2 + u(e_C) - u(e_D)/2\beta$ would choose to attend university A, whereas the migration costs induce some of the student residing relatively close to university B to substitute their demand towards this university. The second condition (8) again simply depicts the condition for the optimal university choice in the case of similar field-of-study options across the universities; that is, half of the students would choose university A, whereas the other half would choose university B. Figure 2 shows that, under similar conditions as in Figure 1, a relatively large proportion of students, represented by area I+II+IV, choose university A instead of B. Moreover, we see that this proportion becomes larger, as the diagonal line becomes less steep, that is, when the importance of migration costs relative to field preferences (ratio α/β) reduces.

In summary, the above-discussed simple theory model suggests that, when field-of-study options vary across universities, migration costs can affect students' choices regarding which university and field to enrol in, but only if these costs are important relative to the disutility that the students experience when having to choose non-preferred fields. This hypothesis serves as a working hypothesis for Chapter 2 of this thesis, which empirically examines the effect of distance on field-of-study choice. The idea that a geographically restricted supply of fields may affect students' choices is also revisited in Chapters 3, 4 and 5.

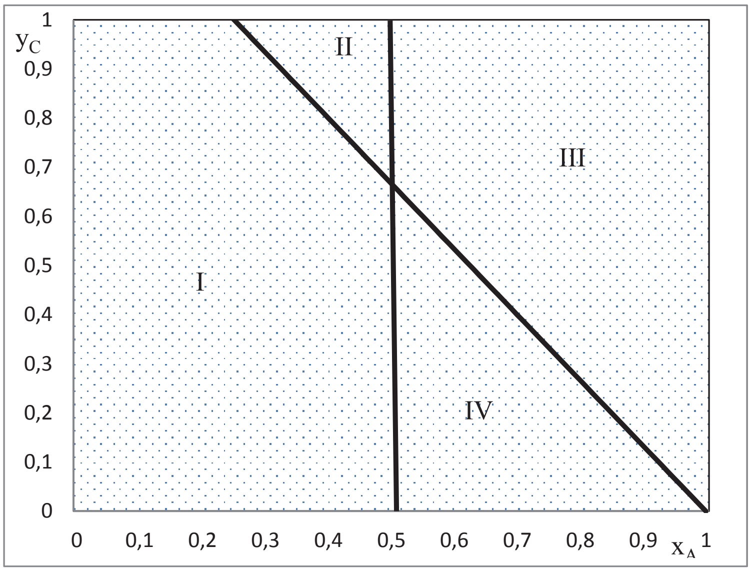


FIGURE 2 The optimal choice of university conditional on the geographical distance to university A (x_A) and the preference distance to field C (y_C). The graph was drawn using the same parameter values as in Figure 1. Area I+II+IV represents the share of students choosing university A, whereas the students in area III choose university B.

4 Outline of the thesis

4.1 Topics and research questions

The four research articles of this thesis, corresponding to Chapters 2–5, can be divided under two broad themes: Chapters 2 and 3 contribute to our understanding of the *mechanisms of field-of-study choice*, whereas Chapters 4 and 5 provide new evidence of the *economic returns to university choice*. Alternatively, the chapters can be divided according to how tightly they are connected to the main theme of the thesis, that is, the role of location in higher education choices and their returns; in this sense, Chapters 2 and 4 can be considered as the 'core' chapters, whereas Chapters 3 and 5 aim to widen our picture of the studied phenomena by focusing on topics that are not directly location-related. None of the covered topics have been studied previously using Finnish data. In addition, each chapter makes distinct contributions to the international literature by approaching its topic from a new perspective and/or with new type of data.

Chapter 2 contributes to the empirical literature investigating the effect of geographical distance on higher education choices. As discussed in Section 3.1, many previous studies have concluded that, in general, a longer distance to a higher education institution is associated with a lower likelihood to enrol in that institution. However, very few studies have thus far examined the effect of distance on the choice of an educational field. In many countries – including Finland – distances between university cities are generally long, and field-of-study options vary across universities; therefore, one may suspect that distances do not only restrict the choice of university but also the choice of field in these countries. This hypothesis is tested empirically in Chapter 2.

Chapter 3 investigates to what extent fields of study in university education are transmitted from parents to children. Although a significant amount of research has been conducted on both intergenerational transmission of educational attainment and the role of family background in higher education choices (see Section 3.1), there appears to be no previous studies examining directly the association between parents' and children's field-of-study choices in the context of university education.

Chapter 4 complements the existing evidence regarding the effect of university choice on earnings by approaching the topic from a locational perspective. The main working hypothesis in this chapter is that universities located in economically active, high-amenity city regions may not only offer their students high-quality university education but also other benefits such as good job opportunities both during studies and after graduation. Hence, students attending a university in a large and attractive city region might subsequently earn an earnings premium for their university location choice. This hypothesis is tested by studying the effect of graduating from a metro area university – that is, a university located in the Helsinki metropolitan area – on students' early-career earnings in Finland.[7] Furthermore, the chapter studies the 'metro area university premium' among individuals who resided in the same region after graduation; as a graduate's residential location choice may partly depend on where he or she attended university, there could be a significant indirect 'post-university region effect' arising from regional earnings differences included in the total earnings premium.

Chapter 5 continues the theme of Chapter 4, focusing on a narrower aspect of the returns to university choice: the effect of institution quality on earnings. Therefore, in this chapter, 'quality' is assumed to be, to some extent, measurable through observables related to, for example, the resources, reputation or selectivity of a university institution. The employed quality measures include the number of teachers per student, the number of publications per researcher and the number of applicants per admitted student. A contribution to the previous literature is made by allowing the quality measures to vary within universities across fields of study, which offers a way to reduce measurement error biases in quality effects estimates. The main question of interest is whether these field-of-study-level quality measures are significantly associated with the students' early-career earnings – through the mechanisms discussed in Section 3.2 – after controlling for the students' pre-university characteristics and 'fixed' differences arising from the choice of university region and field of study.

In summary, the main research questions of the chapters are:

- Chapter 2: Does the geographical distance to a field decrease a student's likelihood of choosing that field?
- Chapter 3: To what extent (if at all) is a student more likely to graduate from a particular field if he or she has a parent with a university degree from that field?
- Chapter 4: Does graduation from a metro area university affect a student's subsequent earnings? Is there a metro area university premium for students who reside in the same region after graduation?
- Chapter 5: Are different aspects of institution quality – e.g., educational resources and selectivity – associated with students' early-career earnings?

---

7 Naturally, it would also be interesting to analyse the earnings effects of university choice at the later stages of students' careers. In Chapters 4 and 5, the analysis is restricted to early-career earnings merely for practical reasons: since students' matriculation examination grades are only available from the mid-1980s onwards in the data, only those who graduated in relatively late years can be included in the study samples to enable the use of these grades as control variables. Almost all of the earlier studies within this topic have also been forced to use early-career outcomes. As an exception, Dale and Krueger (2011) were recently able to analyse earnings data that span more than two decades of the careers of an early cohort of U.S. college students. Their basic results relying on 'selection on observables' suggest that there are significant returns from college selectivity and that these returns even increase considerably over the course of a student's career. However, their results adjusting for students' unobservable characteristics mainly suggest that this return is close to zero and insignificant at all stages of the career; thus, the baseline estimates appear to be biased upwards.

Furthermore, each chapter examines whether there is heterogeneity in the effects of interest, for example, across genders, fields and regions. Given the above-discussed research questions, theory and empirics, the main framework of the thesis can be summarised using a simple picture in Figure 3. The picture highlights four important hypotheses that are either tested or accounted for in the analyses. First, a student's location before university (high school region) may affect his/her choices of university and field of study, for instance, because of migration costs. Second, the student's high school region may also affect the choice of post-university region both directly – for instance, through the mechanisms that induce return migration – and indirectly through the choices of university and field of study. Third, high school region, university and field of study may have both direct effects on labour market outcomes after graduation – for instance, because of differences in human capital, signalling and network effects of education across locations and fields – and indirect effects through the subsequent choices. The fourth and final hypothesis is that a student's family of origin does not only determine his/her location before university but may also significantly affect his/her subsequent locational and educational choices and outcomes.

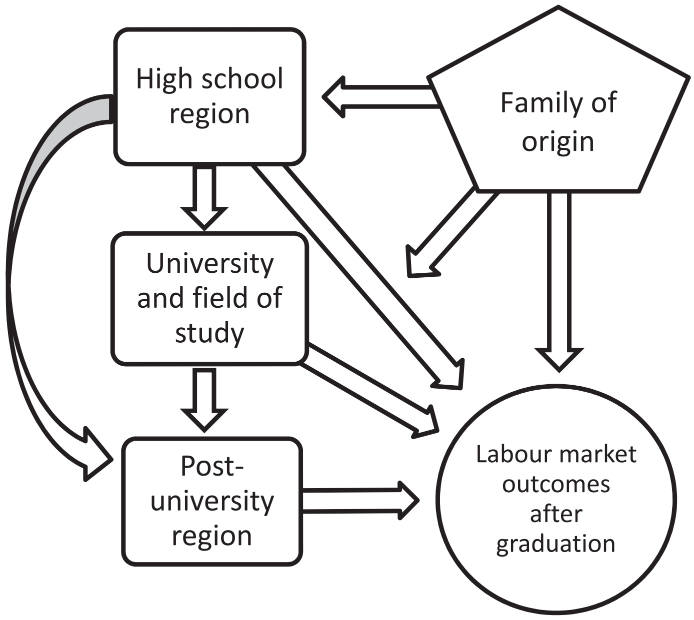


FIGURE 3 Framework of the thesis

4.2 Data and methods

All four studies of this thesis utilise the same main data source: an administrative micro-dataset obtained from Statistics Finland.[8] The dataset has been constructed by drawing a seven per cent random sample from Finland's 2001 population census. The total sample size is 363,643 individuals. The dataset includes rich panel information, for example, on the individuals' labour market outcomes, residential region, job characteristics, educational qualifications and family background from the period between 1970 and 2006. The empirical analyses of the thesis focus on university graduates, and therefore, only a small portion of the full data set is employed during the analyses. The criteria used in restricting the study sample vary slightly across the chapters. As Chapters 2 and 3 examine students' field-of-study choices at the stage of moving from high school to university, these chapters employ a sample of individuals who graduated from high school roughly at the same time – between 1991 and 1996 – and subsequently obtained a higher education degree. Chapters 4 and 5 again study labour market outcomes after graduation, and the samples used in these studies are therefore comprised of individuals who graduated within a few specific years (1994–2000 in Chapter 4 and 1995–2002 in Chapter 5). With regard to analyses of higher education choices and their returns, one of the greatest advantages of the dataset arises from the detailed information available on the students' characteristics before entering university education. In particular, the variables depicting the students' matriculation examination grades in mathematics and first language, region of origin and parents' educational degrees and occupational statuses are highly relevant predictors of the choice of university and field of study and therefore serve as important control variables during the analyses.

Various estimation approaches are utilised in the econometric analyses of this thesis. In Chapters 2 and 3, the dependent variable of interest is a student's field of study, which is categorical. Therefore, these chapters utilise multinomial discrete choice methods, including conditional, multinomial, nested and mixed logit models. In Chapters 4 and 5, the dependent variable, the natural logarithm of a student's earnings after graduation, is continuous, and therefore, linear regression approaches are employed in these chapters. A problematic feature in the econometric modelling of individuals' higher education choices and their returns is that unobserved individual characteristics – such as innate ability, motivation, ambition, etc. – may be correlated with both the explained and explanatory variables of the models, which may result in biased estimates. For instance, in the context of estimating the earnings effect of graduating from a high-quality university by using a linear regression model (see Chapters 4 and 5), one may suspect that there is positive correlation between an indicator of high quality and the error term because students attending high-quality

---

8 As an additional data source, Chapters 2, 4 and 5 use the KOTA database of the Ministry of Education available at https://kotaplus.csc.fi, which includes yearly panel information on the Finnish university institutions from 1980 onwards.

universities may be of higher average ability than students attending low-quality universities. Hence, the estimate for the variable of interest is potentially upward-biased. Although the baseline results in each chapter are obtained by controlling for a highly informative set of observed individual and family-of-origin characteristics, robustness to unobserved heterogeneity is also studied using different techniques. Chapters 2 and 3 utilise nested logit and mixed logit models to allow the substitution patterns in the estimated field-of-study-choice models to depend on unobserved individual- and alternative-specific characteristics. Furthermore, in Chapters 4 and 5, instrumental variables approaches (including two-stage least squares and different control function approaches) are employed to correct for unobserved individual heterogeneity across location- and field-of-study-specific groups. In these approaches, changes in region-specific supplies of student places over time are used as a primary source of exogenous variation; that is, these changes are assumed to have affected individuals' educational and/or locational choices while not being directly related to their early-career earnings.

4.3 Main findings

Chapter 2 studies the effect of distance on the choice of field of study. A particular focus of this chapter is on estimating the effect of the shortest distance to enrol in a field on the likelihood of choosing that field. The estimated distance effect is sizeable: the preferred model specifications suggests that having a 100 kilometre longer distance to enrol is, on average, associated with 15% smaller odds of choosing a field. However, the distance effect appears to vary, to some extent, depending on the choice alternative in question: a student's probability of studying education, arts or medicine is not found be sensitive to distances, whereas considerable distance effects are measured for all other fields.

Chapter 3 investigates the intergenerational transmission of field of study. The obtained results suggest that a significant amount of intergenerational transmission occurs: even according to the most cautious estimate, having a parent with a university degree from a particular field is, on average, associated with a 1.8 times higher odds of graduating from that field among individuals with similar observable characteristics. However, the parental effect on the field-of-study choice appears to be heterogeneous in many ways. In particular, the results suggest that some fields are transmitted from parents to children more frequently than others. The intergenerational transmission appears to be clearly the strongest in the case of law: a student having a parent with a law degree is found to have a more than 5 times higher probability of graduating from law than an observably similar student without that type of parent. Relatively strong intergenerational transmission is also observed in the case of education, business and medicine/health sciences – the estimates for these fields are mainly above the estimated average effect – whereas the remaining fields appear to be transmitted to a lesser extent. The parental effect is also found to differ across genders – the estimated effects are larger for men than for women – and even among individuals with similar observable characteristics. In fact, the results from the mixed logit analysis suggest that a sizeable proportion of individuals place a negative weight on fields chosen by their parents. Furthermore, the results indicate that individuals are more likely to imitate their father's field-of-study choice than the choice of their mother.

Chapter 4 studies the effect of graduating from a metro area university on university students' early-career earnings. The descriptive analysis of this chapter shows that metro area graduates earn an average of 13.6% more than other graduates during the six-year period after graduation. However, according the estimation results, this earnings premium is entirely explained by differences in students' pre-university characteristics and not by the choice of university location itself. The OLS estimates for the average metro area university premium are close to zero and insignificant, whereas the IV estimation provides considerably negative point estimates. However, the IV estimates are also more imprecise, which suggests – along with the findings indicating that the treatment effects are heterogeneous – that the IV results should be interpreted with caution. Thus, we may conclude that despite the locational heterogeneity of universities in Finland, no strong evidence is found that university choice is generally significant in terms of students' labour market success. The findings of Chapter 4 nonetheless suggest that certain subgroups may benefit from locational choices. According to the Heckman-corrected results obtained for 'residents in metro area' and 'residents in other regions', graduation from a metro area university has a sizeable positive average effect on earnings for the latter group but not for the former. Thus, the valuation of a graduate's university choice appears to depend on the regional labour market in which he or she is located after graduation. Furthermore, the results indicate heterogeneity in the university location effects across fields. For example, positive estimates for graduating from a metro area university are obtained for students from business and social sciences, whereas humanities students are found to benefit considerably from attending university outside Helsinki.

In Chapter 5, the association between different measures of university quality and students' monthly earnings four years after graduation are studied. Although most of the results for the employed quality measures are insignificant, suggesting a rather weak quality-earnings relationship, certain significant linkages are also found. In particular, a positive and significant relationship is found between the teachers/student ratio and earnings, suggesting that there may be benefits for students from increasing the amount of educational resources. However, the results also indicate heterogeneity in this relationship: graduating from an institution with a high teachers/student ratio appears, on average, to be beneficial for women but not for men, and in some fields (education, humanities and natural sciences) but not in others. Very little evidence of returns from graduating from a selective institution – proxied by the applicants/admitted ratio – is obtained during the analysis. A significantly positive selectivity-earnings relationship is merely found in the case of natural sciences graduates. The remaining quality measures – the publications/researcher ratio and two 'quality factors' combining the original quality measures – are likewise insignificant in most cases; the results even indicate a negative association between these variables and the early-career earnings of men.

4.4 Concluding remarks

A general conclusion that may be drawn based on the evidence of this thesis is that a student's location before university is an important determinant of higher education choices – the choices regarding both 'where' and 'what' to study – in Finland. This evidence may have particular policy relevance in the context of Finland's current university reform, which aims to increase both stratification and differentiation of universities (see, e.g., Tirronen and Nokkala, 2009). Namely, one should bear in mind that, if the future university system will consist of more specialised universities and universities with larger status differences, it is possible that the region of origin will be an even greater determinant of the type and the quality of the university education that a student receives. Consequently, the utilisation of resources will likely be less efficient, as the universities' student places will be, to a larger extent, allocated to 'non-optimal' students. The increased geographical stratification of higher education could also have adverse effects on regional development, as some regions could, in the presence of limited labour mobility, experience larger shortages of highly educated workers in certain fields and industries. However, it is possible that these adverse effects of the reform will be small compared to the potential benefits that arise from an improved international competitiveness of the university system. This could be the case, particularly, if the interregional mobility of students and graduates was higher in the future. An analysis comparing the costs and benefits of this type of reform would pose an interesting, although highly challenging area of future research.

On the other hand, the obtained evidence suggests that, in general, a university student's location during studies does not matter for his or her early-career labour market outcomes in Finland. Clearly, the choice of field of study, rather than university institution or location, is the primary mechanism behind the heterogeneity of returns to higher education in this country. Although Chapter 5 provides certain evidence of 'institution quality effects', these effects appear not to be strong enough to generate significant returns from university choice, at least during graduates' early careers. From a policy perspective, these results suggest that placing university institutions in relatively heterogeneous locations does not necessarily generate significant inequality among students or graduates, at least when educational resources are divided rather equally across the institutions. Given these conclusions, it is interesting to see whether the on-going university reform will alter the Finnish university system towards a more geographically stratified system with 'good and bad places to be in'.

REFERENCES

Alexander, K.L., Holupka, S. & Pallas, A.M. 1987 Social background and academic determinants of two-year versus four-year college attendance: Evidence from two cohorts a decade apart. American Journal of Education 96, 56–79.

Alm, J. & Winters, J.V. 2009 Distance and intrastate college student migration. Economics of Education Review 28, 728–738.

Angrist, J.D. & Krueger, A.B. 1991. Does compulsory school attendance affect schooling and earnings? Quarterly Journal of Economics 106, 979–1014.

Arcidiacono, P. 2004. Ability sorting and the returns to college major. Journal of Econometrics 121, 343–375.

Avery, C. & Hoxby, C.M. 2004. Do and should financial aid packages affect students' college choices? In C.M. Hoxby (Ed.) College choices: The economics of where to go, when to go, and how to pay for it, University of Chicago Press, 239–301.

Ashenfelter, O. & Krueger, A.B. 1994. Estimating the returns to schooling using a new sample of twins. American Economic Review 84(5), 1157–1171.

Ashenfelter, O. & Rouse, C. 1998. Income, schooling and ability: Evidence from a new sample of identical twins. Quarterly Journal of Economics 113, 253–284.

Becker, G.S. 1962. Investment in human capital: A theoretical analysis. Journal of Political Economy 70(5), 9–49.

Beffy, M., Fougère, D. & Maurel, A. 2012. Choosing the field of study in postsecondary education: Do expected earnings matter? Review of Economics and Statistics 94, 334–347.

Behrman, J., Rosenzweig, M. & Taubman, P. 1996. College choice and wages: Estimates using data on female twins. Review of Economics and Statistics 78 (4), 672−685.

Berger, M. 1988 Predicted future earnings and choice of college major. Industrial and Labour Relations Review 41, 418–429.

Björklund, A., and K. Salvanes. 2011. Education and family background: Mechanisms and policies. In Hanushek, E., Machin, S. & Woessmann, L. (Eds.) Handbook of the Economics of Education, Vol. 3, Elsevier, 201–247,.

Black, D., Kolesnikova, N. & Taylor, L., 2009. Earnings functions when wages and prices vary by location. Journal of Labor Economics 27 (1), 21–47.

Black, D. & Smith, J. 2004. How robust is the evidence on the effects of college quality? Evidence from Matching. Journal of Econometrics 121 (1−2), 99−124.

Black, D. & Smith, J. 2006. Estimating the returns to college quality with multiple proxies for quality. Journal of Labor Economics 24 (3), 701–728.

Bratti, M. 2006. Social class and undergraduate degree subject in the UK. IZA Discussion Paper no. 1979.

Brewer, D., Eide, E. & Ehrenberg, R. 1999. Does it pay to attend an elite private college? Cross-cohort evidence on the effects of college type on earnings. Journal of Human Resources 34 (1), 104–123.

Brunello, G. & Cappellari, L. 2008. The labour market effects of Alma Mater: Evidence from Italy. Economics of Education Review 27(5), 564–574.

Boudarbat, B. 2008. Field of study choice by community college students in Canada. Economics of Education Review 27, 79–93.

Boudatbat, B. & Montmarquette, C. 2009. Choice of fields of study of Canadian university graduates: The role of gender and their parents' education. Education Economics 17, 185–213.

Card, D. 2001. Estimating the return to schooling: Progress on some persistent problems. Econometrica 69(5), 1127–1160.

Checchi, D. 2006. The economics of education. Human capital, family background and inequality. Cambridge University Press, New York.

Dale, S.B. & Krueger, A. 2002. Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. Quarterly Journal of Economics 117 (4), 1491–1527.

Dale, S.B. & Krueger, A. 2011. Estimating the return to college selectivity over the career using administrative earnings data. NBER Working Paper 17159.

Davis, S. & Guppy, N. 1997. Fields of study, college selectivity, and student inequalities in higher education. Social Forces 75(4), 1417–1438.

Chevalier, A. & Conlon, G. 2003. Does it pay to attend a prestigious university? CEP Discussion Paper No. 33. London School of Economics and Political Science.

Cooke, T.J. & Boyle, P. 2011. The migration of high school graduates to college. Educational Evaluation and Policy Analysis 33(2), 202–213.

Eliasson, K., 2006. The role of ability in estimating the returns to college choice: New Swedish evidence. Umeå Economic Studies No. 691, Umeå University.

Freeman, J.A. & Hirsch, B.T. 2008. College majors and the knowledge content of jobs. Economics of Education Review 27, 517–535.

Frenette, M. 2006. Too far to go on? Distance to school and university participation. Education Economics 14, 31–58.

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

Goyette, K.A. & Mullen, A.L. 2006. Who studies the arts and sciences? Social background and the choice and consequences of undergraduate field of study. Journal of Higher Education 77, 497–538.

Haveman, R. & B. Wolfe. 1995. The determinants of children's attainments: A review of methods and findings. Journal of Economic Literature 33, 1829–1878.

Heckman, J. 1979. Sample selection bias as a specification error. Econometrica 47(1), 153–162.

Hershbein, B. 2011. Worker signals among new college graduates: The role of selectivity and GPA. Unpublished manuscript.

Hoekstra, M. 2009. The effect of attending the flagship state university on earnings: A discontinuity-based approach. Review of Economics and Statistics 91(4), 717–724.

Holmlund, L. 2009. The effect of college quality on earnings. Evidence from Sweden. Umeå Economic Studies No. 781, Umeå University.

Hotelling, H. 1929. Stability in competition. Economic Journal 39, 41–57.

James, E., Alsalam, N., Conaty, J.C. & To, D. 1989. College quality and future earnings: where should you send your child to college? American Economic Review 79, 247–252.

Jepsen, C. & Montgomery, M. 2009. Miles to go before I learn: The effect of travel distance on the mature person's choice of a community college. Journal of Urban Economics 65, 64–73.

Klemperer, P. 1987. The competitiveness of markets with switching costs. RAND Journal of Economics 18(1), 138–150.

Lang, K. & Siniver, E. 2011. Why is an elite undergraduate education valuable? Evidence from Israel. Labour Economics 18(6), 767–777.

Leppel, K., Williams, M.L. & Waldauer, C. 2001. The impact of parental occupation and socioeconomic status on choice of college major. Journal of Family and Economic Issues 22, 373–394.

Lindahl, L. & Regnér, H. 2005. College choice and subsequent earnings: Results using Swedish siblings data. Scandinavian Journal of Economics 107(3), 437–457.

Long, B.T. 2004. How have college decisions changed over time? An application of the conditional logistic model. Journal of Econometrics 121, 271–296.

Long, M.C. 2008. College quality and early adult outcomes. Economics of Education Review 27(5), 588–602.

Loury, L.D. & Garman, D. 1995. College selectivity and earnings. Journal of Labor Economics 13(2), 289–308.

Mincer, J. 1974. Schooling, experience, and earnings. Columbia University Press, New York.

Ministry of Education, 2005. OECD thematic review of tertiary education: Country background report for Finland. Publications of the Ministry of Education, Finland 2005:38.

Monks, J. 2000. The returns to individual and college characteristics: Evidence from the National Longitudinal Survey of Youth. Economics of Education Review 19(3), 279–289.

Montmarquette, C., Cannings, K. & Masheredjian, S. 2002. How do young people choose college majors? Economics of Education Review 21, 543–556.

Oosterbeek, H. & Webbink, D. 1997. Is there a hidden technical potential? De Economist 145(2), 159–177.

Ordovensky, J. 1995. Effects of institutional attributes on enrollment choice: Implications for postsecondary vocational education. Economics of Education Review 14, 335−350.

Sá, C., Florax, R. & Rietveld, P. 2004. Determinants of the regional demand for higher education in the Netherlands: A gravity model approach. Regional Studies 38, 375–392.

Sacerdote, B. 2001. Peer effects with random assignment: Results for Dartmouth roommates. Quarterly Journal of Economics 116(2), 681–704.

Schultz, T.W. 1961. Investment in human capital. American Economic Review 1(2), 1–17.

Schwartz, A. 1973. Interpreting the effect of distance on migration. Journal of Political Economy 81(5), 1153–1169.

Shy, O. & Stenbacka, R. 2003. Strategic outsourcing. Journal of Economic Behavior & Organization 50, 203–224.

Sjaastad, L. 1962. The costs and returns of human migration. Journal of Political Economy 70(5), 80–93.

Tirronen, J. & Nokkala, T. 2009. Structural development of Finnish universities: Achieving competitiveness and academic excellence. Higher Education Quarterly 63(3), 219–236.

Triventi, M. & Trivellato, P. 2012. Does graduating from a private university make a difference? Evidence from Italy. European Journal of Education 47(2), 260–276.

Wolniak, G.C., Seifert, T.A., Reed, E.J. & Pascarella, E.T. 2008 College majors and social mobility. Research in Social Stratification and Mobility 26, 123–139.

Wolszczak-Derlacz, J. & Parteka, A. 2011. Efficiency of European higher education institutions: a two-stage multicountry approach. Scientometrics 89, 887–917.

Zimmerman, D.J. 2003. Peer effects in academic outcomes: evidence from a natural experiment. Review of Economics and Statistics 85(1), 9–23.

CHAPTER 2: FIELD-OF-STUDY CHOICE IN HIGHER EDUCATION: DOES DISTANCE MATTER?[9]

When field-of-study options vary across higher education institutions, geographical distances may create barriers to students' study choices. Based on this hypothesis, the present study empirically examines field-of-study decisions in Finland's university system, focusing on the effect of distance. The results of the conditional and nested logit models suggest that a 100-kilometre increase in the shortest distance to enrol in a field is, on average, associated with an approximately 15% reduction in the likelihood of selecting that field. However, the effect of distance varies, to some extent, across the choice alternatives and is insignificant when choosing education, arts or medicine, while large and significant in most other cases.

Keywords: higher education, field of study, geographical distance, discrete choice models

---

9 Financial support from the Academy of Finland (grant number 127049) and the Yrjö Jahnsson Foundation is gratefully acknowledged. The author also wishes to thank Mika Haapanen, Kaisa Kotakorpi, Timo Kuosmanen, Hannu Tervo, Olle Westerlund, the two anonymous referees, and the participants at several conferences (including the ERSA Conference in Palermo, the RSA Conference in Tampere and the Finnish Economic Association meeting in Maarianhamina) for their many helpful comments.

1 Introduction

The field of study pursued in postsecondary education is known to be an important determinant of students' future occupations and earnings. For instance, completing a university degree in business, law or engineering generally results in higher average earnings than a degree in education or the humanities. Thus, it may be argued that given the rapid increase in the general educational level observed recently, individuals' postsecondary field-of-study choices have become increasingly important determinants of earnings inequality and social stratification (Wolniak *et al.*, 2008). Partly on this basis, social scientists have become increasingly interested in studying the mechanisms behind these choices. Numerous empirical studies have extensively examined the roles of a variety of socio-demographic variables (for reviews, see Leppel *et al.*, 2001; Goyette & Mullen, 2006) and expected earnings (Berger, 1988; Montmarquette *et al.*, 2002; Bourdarbat, 2008; Beffy *et al.*, 2012) in determining field-of-study choice. However, whereas many studies have analysed the effect of geographical distance on participation in higher education or the higher education institution selected (e.g., Ordovensky, 1995; Avery & Hoxby, 2004; Long, 2004; Frenette, 2006; Jepsen & Montgomery, 2009; Gibbons & Vignoles, 2012), very little evidence has been presented regarding distance effects in the context of field-of-study selection.[10] In theory, if a particular educational field is unavailable near a student's place of residence, the various costs involved in migrating farther away could reduce the student's willingness to apply to that field, particularly if other study options are available nearby. This 'distance deterrence effect' could be particularly relevant in countries with large distances between higher education institutions and where the institutions' field-of-study options differ. For instance, in Finland – the country studied in this paper – university degrees in some of the smaller fields, such as law and the arts, are only available in a very limited number of locations. Thus, given the relatively low interregional mobility of Finnish students after high school (see Jauhiainen, 2010; Suhonen, 2013), one may suspect that regional disparities exist with respect to enrolments in these fields.

This paper empirically examines students' field-of-study choices in Finland's higher education system with a particular focus on the role of geographical distance as a determinant of these choices. Register-based data on a sample of Finnish high school graduates from 1991–1996 are used in the analysis. More specifically, the study seeks to answer the following primary question: does an increase in the shortest distance required to enrol in a particular field decrease one's likelihood of selecting that field? The Finnish data offer a particularly good opportunity for an analysis of this question because Finland is a relatively vast and sparsely populated country – 5.4 million

10 Thus far, the papers by Bertrand-Cloodt *et al.* (2010) and Denzler & Wolter (2011) appear to be the only ones providing evidence of the effect of geographical distance on field-of-study choice. These papers are briefly discussed in Section 2.

inhabitants and a land area of 338,432 square kilometres (2011) – with a geographically dispersed public higher education system. Thus, the data provide substantial variation in the distances between the high school graduates' residential locations and higher education institutions. Another advantage of this case is that, in Finland's higher education system, studying is free of charge and generously subsidised by the public sector. Therefore, the direct financial costs of higher education attendance, other than those arising from the costs of migration, are relatively low and unlikely to confound the findings regarding the effect of distance on field-of-study decisions.

The Finnish higher education system consists of two distinct sectors: universities and polytechnic institutes. The purpose of the former sector is to provide academic research and education at the Bachelor's, Master's and PhD levels, whereas the latter sector is devoted to more practice-oriented, vocational higher education and awards Bachelor's degrees, for example, in business administration, engineering and nursing. [11] The present study exclusively focuses on the effect of distance on one's choice of field within the university sector. The main reason for this choice is that, of the two sectors, the university sector is far more geographically centralised: while polytechnic education in the most common fields (business, engineering and nursing) is available in all 19 NUTS-3 regions in Finland, the university institutions' main campuses are only located in 10 large- or middle-sized cities and are considerably heterogeneous with respect to the field-of-study options they offer. Therefore, the 'distance issue' is, in general, likely to be more relevant in the case of university enrolment. Another reason relates to the exclusion of students without a high school diploma from the study sample: because higher education enrolment is possible without a high school diploma but far more common in polytechnics than in universities, the resulting sample selection problem is likely to be less severe when focusing on the university-level alternatives.[12] The study options available in polytechnics are, nonetheless, included in the empirical analysis as separate choice alternatives to account for the fact that students may substitute between the two sectors as a result of changes in distances. For instance, being required to travel a long distance to enrol in studies leading to a Master's degree in business (*kauppatieteiden maisteri*) might cause a 'business-oriented'

---

11 The polytechnic sector is the younger of the two higher education sectors: the latest university establishments took place in the late 1970s, while the polytechnics were established during the 1990s on the basis of former vocational colleges. Because the current analysis uses data from this transitional period, some of the lower-level degrees in the study sample are vocational college degrees (*opistoasteen tutkinto*) instead of polytechnic degrees (*ammattikorkeakoulututkinto*). For simplicity, all of the lower-level degrees are nonetheless referred to as polytechnic degrees throughout the article.

12 No statistics were found regarding enrolment in higher education without a high school diploma. However, to study this phenomenon, a sample of 12,526 individuals who graduated from a secondary-level vocational school between 1991 and 1996 was drawn from the Statistics Finland dataset used in this study. Of these graduates, 2,321 (19%) had subsequently completed a polytechnic degree by 2006, whereas only 243 (2%) had completed a university degree. These figures confirm the common perception that vocational school graduates primarily sort into the polytechnic sector when enrolling in higher education.

student to choose business studies at a nearby polytechnic – leading to a lower-level business degree (*tradenomi*) – whereas a long distance to the nearest university with a faculty of medicine might again cause another type of individual to pursue a polytechnic degree in nursing. Provided that these types of between-sector substitution patterns might vary with distance, depending on regional variation in the supply and demand of the study options, it may be necessary to consider the entire set of choice alternatives within the higher education system to obtain unbiased evidence of distance effects.[13]

The article is organised as follows. Section 2 discusses the theoretical explanations and previous empirical evidence regarding distance effects in postsecondary study choices. Section 3 briefly describes Finland's university system, focusing on its geographical structure and admission system. Section 4 discusses the data set used in the study and presents descriptive evidence regarding the determinants of field-of-study choice. Section 5 discusses the methodology used in the empirical analysis: the baseline results are obtained by estimating conditional logit models that control for observable, individual-specific variables such as matriculation grades and parental characteristics, while the estimates' robustness to substitution patterns arising from unobserved heterogeneity is studied using nested logit models. The results of the empirical analysis are presented in Section 6, and Section 7 concludes the paper.

2 Previous literature

Researchers have long recognised that an unequal spatial distribution of educational opportunities combined with substantial geographical distances can restrict individuals' educational choices because of the various costs involved in migrating from home to school. In the context of enrolling in higher education, the selection of a distant university or field-of-study alternative instead of one located nearby involves costs arising from at least three sources (see, for example, Sjaastad, 1962; Schwartz, 1973; Leppel, 1993). First, transportation to the more distant location entails additional time and money (direct costs). Second, information on the more distant alternative may be less available (informational costs). Third, individuals may feel uncomfortable

13 Certain university and polytechnic degrees – particularly in the fields of business and technology – can be considered close substitutes in terms of their subject matter. Therefore, the estimated distance effects should be partly considered a result of both the choice of educational level and the choice of field. However, all degrees leading to a particular profession, such as those of schoolteacher, psychologist, lawyer, architect, doctor and nurse, are sector-specific. In addition, the 'liberal-arts-oriented' university degrees, including most of the degrees in humanities, social sciences and natural sciences, are clearly distinct from the polytechnic degrees. On these grounds, it appears to be a reasonable strategy to treat the study options within the two sectors as separate choice alternatives when examining the mechanisms of field-of-study choice.

about leaving a familiar environment and living far from family and friends (psychic costs). In the empirical literature, very little discussion or evidence is available on the importance of these different types of costs. However, one might suspect that if students' study choices in modern societies are affected by distance, the importance of direct and informational costs in explaining these effects may be rather small because of highly developed transportation, information and communication technologies. The psychic costs of migration could therefore play a major role in explaining potential distance effects.

The role of geographical distance in postsecondary study choices has attracted a significant amount of attention in the literature. Closely related to the current paper are studies that exploit individual-level data to examine the effect of distance on participation in higher education and/or the choice of higher education institution. Most of these studies originate from the U.S. (e.g., Ordovensky, 1995; Avery & Hoxby, 2004; Long, 2004; Jepsen & Montgomery, 2009), but evidence from Canada (Frenette, 2006) and Britain (Gibbons & Vignoles, 2012) is also available. A common methodological approach in these studies has been to measure the shortest travel distance between a student's home and a higher education institution and include this measure as an explanatory variable in a multinomial choice model. The identification of the 'causal' distance effect is then achieved by controlling for various individual characteristics (e.g., high school grades and parental education) and additional regional/institutional attributes such as the tuition and fees charged by the relevant institutions. Most of these studies find that the distance to a particular higher education institution is negatively and significantly associated with the likelihood of selecting that institution (Ordovensky, 1995; Long, 2004, Jepsen & Montgomery 2009, Gibbons & Vignoles, 2012); however, the findings of Avery & Hoxby (2004) from the U.S. suggest that the college choices of high-aptitude students are not sensitive to distance. The evidence regarding the effect of the distance to the nearest higher education institution on enrolment in higher education has, again, been rather mixed across studies and countries, varying from the strong effects observed in Canada (Frenette, 2006) to the weak or insignificant effects observed in Britain (Gibbons & Vignoles, 2012).

In the light of the evidence described above, one might suspect that geographical distance to a higher education institution could also affect the field of study a student elects to pursue, particularly if the variety of fields differs across locations. However, there is reason to believe that distance does not generate a significant deterrence effect: if students' preferences regarding field-of-study choice are strong, they may be willing to migrate long distances to study in their preferred fields. There appears to be very little empirical evidence on this matter. In a recent study, Bertrand-Cloodt *et al.* (2010) examine the effect of distance in the context of selecting a field in upper-secondary education among Dutch students. These authors estimate binary and multinomial logit models with three field-of-study alternatives, concluding that a significant inverse relationship exists between the shortest distances to the institutions offering these fields and the likelihood of selecting these fields. The investigation conducted by Denzler and Wolter (2011), which is based on questionnaire data on Swiss high school leavers, appears to be the only one examining the effect of distance in the context of selecting a field in higher education. The findings of their study suggest that the distances to two particular Swiss universities, located in Zurich and Lausanne, are negatively associated with a student's decision to enrol in these universities and, more importantly for the current study, that the proximity of a specialist university offering a limited number of subjects (the university in Lausanne) has a negative effect on one's decision to study law.

In summary, previous studies have rather systematically supported the view that migration costs affect individuals' study choices, even in the context of selecting an educational field. Thus far, the effect of distance on field-of-study choice has only been studied in geographically small countries (the Netherlands and Switzerland). Therefore, as Finnish university system considered here is highly geographically dispersed, and knowing that Finnish students are relatively immobile after graduation from high school (e.g., Jauhiainen, 2010; Suhonen, 2013), one would expect to observe an even stronger relationship between distance and field-of-study choice in the present analysis.

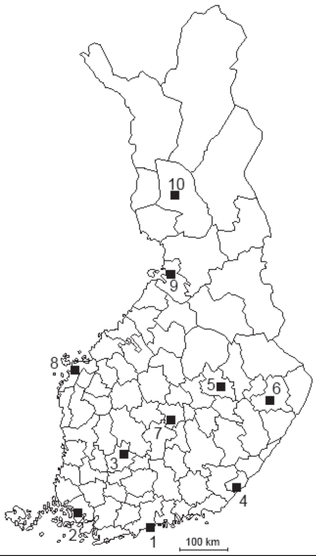
3 Universities and university admissions in Finland

In the 1990s, when high school graduates in the study sample entered higher education, the Finnish university system comprised 16 universities (10 multidisciplinary universities, 3 business schools and 3 universities of technology), 4 art academies and the Finnish National Defence University. All of these institutions still exist, but some of them merged as a result of reforms to the university system than began in the late 2000s, which decreased the number of institutions to 15. Despite these mergers, the geographical structure of the university system remains similar to that of the 1990s depicted in Figure 1.[14] The universities' main campuses are located in 10 urban regions: Helsinki, Turku, Tampere, Lappeenranta, Kuopio, Joensuu, Jyväskylä, Vaasa, Oulu and Rovaniemi. Thus, the system offers a university education that is relatively close to each individual's place of residence, which promotes regional equity in access to university education. However, as we can observe from Figure 1, not

[14] The universities' field-specific educational responsibilities established by the Ministry of Education and Culture also remained highly stable over time. One of the few notable changes occurred in the technology field in 2004, when the right to award university degrees in technology (*diplomi-insinööritutkinto*) was given to the University of Turku and the University of Vaasa. Another change in the geography of the university system involved the establishment of smaller university consortiums – serving as the universities' branch campuses – in 6 other cities. These consortiums are not, however, taken into account in the current empirical analysis when calculating the distances to the fields: because the consortiums were only established in 2004 and are primarily dedicated to adult education, it is unlikely that they had a significant impact on the study choices of the students in the current sample.

all fields of study are available at every university location, and therefore, significant regional variation exists in the distances to particular fields. For example, a university degree in law is only available in three locations: Helsinki, Turku and Rovaniemi. Therefore, some high school graduates must migrate several hundred kilometres to obtain this degree. Thus, a policy-relevant question is whether the unequal geographical distribution of field-of-study options within the university system generates regional disparities in field-of-study decisions.

All Finnish universities are free of charge and publicly subsidised by the provision of study grants, housing allowances and state-guaranteed loans. In addition, as local student housing foundations provide students with inexpensive apartment rentals near the universities, the non-distance-related financial barriers involved in university enrolment or the selection of an institution can be considered fairly small in Finland. In the university admissions process, all the individuals holding a high school diploma or a 3-year vocational qualification are eligible to apply to any university or field of study, regardless of, for example, subject choices made at prior stages of education. When applying to universities, students directly apply for specific majors or broader fields offered by the institutions. Thus, in the Finnish system, the decision regarding the preferred area of concentration must be made relatively early, specifically at the application stage. The supply of university education is regulated by the Ministry of Education and Culture, and therefore, the universities' faculties must apply a *numerus clausus* policy, i.e., they can only admit a limited number of students to each field of study in each year. Typically, admission is based on a subject-specific entrance examination, students' grades from the matriculation examination and final high school grades. Some faculties, particularly in the natural sciences, also admit students with high matriculation grades directly without requiring an entrance examination. As the number of applicants (demand) has, with rare exceptions, always exceeded the number of student places (supply) in all universities and fields (Ministry of Education and Culture, 2013), the Finnish university admission system can be considered fairly competitive. With respect to the current empirical analysis, this feature implies that a student's field of study is not, in general, a result of free choice. This problem is accounted for by controlling for students' matriculation grades, which constitute an important aspect of the admission criteria throughout the university system.



Nr.	City	University institution	Fields of study
1	Helsinki	Uni. Helsinki	Edu, Hum, Soc, Law, Nat, Med, Agr
		Helsinki School of Econ.	Bus
		Helsinki Uni. Technology	Tec
		Swedish School of Econ. (S)	Bus
		Sibelius Academy	Arts
		Uni. Art and Design Helsinki	Arts
		Theatre Academy	Arts
		Finnish Academy of Fine Arts	Arts
		National Defense Uni.	Military
2	Turku	Uni. Turku	Edu, Hum, Soc, Law, Nat, Med, HS
		Turku School of Econ.	Bus
		Åbo Akademi Uni. (S)	Hum, Bus, Soc, Nat, Tec, Med, HS
3	Tampere	Uni. Tampere	Edu, Arts, Hum, Bus, Soc, Nat, Med, HS
		Tampere Uni. Technology	Tec
4	Lappeenranta	Lappeenranta Uni. Technology	Bus, Tec
5	Kuopio	Uni. Kuopio	Soc, Nat, Med, HS
6	Joensuu	Uni. Joensuu	Edu, Hum, Soc, Nat, Agr
7	Jyväskylä	Uni. Jyväskylä	Edu, Hum, Bus, Soc, Nat, HS, Sports
8	Vaasa	Uni. Vaasa	Edu, Hum, Bus, Soc
		Swedish School of Econ. (S)	Bus
		Åbo Akademi Uni. (S)	Edu, Soc
9	Oulu	Uni. Oulu	Edu, Hum, Bus, Nat, Tec, Med, HS
10	Rovaniemi	Uni. Lapland	Edu, Arts, Soc, Law

FIGURE 1 Geography of the Finnish university system. The NUTS-4 regional classification from 2003 is used in the map. The institution names and the field-of-study options reflect the situation in the 1990's. 'S' stands for a Swedish-speaking university institution. Abbreviations: Edu = education; Hum = humanities; Bus = business; Soc = social sciences; Nat = natural sciences; Tec = technology; Med = medicine; HS = health sciences; Agr = agriculture and forestry.

4 Data and descriptive analysis

The study data are based on a 7% random sample drawn from the population of Finland in 2001. The data originate from the registers of Statistics Finland and include yearly panel data, for example, on individuals' labour market outcomes, region of residence, employment characteristics, educational qualifications and family background from the period between 1970 and 2006. Within this large dataset, the study focuses on a subsample of 11,660 individuals who graduated from high school between 1991 and 1996 at a young age (less than 23 years old). Based on the first educational qualification obtained after high school, the individuals can be divided into four groups: 1) 3,398 university graduates (29%), 2) 4,115 polytechnic graduates (35%), 3) 2,058 graduates from secondary-level vocational schools (18%) and 4) 2,089 graduates who did not complete any qualification by the end of 2006 (18%).[15] Only the first two groups – that is, the individuals with a higher education degree – are included in the current estimation sample; thus the final sample size is limited to 7,513 individuals.[16] Based on the information regarding the field in which a degree was obtained, the university degrees are divided into ten categories: 'education', 'arts', 'humanities', 'business', 'social sciences', 'law', 'natural sciences', 'technology', 'medicine' and 'other field' (including agriculture, forestry, military and sports).[17] Furthermore, the polytechnic degrees are divided into five categories:

---

15 Finnish university students, on average, graduate at a relatively old age – approximately 27 years according to the OECD (2010) – and there is substantial variation in time to graduation: for instance, according to Statistics Finland (2013), only 49% of university enrolees were able to complete a degree in 5.5 years, while 82% completed a degree within the 16-year interval. Therefore, as the latest observation year available in the current Statistics Finland dataset is 2006, the study sample must be chosen from a period before the late 1990s to obtain a representative sample of university graduates.

16 Compared to the polytechnic degrees, the secondary-level vocational qualifications are, based on both educational level and subject matter, more distinct from the university degrees and thus less likely to represent distance-relevant substitution patterns in the current context, particularly because polytechnic education is abundantly available across the regions of Finland. Therefore, and to simplify the estimations, the secondary-level graduates were excluded from the sample. Some estimations were also conducted using broader field-of-study categories that combine polytechnic and secondary-level graduates, but the results obtained using this approach were approximately similar to those without the secondary-level graduates. The individuals without a post-high-school qualification are again excluded because it is likely that a large proportion of them were enrolled in higher education – and, thus, chose one of the field-of-study alternatives – but dropped out or were unable to complete a degree by 2006. Thus, as we are interested in individuals' choices, rather than their success or graduation, treating these individuals as a separate, homogeneous group would likely result in biased results.

17 Aside from the field of study, other information regarding the completed degrees can also be observed from the data. Based on the information, only 14% of the university degrees were Bachelor's degrees, and the remaining 86% were Master's degrees. This result is not surprising, as completing a Bachelor's degree before a Master's degree was optional – and therefore rare – in the Finnish university system until 2005. The data also show that time to graduation clearly differs across the university and polytechnic sectors: the average university graduate completed his/her degree at the age of 26, 7.2 years after graduating from high school, whereas the corresponding

'humanities', 'business', 'technology', 'health' and 'other field'. These 15 field categories – which essentially cover all of the study options available in the Finnish higher education system – constitute a decision maker's choice set in the discrete choice models estimated below.[18]

To study the impact of geographical distance on the choice of a university-level field alternative, the road distances from 81 NUTS-4 regions to the 10 university cities are matched with the micro data on the basis of a student's residential location one year prior to the year in which he or she graduated from high school. Based on these distances, the shortest distance to each field of study is determined to generate the main explanatory variable: the *shortest distance to enrol*. To provide a concrete example, for a high school graduate from the NUTS-4 region of Mikkeli (located in the southeast), the shortest distance to business and technology (available in Lappeenranta) is 105 kilometres, the shortest distance to education, humanities, social sciences, natural sciences, medicine and 'other field' (available in Jyväskylä) is 114 kilometres and the shortest distance to arts and law (available in Helsinki) is 231 kilometres.

Table 1 presents the average shortest distances to enrol for the 10 field alternatives among the 'choosers' and 'non-choosers', i.e., those who graduated from these fields and those who did not, with and without including the polytechnic graduates in the latter group. With only two exceptions – education and humanities – the average distance to a field alternative is shorter among the choosers than among the non-choosers, regardless of whether the polytechnic graduates are included in the sample. Thus, the descriptive evidence generally suggests a negative association between distance and field-of-study choice. In five cases – business, social sciences, law, technology and 'other field' – the negative distance gap between the choosers and non-choosers is also statistically significant, while a significantly positive distance gap is only observed in the case of education. The association between distance and the choice is clearly the strongest in the case of law: the average student who graduated from a law programme had a 41% shorter distance to enrol in this field relative to the average student without a law degree; the remaining distance gaps are somewhat smaller in magnitude, varying between –25% (business) and +17% (education) in the sample including both the university and polytechnic graduates.

Table 2 describes the relationship between the shortest distance to enrol and four other regional variables, including a six-category variable depicting the urbanisation level of an individual's region of origin (ranging from a 'sparsely populated region' to a 'metropolitan area') and three institutional

---

figures for the average polytechnic graduate are 24 and 4.9. Furthermore, the data provide evidence that Finnish students are rather immobile: 44% of university graduates completed their degrees at the study location nearest their places of origin, and 68% graduated from one of the three nearest study locations. The average distance between a university graduate's study location and region of origin is 167 kilometres.

18 Examples of narrower majors and study programmes within the broad field categories are presented in Appendix 2.

TABLE 1 Descriptive evidence of the impact of distance on field-of-study choice: the average shortest distance to enrol in a field for those who chose the field (choosers) and for those who did not (non-choosers)

Field of study	Average shortest distance to enrol				
		Non-choosers			
	Choosers	University graduates only		University and polytechnic graduates	
Education	72	57	***	61	***
	(473)	(2925)		(7040)	
Arts	127	143		144	
	(94)	(3304)		(7419)	
Humanities	68	64		68	
	(511)	(2887)		(7002)	
Business	53	69	***	71	***
	(349)	(3049)		(7164)	
Social sciences	56	67	**	70	***
	(310)	(3088)		(7203)	
Law	99	169	***	167	***
	(148)	(3250)		(7365)	
Natural sciences	71	72		75	
	(417)	(2981)		(7096)	
Technology	84	96	***	96	***
	(732)	(2666)		(6781)	
Medicine	75	77		80	
	(243)	(3155)		(7270)	
Other field	113	147	***	149	***
	(121)	(3277)		(7392)	

Notes: The sample sizes for the groups are in parentheses. A significant difference in a mean between the choosers and non-choosers (according to a t-test with unequal variances) is indicated by * (p<.1), ** (p<.05) or *** (p<.01).

attributes of the nearest study location of a field: the number of student places, the admission percentage and the number of majors.[19] In most cases, a negative

[19] The information on student places and admission percentages was collected from the KOTA database of the Ministry of Education and Culture, which includes panel data on Finnish universities at the field-of-study level. The number of admitted students was used to approximate the number of student places, which is justified because of the constant excess demand for student places throughout the university system. The admission percentage was again constructed by dividing the number of admitted students by the number of applicants. These variables were constructed separately for each year between 1991 and 1996, after which they were matched with the micro-data based on the individuals' high school graduation years; thus, the estimates for these variables are partly identified through variation over time. Compared to the first two variables, the variable depicting the number of majors at the nearest study location is cruder: because there was no collective, historical dataset available, this variable was constructed based on the current year's (2013) information obtained from www.koulutusnetti.fi regarding the application options in different universities and fields. A brief analysis of the regional variation in these institutional variables reveals that the primary difference was between those having Helsinki as the nearest study location and the others: both the average number of student places and the

TABLE 2 Regional variables and their correlation with the shortest distance to enrol

	Region-of-origin urbanisation level		Attributes of the nearest study location					
			Number of student places		Admission percentage		Number of majors	
Mean	3.06		418		0.23		9.77	
Standard deviation	1.58		426		0.15		10.09	
Correlation with the shortest distance to enrol								
Education	-0.60	***	-0.12	***	-0.13	***	-0.17	***
Arts	-0.52	***	-0.42	***	0.07	***	-0.44	***
Humanities	-0.61	***	-0.39	***	0.17	***	-0.32	***
Business	-0.59	***	-0.45	***	-0.19	***	0.12	***
Social sciences	-0.53	***	-0.42	***	0.26	***	-0.36	***
Law	-0.54	***	-0.26	***	0.16	***	-0.31	***
Natural sciences	-0.60	***	-0.28	***	-0.02		-0.29	***
Technology	-0.59	***	-0.31	***	0.25	***	-0.35	***
Medicine	-0.60	***	-0.23	***	0.18	***	-0.10	***
Other field	-0.48	***	-0.47	***	-0.38	***	-0.47	***

Notes: A significant correlation is indicated by * (p<0.1), ** (p<0.05) or *** (p<0.01). [a] For the calculations, integer values were assigned to the urbanisation level categories as follows: 0 = sparsely populated region; 1 = rural region; 2 = industrial centre; 3 = regional centre; 4 = large university city; 5 = metropolitan area.

and significant correlation exists between the shortest distance to enrol and the other regional variables. In particular, because all of the universities are located in large- or middle-sized cities, a high urbanisation level in the region of origin is strongly and systematically associated with a short distance to a field. Furthermore, the correlations in almost all cases indicate that if the nearest study location is far away, it is likely that the university unit in this location is relatively small in size and has a relatively small number of majors; as the only exception, the correlation for the number of majors is positive in the case of business. However, most of the correlations for the admission percentage suggest that students from remote regions, on average, face a less competitive admissions process in the nearest study location than students from less remote regions; however, in the case of three fields (education, business and 'other field'), the opposite (negative) relationship between the admission percentage and the shortest distance to enrol in observed, and for natural sciences, the correlation is approximately zero. Thus, given their significant correlations with distance, the additional institutional attributes may serve as relevant controls for other supply restrictions that may affect students' study choices. For instance, a low admission percentage can be assumed to be positively

average number of majors was more than twice as large for the first group. However, variations in the admission percentage were rather modest across regions.

associated with the 'selectivity' or 'prestige' of the institution, and thus, this variable can be used to control for regional variation in institutional quality.[20]

An important fact that must be considered in the current analysis is that the study options treated as choice alternatives are fundamentally different in many ways, for example, with respect to expected pecuniary and non-pecuniary job characteristics, skill sets required and the difficulty of admission, resulting in significant sorting based on individual characteristics across these characteristics. This heterogeneity is illustrated in Table 2, which presents the average annual earnings in 2006[21] and the sample averages for several individual characteristics of the 15 choice alternatives. Based on the earnings information, large differences exist in the 'lucrativeness' of the alternatives. The university-level field alternatives can be roughly divided into three categories: business, law, technology and medicine are highly lucrative fields (with average earnings of approximately 50,000 €); social sciences, natural sciences and 'other field' are medium-pay fields (with average earnings of approximately 35,000 €); and education, arts and humanities are non-lucrative fields (with average earnings below 30,000 €). Of the polytechnic-level alternatives, technology is the only one with relatively high average earnings (37,300 €), whereas the remaining fields clearly belong to the non-lucrative category (with average earnings of approximately 25,000 €). For the sake of brevity, a comprehensive analysis of the effects of different individual characteristics on study choices is omitted from the paper. In brief, the sample means reported in Table 3 suggest that gender, first language, matriculation grades, family background and the type of the region of origin may all be significantly associated with individuals' choices. However, these results primarily reflect the considerable sorting of students across the university and polytechnic sectors based on ability and family background: compared to the average polytechnic student, the average university student has, almost regardless of the field of study, higher grades in both first language and mathematics and parents with higher socioeconomic status (as indicated by the higher average educational levels and lower frequencies observed for the category 'worker or other').

---

20 The state-funded university system has clearly promoted small quality differences between universities; nevertheless, certain evidence of regional differences, for example, in university students' average matriculation grades has been presented (see Suhonen, 2013). In particular, the University of Helsinki is generally considered Finland's flagship university, and consequently, it has been more able to attract students with high grades than the other universities. Thus, in the presence of peer effects, it is unlikely that the quality of university education has been equally distributed across space.

21 The mean earnings were estimated from the full Statistics Finland micro-dataset by including all the individuals under the age of 55 in the estimation sample. The earnings include both wage and entrepreneurial income.

TABLE 3 Average annual earnings (2006) and individual characteristics by field-of-study alternative

	University-level alternatives										Polytechnic-level alternatives				
	Educ.	Arts	Hum.	Bus.	Soc.	Law	Nat.	Tech.	Med.	Other	Hum.	Bus.	Tech.	Health	Other
Annual earnings in 2006 [a]	28,592	25,039	27,419	49,698	33,244	51,815	35,764	51,705	54,231	39,323	19,917	26,602	37,303	23,868	25,900
Female	0.83	0.59	0.79	0.49	0.72	0.53	0.52	0.20	0.67	0.44	0.71	0.68	0.17	0.87	0.65
Swedish speaker	0.07	0.02	0.07	0.12	0.09	0.07	0.07	0.05	0.09	0.03	0.05	0.05	0.06	0.04	0.05
First language grade [b]	4.89	5.17	5.47	5.04	5.33	5.29	5.08	5.01	5.34	5.04	4.84	4.37	4.00	4.51	4.44
A-level math grade [b]	3.67	4.52	4.18	4.60	4.51	4.40	4.89	5.30	4.92	4.31	3.93	3.16	3.53	3.31	3.37
B-level math grade [b]	4.03	3.95	4.55	4.94	4.56	4.81	4.84	4.79	5.11	4.59	3.90	3.60	3.97	3.39	3.46
Mother's educational level [c]	3.22	3.95	3.50	3.58	3.59	3.95	3.65	3.79	3.89	3.97	3.10	2.83	2.98	2.88	2.99
Mother's occupational status															
Farmer/entrepreneur	0.10	0.10	0.10	0.13	0.11	0.11	0.13	0.14	0.14	0.16	0.11	0.13	0.12	0.10	0.13
High-/low-ranking official	0.72	0.82	0.74	0.75	0.76	0.82	0.75	0.74	0.76	0.74	0.70	0.64	0.66	0.63	0.65
Worker/other	0.18	0.09	0.16	0.12	0.13	0.07	0.12	0.12	0.10	0.10	0.19	0.22	0.23	0.26	0.23
Father's educational level [c]	3.33	4.01	3.66	4.05	3.95	4.16	3.86	4.20	4.07	3.97	3.32	2.91	3.14	2.85	2.97
Father's occupational status															
Farmer/entrepreneur	0.20	0.16	0.16	0.20	0.16	0.18	0.16	0.16	0.19	0.21	0.13	0.21	0.19	0.18	0.20
High-/low-ranking official	0.53	0.67	0.60	0.63	0.64	0.67	0.63	0.66	0.64	0.58	0.56	0.45	0.48	0.42	0.46
Worker/other	0.27	0.17	0.24	0.17	0.20	0.16	0.21	0.18	0.16	0.22	0.31	0.34	0.33	0.40	0.34
Region of origin															
Metropolitan area	0.13	0.34	0.19	0.34	0.26	0.43	0.17	0.26	0.28	0.26	0.26	0.27	0.19	0.20	0.25
Urban region	0.47	0.44	0.44	0.37	0.48	0.42	0.50	0.46	0.43	0.38	0.44	0.40	0.39	0.45	0.42
Small region	0.40	0.22	0.36	0.29	0.26	0.16	0.33	0.28	0.29	0.36	0.31	0.32	0.41	0.36	0.33
N	473	94	511	349	310	148	417	732	243	121	250	1578	933	995	359

Notes: [a] The average earnings were estimated from the full Statistics Finland dataset (7% random sample from Finland's population) with including all the individuals under the age of 55 in the estimation sample. [b] The grades are from the Finnish matriculation examination. For the calculations, integer values were assigned to the grades as follows: I = 1; A = 2; B = 3; C = 4; M = 5; E/L = 6. [c] Mother's and father's education levels take values from 1 (primary school) to 7 (doctoral degree).

5 Methodology

Although the descriptive evidence presented in Section 4 suggests that individuals' study choices are associated with geographical distance, the economic magnitude of the effect of distance cannot be directly inferred from this evidence. In addition, one may suspect that the evidence based on group means is confounded because only the outcome of a constrained choice – that is, a student's educational field and sector after being admitted to study – is observed in the data. Therefore, the differences in the group means could partly arise from differences in abilities or opportunities rather than the students' revealed preferences across the choice alternatives. A particular problem is that a student's region of origin is not randomly assigned but rather the result of a choice made by his or her family of origin. Thus, correlations between the distances and study choices could partly reflect an uneven spatial distribution of various individual characteristics – related to either students' preferences or opportunities – resulting from families' selective locational choices (e.g., Denzler & Wolter, 2011; Gibbons & Vignoles, 2012). Furthermore, as illustrated in Section 4, regional heterogeneity in the universities' institutional characteristics, such as the difficulty of admission and diversity of programmes available, constitute an additional potential problem for the identification of the shortest distance effect.

Similar to the studies discussed above, the current analysis utilises multinomial discrete choice methods to assess the distance effect. As most of the earlier work, the current analysis is limited to a cross-sectional analysis: the choice of an educational alternative is modelled in a static discrete choice framework, and individual- and alternative-specific control variables are used to alleviate problems arising from regional heterogeneity in demand- and supply-side factors. In the standard manner, the choice model employed in the analysis is specified by assuming that individual i's study choice y_i is determined by the maximisation of latent utility U_{ij} over J alternatives indexed $j = 1,2, \dots, J$. That is, the observed choice $y_i = k$ is assumed to satisfy $U_{ik} > U_{ij}$ for all $j \neq k$. The baseline results are obtained by assuming that the utilities are of the following form:

$$U_{ij} = \alpha_j + \beta d_{ij} + \gamma_j' x_i + \vartheta_{ij}, \tag{1}$$

where α_j is an alternative-specific constant term; d_{ij} is the shortest distance from i's region of origin to alternative j; x_i is a vector of individual-level control variables; and ϑ_{ij} is the error term. The parameter of interest in equation (1) is β; based on the discussion above, an increase in the distance to an educational alternative is likely to decrease the attractiveness of that alternative, and thus, this parameter is expected to take a negative sign. To assess the heterogeneity of the distance effect across the choice alternatives, a version of equation (1) in which β is replaced with an alternative-specific coefficient β_j is also considered during the analysis.

As discussed in Section 4, the choice model includes two types of alternatives: 10 university-level alternatives and 5 polytechnic-level alternatives. As the current focus in on the effect of distance on one's choice of a university-level alternative, parameter β is only identified through variation in d_{ij} across the university-level alternatives by restricting the value of this variable to zero for each polytechnic-level alternative. This restriction might be problematic if significant regional variation existed in the supply of the polytechnic-level alternatives and if this variation were correlated with that in the university-level alternatives. However, this problem appears unlikely because all of the polytechnic-level alternatives were available in nearly all of the 19 NUTS-3 regions in Finland; as the only two exceptions, the regions of Kainuu and Ahvenenmaa have not provided polytechnic education in the humanities (Statistics Finland, 2013). The inclusion of additional 'polytechnic dummy variables', i.e., indicators for the presence of a polytechnic offering a particular field in one's region of origin, was also attempted, but these indicators were insignificant and did not affect the results regarding the distance effects.[22]

When estimating the choice model, a rich variety of individual-specific variables is employed to control for individual heterogeneity by region of origin. Detailed descriptions of these variables are included in Appendix 1. The basic controls include a female indicator, a Swedish speaker indicator, the high school graduation year, the matriculation grades in first language and mathematics, the mother's and father's educational levels and occupational statuses and indicators for whether one of the parents received a university degree in a particular field. The first three variables can be considered rather neutral controls, the purpose of which is to smooth differences in the basic demographic characteristics and the timing of the choice.[23] The role of grades and parental characteristics may again be crucial for identifying the distance effect: the parental characteristics in particular may capture the intergenerational transmission of educational outcomes through genetic and environmental factors (see, e.g., Haveman & Wolfe, 1995), whereas the matriculation grades may effectively control for the remaining regional variation, for example, in students' chances of being admitted to different fields and preferences across different types of alternatives (e.g., mathematical and non-mathematical fields). Aside from the individual and parental characteristics, the regional variables discussed in Section 4 are employed as

---

22 The information on the supply of polytechnic education in the NUTS-4 regions was collected from the StatFin database of Statistics Finland. The proportion of individuals originating from a region offering polytechnic education in a particular field was generally high – 68% for humanities, 85% for business, 79% for technology, 77% for health and 73% for 'other field' – which reflects the highly decentralised nature of the polytechnic sector.

23 Age at the time of high school graduation is not included in the demographic controls, as this variable varies very little in the sample: 98% of the students were either 19 or 20 years old in that year.

additional controls for supply and demand conditions in the students' neighbourhoods.

Two alternative approaches are employed to estimate the choice model: a conditional logit approach and a nested logit approach. In the conditional logit approach, the unobserved utility components ϑ_{ij} are assumed to be independently and identically distributed (type 1 extreme value), which implies that the choice probability of alternative j is of the following form:

$$\mathrm{P}_{ij}^{CL} = \frac{e^{V_{ij}}}{\sum_{k=1}^{J} e^{V_{ik}}}, \tag{2}$$

where $V_{ij} = \alpha_j + \beta d_{ij} + \gamma_j' x_i$, that is, the deterministic component of utility U_{ij}. While the conditional logit approach enables rapid and simple estimation, the choice probabilities in this approach are known to have the dubious independence-of-irrelevant-alternatives (IIA) property. In the current context, the IIA implies that when the distance to a choice alternative changes, the probabilities of the remaining alternatives change in equal proportion. Obviously, this assumption is strong given that the substitutability between the alternatives likely varies according to certain unobserved characteristics of the individuals and alternatives. For instance, an increase in the distance to humanities could cause individuals having strong preferences for 'human sciences' to substitute education or social sciences, rather than natural sciences or technology, for humanities. If the control variables do not adequately capture these types of substitution patterns, then the IIA is violated.

The nested logit approach is employed to assess the robustness of the results to substitution patterns arising from unobservables, that is, to relaxing the IIA. The basic concept of this approach is to group the choice alternatives into M nests based on certain 'similarity criteria' and assume that the choice probability of alternative j belonging to nest m is of the following form:

$$P_{ij}^{NL} = P_{im} \times P_{ij|m} = \frac{e^{\tau_m IV_m}}{\sum_{l=1}^{M} e^{\tau_l IV_l}} \times \frac{e^{V_{ij}/\tau_m}}{\sum_{k=1}^{J_m} e^{V_{ik}/\tau_m}}, \tag{3}$$

where τ_m is a 'dissimilarity parameter', i.e., an inverse measure of the correlation among the J_m alternatives in nest m, whereas $IV_m = ln \sum_{l=1}^{J_m} e^{V_{il}/\tau_m}$ is an 'inclusive value' depicting the expected utility value that i obtains by selecting nest m (see, e.g., Heiss, 2002; Train, 2003). In contrast to conditional logit models, estimating nested logit models is generally very time-consuming, and the estimation time is highly sensitive to the complexity of the nest structure. Therefore, the current analysis exclusively focuses on simple models with two nests. Three alternative nesting structures are considered. The first nesting structure loosely follows the classical division of academic fields into 'soft and hard sciences' based on the type of knowledge that the fields produce (see, e.g., Neumann *et al.*, 2002); six of the university-level alternatives, education, arts, humanities, business, social sciences and law, and two polytechnic-level alternatives, humanities and business, are specified as soft fields; whereas the remaining alternatives are specified as hard fields (see Appendix 2). The second nesting structure is again based on the pecuniary aspect of field-of-study choice discussed in many previous studies (Berger, 1988; Montmarquette *et al.*, 2002; Bourdarbat, 2008; Beffy *et al.*, 2012): based on the figures reported earlier in Table 3, the fields with average earnings above 35,000 € are specified as lucrative fields; whereas the remaining fields are specified as non-lucrative fields. In the last nesting approach, the field alternatives are again divided based on the educational level/sector into university-level and polytechnic-level fields. As individuals' unobserved abilities and preferences may vary across space and match heterogeneously with the type of knowledge, economic prestige and educational level/sector of the choice alternatives, all three nesting structures could represent substitution patterns that confound the baseline estimates.

As a final methodological note, a word of caution is in order because the current analysis is based on individuals' completed education, instead of initial enrolment decisions: if there were mechanisms generating correlation between an individual's region of origin and the probability to drop out and/or to switch programmes prior to graduation, then the estimates of the distance effects could suffer from selection bias.[24] However, one may argue that, in the context of the Finnish university system, the determinants of success in enrolment and success in graduation are likely to be relatively similar due to the competitive admission system, in which a student's success is based on his/her ability and motivation demonstrated in the matriculation examination and entrance examinations. Therefore, the solutions to both selection problems – selection in enrolment and selection in graduation – are likely to depend on the question of whether the observed individual characteristics (such as the matriculation grades) adequately capture variation, for example, in ability, motivation and preferences, across students originating from different regions. It is, of course, possible that the severity of the sample selection problem differs across fields, depending on how difficult completing a degree is relative to earning admission. Based on this hypothesis, the selection problem is likely to be the most severe in the case of natural sciences because the studies in this field are, in the Finnish university system, generally considered 'easy to get in but difficult to complete'.[25]

---

24 Based on the current data on 1991–1996 high school graduates, the likelihood of lacking a subsequent educational qualification after high school is the lowest (13%) among individuals originating from sparsely populated regions, whereas this propensity increases almost linearly in the degree of urbanisation, being highest (23%) among high school graduates from the Helsinki metropolitan area. Consequently, a small, positive correlation exists between the distances to enrol and the probability of a post-secondary degree, suggesting that the estimated distance effects could be downward biased.

25 The statistics reveal that the discontinuation of studies is more common in natural sciences than in any of the other fields (Statistics Finland, 2013). For instance, in the 2004–2005 academic year, the discontinuation rate was 11% among natural sciences students but 7% among all university students.

6 Results

This section reports the results of the discrete choice models. The interpretation of the conditional logit results is based on the following transformation of the estimated coefficients: $exp(\hat{\beta}) - 1$, which yields the relative change in $P_{ij}/(1 - P_{ij})$, i.e., the odds of choosing alternative j, associated with a 100-kilometre increase in the shortest distance to enrol. Because the number of choice alternatives is relatively large in the current models, $1 - P_{ij}$ is generally very close to one, and thus the relative change in the odds fairly accurately approximates the 'semi-elasticity' $\partial lnP_{ij}/\partial d_{ij}$. However, in the nested logit models, the magnitudes of the effects are less straightforward to interpret because they are partly determined by the strength of the correlations within the nests; therefore, the average semi-elasticities for the study sample are presented, along with the (non-transformed) coefficient estimates, when reporting the nested logit results.

First, simple conditional logit models are employed to assess the average effect of the shortest distance to enrol on the choice of a university-level field alternative. Table 4 reports the sensitivity of this estimate to different sets of control variables and the inclusion of the polytechnic-level alternatives in the model's choice set. The simplest model specification (column 1, row 1) only controls for the alternative-specific constant terms, while the polytechnic-level alternatives are excluded from the choice set; according to the estimates obtained from this specification, a 100-kilometre increase in the shortest distance to enrol in a field is associated with a 20.6% average decrease in the odds of choosing that field. However, when the polytechnic-level alternatives are included in the choice set (column 1, row 2), the implied effect of the distance variable is somewhat smaller, –16.6%. This result indicates that the estimate obtained using the restricted choice set (the university-level alternatives only) is upward biased, reflecting that many individuals ultimately pursued their educations in the polytechnic sector despite residing relatively near the university-level alternatives.

A comparison of the estimates across the first three columns of Table 4 reveals that the estimate for the shortest distance to enrol is fairly robust to controls for the basic demographic characteristics (gender, mother tongue and year of high school graduation) and the matriculation grades. Thus, for example, regional variation in mathematical and linguistic skills does not appear to significantly explain the association between distance and study choice. However, column 4 demonstrates that the results are somewhat sensitive to the inclusion of the parental characteristics, but only when the polytechnic-level alternatives are included in the choice set; in this case, a clearly smaller estimate, –9.1%, is obtained for the shortest distance to enrol than previously. However, when the restricted choice set is employed, only a small change (from –21.1% to –20.1%) is observed in this estimate. These findings suggest that there are family-background-related mechanisms – such as the intergenerational transmission of educational outcomes through 'nature' or 'nurture' (see Haveman and Wolfe, 1995) – that partly explain the relationship between distance and individuals' study choices. However, as the estimate obtained using the restricted choice set is rather insensitive to the parental controls, it appears that family-related heterogeneity is more strongly associated with the 'vertical' choice between the university- and polytechnic-level alternatives than the 'horizontal' choice among the university-level alternatives.

TABLE 4 Relative change in the odds of choosing a field associated with a 100-kilometre increase in the shortest distance to enrol in the field: estimates from a conditional logit model

Sample used	(1)		(2)		(3)		(4)		(5)	
University graduates	-.206	***	-.200	***	-.212	***	-.201	***	-.195	***
(10 choice alternatives, 3398 individuals)	(.039)		(.040)		(.037)		(.034)		(.033)	
University and polytechnic graduates	-.166	***	-.158	***	-.146	***	-.091	***	-.145	***
(15 choice alternatives, 7513 individuals)	(.021)		(.021)		(.021)		(.031)		(.029)	
Controls included:										
Alternative-specific constants	Yes		Yes		Yes		Yes		Yes	
Demographic variables	No		Yes		Yes		Yes		Yes	
Matriculation grades	No		No		Yes		Yes		Yes	
Parents' characteristics	No		No		No		Yes		Yes	
Region-of-origin urbanisation level	No		No		No		No		Yes	

Notes: The standard errors (in parentheses) are clustered at the level of 81 regions of origin. Significant estimates are indicated by * (p<.1), ** (p<.05) or *** (p<.01). The set of demographic variables includes a woman indicator, a Swedish speaker indicator and high school graduation year. The set of parents' characteristics includes mother's and father's educational levels and occupational statuses and indicators for whether one of the parents graduated with a university degree from a particular field.

In the fifth column of Table 4, the choice model is further augmented with a six-category variable depicting the degree of urbanisation in the individual's region of origin. Unlike the parental controls, the regional controls shift the estimate for the shortest distance to enrol upwards in the model that includes both the university- and polytechnic-level alternatives; in this specification, the implied effect of a 100-kilometre increase in the shortest distance to enrol is –14.5%. When only including the university-level alternatives in the choice set, the estimate is again rather insensitive to the additional controls (the estimate is reduced to –19.5%). Thus, there appear to be certain fundamental differences in study choices between urban and rural students, but these are primarily reflected in the 'vertical' choice between the university- and polytechnic-level alternatives. The finding that omitting the degree of urbanisation results in a downward bias in the estimated distance effect implies that a positive relationship must exist between the urbanisation level and one's propensity to substitute a polytechnic-level alternative for a university-level alternative. This interpretation is consistent with the observation that the urbanisation level is negatively correlated with the distances to the university-level alternatives. This behavioural pattern could be explained by differences in mobility across urban and rural areas; in particular, students from the Helsinki metropolitan area are known to be highly unlikely to change their residential location after high school (see Suhonen, 2013) and, to avoid relocation, could also be relatively eager to substitute between the higher education sectors. However, explanations based on regional differences in individuals' educational preferences or supply-side factors cannot be dismissed.[26]

An important conclusion from Table 4 is that the results regarding the distance effect are somewhat sensitive to whether the study options in the polytechnic sector are included in the analysis. Therefore, these polytechnic-level alternatives are included in all of the remaining estimations. The first supplemental analysis assesses the robustness of the shortest distance effect to controlling for three other attributes of the nearest study location – the number of student places, the admission percentage and the number of majors. Table 5 demonstrates that the association of these additional variables with field-of-study choice is rather weak. When the model does not control for the degree of urbanisation in the region of origin (the left column), the number of student places has a significant and negative coefficient estimate, indicating (counter-intuitively) that increasing the number of student places in a given field at the nearest study location by 100 is associated with a 2.5% decrease in the odds of selecting that field. However, after controlling for the urbanisation level (the right column), this estimate is reduced by half and becomes insignificant. The estimates for the two remaining variables (the admission percentage and the number of majors) are again approximately zero and insignificant in both specifications. Consequently, the impact of the additional controls on the estimated shortest distance effect is also negligible: in the left column of Table 5, this estimate is slightly larger than previously observed (–10.1%) and that in the right column is approximately identical to that obtained previously. Thus, the results suggest that, in line with the results presented above, if the estimated shortest distance effect is biased because of omitted regional variables, then the bias is likely to be downward. Furthermore, the results indicate that the additional regional controls are redundant after controlling for the degree of urbanisation in the student's region of origin; therefore, these variables are excluded from the remaining estimations.

---

26 In addition to the models in Tables 4 and 5, several other versions of the baseline conditional logit model were estimated as robustness checks. For instance, in place of the shortest distance to enrol, an indicator for whether a given field is available in the individual's nearest university region was used. In line with the baseline model, the coefficient estimate for this variable was significantly positive and implied that having a field available in the nearest university region is associated with a 26% increase in the odds of selecting that field. In another supplemental analysis, the shortest distance effect was allowed to differ across genders; the estimate obtained was slightly larger for women (–15.2%) than for men (–13.7%), but the gender difference was not statistically significant. Furthermore, when both the linear and quadratic terms of the shortest distance to enrol were included in the choice model, the estimate for the latter term was close to zero and insignificant, yielding no strong support for nonlinearity in the shortest distance effect.

TABLE 5 Estimates from conditional logit models controlling for other attributes of the nearest study location (sample size: 15 choice alternatives, 7513 individuals)

Variable	(1)		(2)	
Shortest distance to enrol (/100 km)	-.101	***	-.144	***
	(.023)		(.030)	
Student places in the nearest study location (/100)	-.025	***	-.012	
	(.008)		(.009)	
Admission percentage in the nearest study location	.001		.002	
	(.003)		(.003)	
Majors in the nearest study location	-.003		.005	
	(.004)		(.004)	
Region-of-origin urbanisation level controlled for	No		Yes	

Notes: The estimates are to be interpreted as relative changes in the odds of choosing a field associated with unit increases in the variables. The standard errors (in parentheses) are clustered at the level of 81 pre-university regions. Significant estimates are indicated by * (p<.1), ** (p<.05) or *** (p<.01). The basic set of control variables is the same as in column 4 of Table 4.

In the second supplemental analysis, heterogeneity in the distance effect across fields is examined by estimating a conditional logit model, in which the coefficient for the distance variable is allowed to vary across the alternatives. In most cases, the findings from this model – reported in Table 6 – are consistent with a 'distance deterrence effect' in field-of-study choice; however, a certain degree of heterogeneity across fields is also detected. A significantly negative distance effect (at the 10% level) is observed for humanities (–14.9%), social sciences (–19.6%), law (–20.4%), natural sciences (–25.3%) and 'other field' (–31.5%). Sizeable negative estimates are also obtained for business and technology (–11.6% and –10.6%), but the p-values for these estimates are above .10 because of the large standard errors. For the three remaining fields – education, arts and medicine – the estimates are again close to zero and insignificant, suggesting that distances do not affect enrolment in these fields. As indicated by the superscripts in Table 6, many of the differences between the field-specific estimates are also statistically significant; in particular, the estimates for natural sciences and 'other field' are significantly different (at the 10% level) from those for education, arts, technology and medicine.

In the final stage of the analysis, the robustness of the conditional logit estimates to relaxing the IIA assumption is studied using the nested logit approach. Based on the previous findings regarding the heterogeneity in the distance effect, it is reasonable, at this stage, to concentrate on models that allow the effects to vary across fields. Table 7 presents the coefficient estimates for the

TABLE 6 Heterogeneity in the shortest distance effect across fields: estimates from a conditional logit model (sample size: 15 choice alternatives, 7513 individuals)

Shortest distance to enrol X			
Education	.004		n, o
	(.100)		
Arts	.046		h, l, n, o, s
	(.109)		
Humanities	-.149	*	a
	(.078)		
Business	-.114		
	(.085)		
Social sciences	-.196	*	a
	(.095)		
Law	-.204	**	a
	(.082)		
Natural sciences	-.253	***	a, e, m, t
	(.050)		
Technology	-.106		n, o
	(.065)		
Medicine	.020		n, o
	(.105)		
Other field	-.315	***	a, e, m, t
	(.095)		

Notes: The estimates are to be interpreted as relative changes in the odds of choosing a field associated with a 100-kilometre increase in the shortest distance to enrol in the field. Standard errors (in parentheses) are clustered at the level of 81 regions of origin. A significant estimate is indicated by * (p<.1), ** (p<.05) or *** (p<.01). A significant difference (p < .1) to the estimate of another field is indicated by [a] (arts), [b] (business), [e] (education), [h] (humanities), [l] (law), [m] (medicine and health sciences), [n] (natural sciences), [o] (other field), [s] (social sciences) or [t] (technology). The set of control variables is the same as in column 5 of Table 4.

shortest distance to enrol and the dissimilarity parameters obtained with three alternative nesting structures: 'hard and soft fields', 'lucrative and non-lucrative fields' and 'university- and polytechnic-level fields'. As the dissimilarity parameters are not significantly different from one, separately or jointly, in any of the three specifications, no strong evidence against the IIA is obtained. Only in two cases – the nests for 'hard fields' and 'university-level fields' – is the point estimate for the dissimilarity parameter notably below one, an indication of a within-nest correlation in unobservables. However, because the confidence intervals for the dissimilarity parameters are considerably large, the existence of

TABLE 7 Estimates from nested logit models with alternative nesting structures (sample size: 15 choice alternatives, 7513 individuals)

Variable	Nesting structure					
	N1: soft sciences N2: hard sciences		N1: lucrative N2: non-lucrative		N1: university N2: polytechnic	
Shortest distance to enrol X						
Education	.006		.019		-.009	
	(.110)		(.177)		(.095)	
Arts	.046		.078		.027	
	(.121)		(.201)		(.102)	
Humanities	-.163		-.267		-.144	
	(.110)		(.172)		(.091)	
Business	-.122		-.144		-.117	
	(.115)		(.131)		(.095)	
Social sciences	-.222	*	-.334	*	-.191	*
	(.121)		(.187)		(.103)	
Law	-.234	**	-.264	**	-.199	**
	(.106)		(.110)		(.090)	
Natural sciences	-.261	***	-.342	***	-.255	**
	(.090)		(.120)		(.101)	
Technology	-.103		-.115		-.106	*
	(.062)		(.079)		(.060)	
Medicine	.017		.028		.008	
	(.100)		(.131)		(.096)	
Other field	-.337	***	-.453	***	-.321	**
	(.110)		(.158)		(.134)	
Dissimilarity parameter						
Nest 1	1.030		1.186		.840	
Nest 2	.881		1.732		.993	
Likelihood ratio test for the IIA (p-value)	.463		.191		.668	

Notes: The coefficient estimates are to be interpreted as changes in the utility from choosing a field associated with a 100-kilometre increase in the shortest distance to enrol in the field. Standard errors are in parentheses. A significant estimate is indicated by * (p<.1), ** (p<.05) or *** (p<.01). The set of control variables is the same as in column 5 of Table 4.

these correlations cannot be confirmed.[27] In the case of the second nesting structure – the division of the fields into 'lucrative and non-lucrative fields' –

27 Obviously, the lack of correlation in unobservables could arise either because the alternatives within the nests are not very similar initially or because the control variables included in the models already capture the primary substitution patterns across alternatives. To assess which explanation is more likely, the nested logit models were also estimated without the use of control variables. These models suggested strong correlation in unobservables within three nests: 'university-level fields', 'polytechnic-level fields' and 'lucrative fields'. For the remaining nests, the dissimilarity parameters were again above one, suggesting no correlation. Thus, it appears that at least some of the nests contain highly similar alternatives, and therefore, the lack of correlation partly arises from the use of appropriate control variables.

the dissimilarity parameters are clearly above one, indicating that this model specification is inconsistent with random utility maximisation (see Heiss, 2002). Naturally, because of the lack of strong correlation in unobservables, the nested logit results regarding the distance effects are also qualitatively very similar to the conditional logit results presented above (Table 6). The results of each specification suggest that, in the case of social sciences, law, natural sciences and 'other field', an increase in the shortest distance to enrol is negatively and significantly associated with the utility obtained from choosing the field. The negative estimates for humanities and technology are also weakly significant in each specification (with p-values systematically below .15).

To obtain a better sense of the effects and substitution patterns implied by the nested logit models, the average semi-elasticities $\partial lnP_{ij}/\partial d_{ij}$ are calculated based on two models – those using the 'soft-hard' and 'university-polytechnic' nesting structures. From Table 8, we observe that the implied, average semi-elasticities are nearly identical in the two models because of the low correlations in unobservables. In most cases, the 'university-polytechnic' model yields a slightly larger average own-distance elasticity for a university-level alternative than the 'soft-hard' model; the average elasticities across alternatives are –14.7% and –13.8% in the two models. Obviously, this result arises because more substitution between the university-level alternatives occurs in the 'university-polytechnic' model. In comparison with the nested logit models, the average semi-elasticity is slightly smaller, –13.7%, in the corresponding conditional logit model (Table 6). Thus, there is some indication of downward bias in the distance effect estimates relying on the IIA assumption.

7 Summary and concluding remarks

A vast literature has studied the determinants of educational choices. This paper contributes to this literature by presenting evidence of a 'distance deterrence effect' in the context of selecting a field of study in the Finnish university system. The estimated distance effect is sizeable: the baseline conditional logit model controlling for various individual, parental and regional characteristics suggested that a 100 kilometre longer distance required to enrol in a field is, on average, associated with a 15% lower likelihood of selecting that field. However, the results also indicated that the impact of distance varies, to some extent, depending on the choice alternative in question: a student's decision to study education, arts or medicine was not found to be sensitive to the distances to these fields; whereas for the remaining fields, the estimated distance effects were generally large and statistically significant.

TABLE 8 Shortest distance effects implied by two nested logit models: average semi-elasticities

Shortest distance to enrol X	Nested logit model	Average change in log choice probability														
		University-level alternatives										Polytechnic-level alternatives				
		Educ.	Arts	Hum.	Bus.	Soc.	Law	Nat.	Tech.	Med.	Other	Hum.	Bus.	Tech.	Health	Other
Education	soft-hard	**.006**	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	univ-poly	**-.009**	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001
Arts	soft-hard	-.001	**.044**	-.001	-.001	-.001	-.001	-.001	-.001	-.001	-.001	-.001	-.001	-.001	-.001	-.001
	univ-poly	-.001	**.032**	-.001	-.001	-.001	-.001	-.001	-.001	-.001	-.001	.000	.000	.000	.000	.000
Humanities	soft-hard	.011	.011	**-.148**	.011	.011	.011	.011	.011	.011	.011	.011	.011	.011	.011	.011
	univ-poly	.014	.014	**-.157**	.014	.014	.014	.014	.014	.014	.014	.010	.010	.010	.010	.010
Business	soft-hard	.005	.005	.005	**-.113**	.005	.005	.006	.006	.006	.006	.005	.005	.006	.006	.006
	univ-poly	.008	.008	.008	**-.131**	.008	.008	.008	.008	.008	.008	.005	.005	.005	.005	.005
Social sciences	soft-hard	.009	.009	.009	.009	**-.207**	.009	.009	.009	.009	.009	.009	.009	.009	.009	.009
	univ-poly	.011	.011	.011	.011	**-.216**	.011	.011	.011	.011	.011	.008	.008	.008	.008	.008
Law	soft-hard	.004	.004	.004	.004	.004	**-.223**	.005	.005	.005	.005	.004	.004	.005	.005	.005
	univ-poly	.005	.005	.005	.005	.005	**-.231**	.005	.005	.005	.005	.004	.004	.004	.004	.004
Natural sciences	soft-hard	.014	.014	.014	.014	.014	.014	**-.279**	.018	.018	.018	.014	.014	.018	.018	.018
	univ-poly	.019	.019	.019	.019	.019	.019	**-.284**	.019	.019	.019	.014	.014	.014	.014	.014
Technology	soft-hard	.010	.010	.010	.010	.010	.010	.012	**-.105**	.012	.012	.010	.010	.012	.012	.012
	univ-poly	.013	.013	.013	.013	.013	.013	.013	**-.113**	.013	.013	.010	.010	.010	.010	.010
Medicine	soft-hard	-.001	-.001	-.001	-.001	-.001	-.001	-.001	-.001	**.019**	-.001	-.001	-.001	-.001	-.001	-.001
	univ-poly	.000	.000	.000	.000	.000	.000	.000	.000	**.009**	.000	.000	.000	.000	.000	.000
Other field	soft-hard	.005	.005	.005	.005	.005	.005	.007	.007	.007	**-.376**	.005	.005	.007	.007	.007
	univ-poly	.007	.007	.007	.007	.007	.007	.007	.007	.007	**-.375**	.005	.005	.005	.005	.005

Note: The own-distance semi-elasticities for the fields are emboldened.

As the distance effect appears to differ across fields, potential explanations for this heterogeneity may be briefly discussed. A common feature of the zero-distance-effect fields – education, arts and medicine – is that the occupations associated with these fields, such as those of schoolteacher, artist and doctor, are often considered to require particular vocational commitment, that is, strong preferences for the type of work they involve. However, university degrees in many of the negative-distance-effect fields, such as social sciences and natural sciences, are usually more general in terms of occupational possibilities; thus, as choice alternatives, these fields are likely to require a lesser amount of vocational commitment and are likely to be more substitutable with other alternatives. Therefore, a potential explanation for the heterogeneity in the distance effect is that enrolment in education, arts or medicine requires relatively strong vocational preferences and is therefore unaffected by 'secondary factors' such as the distance to the nearest study location. The substantial distance effects observed for the remaining fields again appear to suggest that Finnish students are, in general, fairly uncertain regarding their choice of field after high school and, therefore, often ultimately select one of the alternatives available at the nearest university.

Finally, certain implications for educational policy may be derived based on the results obtained. In particular, one may argue that although transforming a small and remote university from a multidisciplinary institution into a more specialised one with fewer field-of-study options could be beneficial in terms of its competitiveness, regional disparities in access to particular fields may increase as a result of such a policy. For example, one might consider the consequences of ceasing to supply university degrees in certain fields at the universities located in northern Finland. The results of the current analysis suggest that this action would significantly reduce the likelihood of high school graduates from northern Finland moving into these fields.

REFERENCES

Antikainen, J. (2001) Functional regions of the future – The Finnish Urban Network Study 2001, *Journal of Nordregio*, 1, 20–23.

Avery, C. & Hoxby, C.M. (2004) Do and should financial aid packages affect students' college choices? In C.M. Hoxby, (ed) *College choices: The economics of where to go, when to go, and how to pay for it*, University of Chicago Press, pp. 239–301.

Beffy, M., Fougère, D. & Maurel, A. (2012) Choosing the field of study in postsecondary education: Do expected earnings matter? *Review of Economics and Statistics*, 94, 334–347.

Berger, M. (1988) Predicted future earnings and choice of college major, *Industrial and Labour Relations Review*, 41, 418–429.

Bertrand-Cloodt, D., Cörvers, F., Heijke, H. & Van Thor, J. (2010) The impact of distance deterrence on the choice of field of study in vocational education in the Netherlands, in: M.J. Mancebón-Torrubia, D.P. Ximénez-de-Embún, J.M. Gómez-Sancho & G.G. Esteban, (eds) *Investigaciones de Economía de la Educacíon*, Asociación de Economía de la Educación, pp. 605–624.

Boudarbat, B. (2008) Field of study choice by community college students in Canada, *Economics of Education Review*, 27, 79–93.

Denzler, S. & Wolter, S.C. (2011) Too far to go? Does distance determine study choices? IZA Discussion Paper No. 5712.

Frenette, M. (2006) Too far to go on? Distance to school and university participation, *Education Economics*, 14, 31–58.

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

Goyette, K.A. & Mullen, A.L. (2006) Who studies the arts and sciences? Social background and the choice and consequences of undergraduate field of study, *Journal of Higher Education*, 77, 497–538.

Haveman, R. & Wolfe, B. (1995) The determinants of children's attainments: A review of methods and findings, *Journal of Economic Literature*, 33, 1829–1878.

Heiss, F. (2002) Structural choice analysis with nested logit models, *Stata Journal*, 2, 227–252.

Jauhiainen, S. (2010) Studies on human capital flows and spatial labour markets, Jyväskylä Studies in Business and Economics 94, University of Jyväskylä, Finland.

Jepsen, C. & Montgomery, M. (2009) Miles to go before I learn: The effect of travel distance on the mature person's choice of a community college, *Journal of Urban Economics*, 65, 64–73.

Leppel, K. (1993) Logit estimation of a gravity model of the college enrolment decision, *Research in Higher Education*, 34, 387–398.

Leppel, K., Williams, M.L. & Waldauer, C. (2001) The impact of parental occupation and socioeconomic status on choice of college major, *Journal of Family and Economic Issues*, 22, 373–394.

Long, B.T. (2004) How have college decisions changed over time? An application of the conditional logistic model, *Journal of Econometrics*, 121, 271–296.

Ministry of Education and Culture. (2013) KOTA online database service. Available at https://kotaplus.csc.fi.

Montmarquette, C., Cannings, K. & Masheredjian, S. (2002) How do young people choose college majors? *Economics of Education Review*, 21, 543–556.

Neumann, R., Parry, S. & Becher, T. (2002) Teaching and learning in their disciplinary context: A conceptual analysis, *Studies in Higher Education*, 27, 405–417.

Ordovensky, J. (1995) Effects of institutional attributes on enrollment choice: Implications for postsecondary vocational education, *Economics of Education Review*, 14, 335−350.

Schwartz, A. (1973) Interpreting the effect of distance on migration, *Journal of Political Economy*, 81, 1153–1169.
Sjaastad, L. (1962) The costs and returns of human migration, *Journal of Political Economy*, 70, 80–93.
Statistics Finland (2013) Official Statistics of Finland (OSF). Available at http://www.stat.fi/meta/svt.
Suhonen, T. (2013) Are there returns from university location in a state-funded university system? *Regional Science and Urban Economics*, 43, 465–478.
Train, K. (2003) Discrete choice methods with simulation, New York: Cambridge University Press.
Wolniak, G.C., Seifert, T.A., Reed, E.J. & Pascarella, E.T. (2008) College majors and social mobility, *Research in Social Stratification and Mobility*, 26, 123–139.

APPENDIX 1 List of variables used in the analysis

- *Field of study.* The field of the first completed higher education degree (Bachelor's or Master's) after graduation from high school. This variable includes 10 university-level categories: education; arts; humanities; business; social sciences; law; natural sciences; technology; medicine/health sciences; other field (agriculture, forestry, military sciences and sports sciences). In addition, the variable includes 5 polytechnic-level categories: humanities; business; technology; health; other field.
- *Shortest distance to enrol.* The road distance from the individual's residential location one year before high school graduation to the nearest city in which the field is available. The distances are measured from the central municipality (usually the largest municipality) of each NUTS-4 region. The matrix of road distances was provided by the Finnish Road Administration.
- *Student places in the nearest study location.* The total number of admitted students in the field in the nearest city in which the field is available at the year of high school graduation. This variable is based on the information collected from the KOTA database (Ministry of Education and Culture, 2013).
- *Admission percentage in the nearest study location.* The number of admitted students per the number of applicants in the field in the nearest city in which the field is available at the year of high school graduation. This variable is based on the information collected from the KOTA database (Ministry of Education and Culture, 2013).
- *Majors in the nearest study location.* The number of application options (usually majors or special study programmes) within the field in the nearest city in which the field is available. This variable is based on the information collected from an Internet site of the National Board of Education (www.koulutusnetti.fi).
- *Region-of-origin urbanisation level.* The urbanisation level of the NUTS-4 region in which the individual resided one year before the year of high

school graduation. This variable includes 6 categories: sparsely populated region; rural region; industrial centre; regional centre; many-sided university region; metropolitan area. The urbanisation classification of the NUTS-4 regions is based on the Finnish Urban Network Study (Antikainen, 2001).

- *Female* takes a value of 1 if a woman and 0 otherwise.
- *Swedish speaker* takes a value of 1 if the individual's mother tongue is Swedish and 0 otherwise.
- *High school graduation year.* This variable includes 6 categories (years from 1991 to 1996).
- *First language grade.* The highest grade obtained in the first language test of the Finnish matriculation examination. The variable includes five categories: improbatur (I), approbatur (A) or lubenter approbatur (B); cum laude approbatur (C); magna cum laude approbatur (M); eximia cum laude approbatur (E) or laudatur (L); no grade available.
- *Math grade.* The highest grade obtained in the mathematics test of the Finnish matriculation examination. This variable includes 11 categories: B-level improbatur (I) or approbatur (A); B-level lubenter approbatur (B); B-level cum laude approbatur (C); B-level magna cum laude approbatur (M); B-level eximia cum laude approbatur (E) or laudatur (L); A-level improbatur (I) or approbatur (A); A-level lubenter approbatur (B); A-level cum laude approbatur (C); A-level magna cum laude approbatur (M); A-level eximia cum laude approbatur (E) or laudatur (L); no grade available.
- *Mother's/Father's education level.* The level of mother's/father's highest educational qualification. These variables include 7 categories: primary school or no classification available; high school diploma; vocational diploma (at the secondary level); lowest tertiary-level qualification; bachelor's degree; master's degree; doctoral degree.
- *Mother's/Father's occupational status.* The latest observed type of the individual's mother's/father's vocation (observations in 1970, 1980, or 1990). These variables include 5 categories: farmer; entrepreneur; high-ranking official; low-ranking official; worker or no classification available.
- *Parent's field j* takes a value of 1 if at least one of the individual's parents has a university degree (Bachelor's, Master's or PhD) in field j and 0 otherwise. Each field-of-study-specific equation includes 9 parent's field indicators: parent's field education; parent's field arts or humanities; parent's field business; parent's field social sciences; parent's field law; parent's field natural sciences; parent's field technology; parent's field medicine/health sciences; parent's field other.

APPENDIX 2

TABLE A1 Field-of-study categories and their nests in the nested logit analysis

		Nest			
Field-of-study alternative	Examples of majors/study programmes in the field	Soft science	Hard science	Lucrative	Non-lucrative
University-level alternatives					
Education	education, special pedagogy, class teacher, kindergarten teacher	x			x
Arts	theatre, fine arts, industrial arts, music	x			x
Humanities	literature, languages, history, theology, archaeology, ethnology, philology, communication, language teacher	x			x
Business	business administration, leadership, marketing, accounting, entrepreneurship, economics, information system science	x		x	
Social sciences	philosophy, political science, sociology, social policy, economics, administrative science, psychology	x			x
Law	law, notary	x		x	
Natural sciences	biology, physics, chemistry, mathematics, statistics, computer science, geography, science teacher		x	x	
Technology	architecture, engineering & technology, industrial management		x	x	
Medicine	medicine, dentistry, pharmacy, veterinary medicine, physiotherapy, health education, health economics, gerontology		x	x	
Other field	sport sciences, agriculture, forestry, military officer		x	x	
Polytechnic-level alternatives					
Humanities	cultural producer, musician, designer, artist, community pedagog, dance teacher, conservator	x			x
Business	business administration, sales work, international trade, security services, management assistant	x			x
Technology	data processing, engineering & technology, industrial management		x	x	
Health	nurse, physiotherapist, midwife, paramedic, socionom, dental hygienist		x		x
Other field	tourism, hotel and restaurant business, horticulturalist, agrologist		x		x

CHAPTER 3: INTERGENERATIONAL TRANSMISSION OF FIELD OF STUDY: EVIDENCE FROM THE FIELD-OF-STUDY CHOICE OF FINNISH UNIVERSITY STUDENTS[28]

This paper contributes to the research on intergenerational transmission of education by providing evidence from Finland regarding the association between parents' and children's field-of-study choices in the context of university education. The results indicate a considerable intergenerational transmission of field: having a parent with a university degree from a particular field is, on average, found to be associated with twice higher odds of graduating that field. However, the parental effect is found to be heterogeneous in many ways. In particular, the law field appears to be transmitted from parents to children more frequently than any other field.

Keywords: field of study; intergenerational transmission; higher education

28 Financial support from the Academy of Finland is gratefully acknowledged (grant number 127049). I also wish to thank Mika Haapanen, Kaisa Kotakorpi, Timo Kuosmanen, Jukka Lahtonen, Hannu Tervo, Olle Westerlund and participants at the ERSA 2012 Conference (Bratislava, Slovakia) for their many helpful comments.

1 Introduction

Empirical studies frequently conclude that children with highly educated parents achieve a higher average educational level compared to children with low-educated parents, and that partly for this reason, socioeconomic status is transmitted from a generation to another (for literature reviews, see Haveman and Wolfe 1995; Björklund and Salvanes 2011). In a broad sense, there are two competing explanations for an intergenerational correlation of education: that arising from a correlation between parents' and children's inherited cognitive abilities and other traits ('nature') and that arising from environmental factors such as parental investment in a child's human capital ('nurture'). In recent empirical studies, there has been a heavy focus on distinguishing between these different mechanisms; some researchers have found evidence of a causal effect of parents on a child's educational attainment – that is, an effect arising from post-birth environmental factors rather than genetic endowments (Sacerdote 2007; Björklund et al. 2006; De Haan 2011) – whereas some others have not been supportive of such an effect (Behrman and Rosenzweig 2002; Plug 2004).

Thus far, the research on intergenerational transmission of education has almost entirely focused on a child's educational level, whereas very little evidence on the transmission of educational field appears to exist.[29] Knowing that future occupations and earnings vary considerably across students from different fields,[30] the transmission of fields from parents to children – the 'doctor's son becomes a doctor' effect – could be another important factor generating social stratification in modern societies (see Van de Werfhorst et al. 2001; Wolniak et al. 2008). In today's context, analyses of the intergenerational transmission of field in postsecondary education – that is, among university and college students – using nationally representative data sets may be particularly interesting because, with the recent rapid rise of the general educational level, students' postsecondary field-of-study choices have apparently become increasingly important determinants of overall earnings inequality (see Wolniak et al. 2008). However, no previous studies focusing on this particular topic appear to exist, which may be partly due to data limitations:

---

29 A search through the literature revealed only two paper – those by Dryler (1998) and Van de Werfhorst et al. (2001) – that provide theoretical discussion and empirical evidence on intergenerational transmission of field. These studies are discussed in Section 2.1.

30 For instance, completing a university degree in business, law or engineering generally results in higher average earnings than a degree in education or the humanities. Naturally, part of these earnings differences are explained by compositional differences – such as differences in gender and ability – across the fields. However, according to the existing empirical evidence, a causal relationship between earnings and field-of-study choice also exists. For example, Arcidiacono (2004) finds large earnings premiums for certain college fields in the US even after accounting for the field-of-study differences in individual ability.

for such an analysis, accurate information on individuals' mothers' and fathers' postsecondary educational degrees must be available.[31]

This paper contributes to the topic of intergenerational transmission of field by providing evidence of this phenomenon in the context of Finnish university education. The analysis is conducted using administrative data on a representative sample of Finnish students who graduated from high school between 1991 and 1996 and subsequently obtained a university degree. The primary goal of the paper is to answer the following question: to what extent (if at all) is a student more likely to graduate from a particular field if he or she has a parent with a university degree from that field? The current data set allows the students' and their parents' university degrees to be divided into nine broad field-of-study categories with a sufficient amount of observations in each category; thus, the data enable a relatively accurate analysis of the similarity between parents' and children's choices. By examining the sensitivity of the parental effect estimates to different functional form assumptions and sets of controlling covariates, the mechanisms and heterogeneity of the intergenerational relationships can be assessed to some extent, whereas a comprehensive attempt to identify a 'causal' parental effect on field-of-study choice is not possible due to limitations of the current data set.[32] Nevertheless, new evidence and insights regarding the persistence of educational outcomes across generations will be provided.

The paper is organised as follows. Section 2 discusses theoretical explanations and previous empirical evidence regarding intergenerational transmission of field and briefly describes the institutional setting of the Finnish university system. In Section 3, the data set of the study is discussed and descriptive evidence of the determinants of field-of-study choice is shown. Section 4 discusses the methodology used in the empirical analysis: the baseline results are obtained by the estimation of a conditional logit model assuming a fixed parental effect across both individuals and choice alternatives, whereas robustness to different sources of heterogeneity is studied using mixed logit and multinomial logit approaches. The results of the empirical analysis are presented in Section 5, and Section 6 concludes the paper.

---

31 Over the years, the mechanisms behind postsecondary students' field/subject choices have attracted a certain amount of interest among social scientists. For example, many studies have investigated the effect of parents' socioeconomic status on these choices (e.g., Leppel et al. 2001; Bratti 2006; Wolniak et al. 2008). Furthermore, several papers have examined whether the students' field choices respond to changes in the expected earnings that these choices yield (Berger, 1988; Montmarquette et al. 2002; Bourdarbat 2008; Beffy et al. 2012).

32 In previous empirical studies on intergenerational transmission of education, different strategies have been applied for separating effects arising from 'nature' and 'nurture'. These strategies include 1) using data on twin parents, 2) using data on adoptees and 3) using instrumental variables such as structural changes in educational systems (for a review of the methods, see Björklund and Salvanes 2011). With the current data set, the first two strategies are out of the question, and the possibilities to employ the last strategy are very limited: information on parents' characteristics (education, place of residence, etc.) is only available from 1970 onwards due to which potential instrumental variables affecting parents' educational choices in their early lives cannot, in most cases, be formed in a reliable manner.

2 Background

2.1 Theoretical considerations and previous research

A considerable amount of theoretical and empirical work has been conducted to explain intergenerational transmission processes (see, e.g., Haveman and Wolfe 1995). However, the discussion in this literature has rarely focused explicitly on the transmission of educational fields. In the classical 'nature-nurture' framework, a student is thought to be inclined to choose a similar field as his/her parent because he/she shares, for example, similar inborn cognitive abilities and preferences regarding educational and occupational possibilities with the parent and/or because of various post-birth processes that are involved in having a parent with a particular type of educational background. In previous studies on intergenerational transmission of field, certain hypotheses have been presented regarding what these post-birth processes might be. For example, Dryler (1998) links her analysis of the field-of-study choice of Swedish upper secondary students to role model theories, stating that a child learns behavioural patterns from adult role models and, in particular, from his/her parents. Therefore, the child may be inclined to choose the same field as one of the parents ('same-sector effect') and, in particular, the field chosen by the parent of the same sex ('same-sex effect'). Alternatively, the child could be inclined to imitate the more influential parent ('dominance effect'). In their analysis from the Netherlands, Van de Werfhorst et al. (2001) provide several other explanations for the intergenerational transmission. First, cultural capital acquired through family-of-origin may provide a child with extra skills and motivation for studies in fields chosen by his/her parents. Second, a child may take a parent's social status as a reference point for his/her own career, and therefore, to achieve at least the same status, the child selects the field that established the parent's status. Finally, a parent may encourage a child to choose his/her own field, for example, by providing study counsel and information regarding job opportunities in the field.

Although the discussion of both Dryler (1998) and Van de Werfhorst et al. (2001) mainly concerns the choice of field at an earlier stage of life – at the secondary level of education – the mechanisms behind the intergenerational transmission of postsecondary field are likely to be at least partly similar. After all, many students are likely to contemplate which university or college to attend and which subjects to study after high school already when they are still living with their family-of-origin. Moreover, students' subject choices at high school are likely to be partly reflected in their subsequent choices. Aside from the above-discussed mechanisms, families' locational choices represent another potential source of intergenerational transmission, particularly in the context of postsecondary education: students having parents from a particular field could, on average, live closer to universities or colleges offering education in that field than other students and, thus, under a limited spatial mobility of students, could be more likely to enrol as a student in that field.[33]

The papers by Dryler (1998) and Van de Werfhorst et al. (2001) also appear to be the only ones providing empirical evidence related directly to the current topic. Both studies find considerable similarity between parents' and children's field-of-study choices. Furthermore, the findings of Dryler yield support for the same-sector effect in the choices of both boys and girls, whereas the same-sex hypothesis is supported only in the case of boys. Dryler also finds weak support for the dominance effect. Van de Werfhorst et al. again find that the strength of the transmission of a field from father to child varies, to some extent, across the fields, being strong in the general, teacher/educational and agricultural fields, while insignificant in the cultural, engineering and medical/caring fields. Methodologically, the current analysis highly resembles those by Dryler and Van de Werfhorst et al., whereas the educational level at which field-of-study choice is studied is different: Dryler studies upper secondary students, whereas Van de Werfhorst et al. use field-of-study categories that include degrees from all educational levels.

2.2 Institutional setting

This paper focuses on students' field-of-study choices at the top level of Finland's educational system, that is, within the Finnish university system, which is a publicly financed and administered system responsible for the provision of academic research and education at the bachelor's, master's and PhD levels.[34] In the 1990s, when high school graduates in the study sample entered postsecondary studies, the Finnish university system comprised 16 universities (10 multidisciplinary universities, 3 business schools and 3 universities of technology), 4 art academies and the Finnish National Defence University; all of these institutions continue to exist today, but some of the institutions have merged. The system has been distributed across 10 city regions countrywide: Helsinki, Turku, Tampere, Lappeenranta, Kuopio, Joensuu, Jyväskylä, Vaasa, Oulu and Rovaniemi.[35] Thus, the system offers a university education relatively close to each resident's place of residence, which promotes regional equality in access to university education. Admission to universities is generally based on the applicants' high school grades, particularly grades from the matriculation examination, and a field-of-study-

33 Previously, the link between residential choices and intergenerational transmission of education has been noted, e.g., by Checchi (2006). In a separate paper – parallel to this one – the effect of geographical distance on the field-of-study choice of Finnish students is examined thoroughly.

34 In addition to the university sector, the Finnish higher education system has – since 1996 – included another sector, the polytechnics (also known as the universities of applied sciences), which consists of 28 schools located across the country. These schools award vocational degrees that are usually paralleled by the bachelor's degree.

35 In addition, smaller university consortiums are currently located in 6 other cities. These consortiums serve as branch campuses of the university institutions and are primarily dedicated to adult education.

specific entrance examination. Studying is free of charge and publicly subsidised by the provision of study grants, housing allowances and state-guaranteed loans. Thus, virtually no financial barriers are involved in university participation or in the choice of institution.

In Finland's educational system, a student's road to university usually consists of nine years of comprehensive school and three years of upper secondary school.[36] The transition to university marks a large increase in specialisation: whereas education at the pre-university levels is highly multidisciplinary – including compulsory courses in nearly all disciplines – the university studies are, already from the beginning, focused on a few narrow academic subjects. When applying for university studies, students apply directly to a specific major subject or discipline within a faculty of a university institution. Thus, in the Finnish system, the choice of concentration must be made at a relatively early stage in comparison with some other higher education systems (e.g., the US college system). Given that this 'sudden' choice of concentration has potentially large consequences for a student's future life, knowledge on the underlying mechanisms of the choice is relevant from both policy and scientific perspective.

3 Data and descriptive evidence

The study data are based on a 7% random sample drawn from the population of Finland in 2001. The data originate from the registers of Statistics Finland and include yearly panel data, e.g., on individuals' labour market outcomes, residential region, job characteristics, educational qualifications and family background from the period between 1970 and 2006. From these data, the study utilises a subsample consisting of individuals who graduated from high school between 1991 and 1996 at a young age (less than 23 years old) and, subsequently, completed a university degree no later than 2006; using these criteria, the sample size is limited to 3,869 observations.[37] The analysis of the field-of-study choice is based on the first university degree that a student completed after high school. Based on a three-digit code depicting the level and the field of the obtained degree, the degrees are divided into nine field-of-study categories: 'education', 'arts/humanities', 'business', 'social sciences', 'law', 'natural sciences', 'technology', 'medicine/health sciences' and 'other field' (including agriculture, forestry, military sciences and sports sciences).[38]

---

36 In this paper, the term 'high school' is used in many cases for conciseness.

37 Four individuals were also excluded from the sample because of missing information on pre-university region, which is used as a control variable in the analysis.

38 Aside from the field of study, some other information regarding the completed degrees can also be observed from the data. Based on the information, only 15% of the degrees were bachelor's degrees, and the remaining 85% were master's degrees. This result is not surprising, as completing a bachelor's degree before a master's degree was optional – and therefore rare – in the Finnish university system until 2005. The data also show that the average student in the sample graduated from university

For examining the intergenerational transmission of the fields, the data include the students' mothers' and fathers' highest educational qualifications at the same level of accuracy as the students' own qualifications. In the data, there are in total 1,532 individuals who have at least one parent with a university degree in one of the nine field categories. Thus, a large number of observations (comprising 40% of all observations) actually convey information on whether a field was transmitted from parent to child or not, which is crucial in the identification of the 'parental effect'.[39] The first picture of the similarity between parents' and children's field-of-study choices is provided in Table 1. The figures in this table represent proportions of graduates in a field (columns) having a parent with a particular type of degree (rows). From these figures, we see that, in each field, the group of students having a parent from that particular field is clearly overrepresented; the proportion of this group varies from 5.7% ('other field') to 14.4% (medicine/health sciences). The intergenerational immobility appears to be the strongest in the case of law: 11.9% of the law students have at least one parent with a law degree, whereas this percentage is only 2.1% among the other students. Thus, in line with predictions, having a parent from a particular field appears to be associated with a higher likelihood of graduating from that field. However, certain large 'cross-field transitions' are also observed; for instance, students having a parent from arts/humanities are also very likely to graduate from social sciences, law and 'other field'.

In addition to parents' field-of-study choices, a number of other variables are employed as explanatory variables during the analysis. Table 2 shows the field-of-study differences in several individual and family-of-origin characteristics. Of the individual-level observables, gender is clearly one of the strongest determinants of field-of-study choice. Based on the reported female shares, education (83%), arts/humanities (78%), social sciences (75%) and medicine/health sciences (71%) are clearly female-dominated fields, whereas technology (22%) is a male-dominated field; in the remaining fields, the gender distribution is more even. Certain noteworthy differences also exist in mother tongue: Swedish-speaking students are clearly overrepresented in business and social sciences with shares of 12% and 10% of all students in these fields, respectively, whereas the share of this language group is relatively small in

---

7.4 years after graduating from high school; thus, the average time to graduation is relatively long among Finnish students. Furthermore, the data provide evidence that the Finnish students are rather immobile after graduating from high school: 46% of the students graduated from university in their nearest university city, and 68% graduated from one of the three nearest cities.

39 However, the sample size is still rather small with regard to analysing heterogeneity in the parental effect. For example, in most cases, we do not attempt to estimate the mother's and father's effects separately, as the number of identifying observations would be insufficient for that purpose. In addition, we only use a pooled sample of men and women because with a division of the sample by gender, the number of both identifying observations and observations used for estimating control variable coefficients would reduce dramatically.

TABLE 1 Intergenerational similarity of fields: percentages of students in each field having a parent with a university degree in a particular field

	Child's field									
Parent's field	Education	Arts/ humanities	Business	Social sciences	Law	Natural sciences	Technology	Medicine/ health sciences	Other field	Any field
Education	**9.7 %**	6.4 %	4.5 %	6.4 %	4.4 %	7.3 %	7.9 %	6.2 %	9.3 %	7.1 %
Arts/humanities	4.0 %	**12.1 %**	6.9 %	9.8 %	10.0 %	6.6 %	5.7 %	8.2 %	12.1 %	7.8 %
Business	2.0 %	4.8 %	**12.6 %**	3.7 %	8.8 %	2.9 %	6.2 %	3.4 %	2.9 %	5.2 %
Social sciences	3.8 %	6.9 %	6.9 %	**8.8 %**	6.9 %	3.7 %	4.9 %	6.9 %	5.7 %	5.8 %
Law	1.1 %	2.0 %	2.8 %	2.7 %	**11.9 %**	1.3 %	2.6 %	2.1 %	2.9 %	2.5 %
Natural sciences	4.4 %	3.8 %	6.2 %	4.8 %	4.4 %	**9.2 %**	8.7 %	5.2 %	7.1 %	6.1 %
Technology	3.1 %	5.3 %	6.9 %	4.0 %	6.3 %	5.1 %	**12.5 %**	7.2 %	3.6 %	6.6 %
Medicine/health sciences	2.9 %	4.8 %	5.2 %	4.3 %	10.6 %	7.3 %	7.1 %	**14.4 %**	5.0 %	6.3 %
Other field	3.3 %	3.6 %	3.3 %	2.1 %	2.5 %	3.3 %	3.1 %	4.5 %	**5.7 %**	3.4 %
N	546	685	422	376	160	455	794	291	140	3869

Note: The percentages of students having a similar field with one of the parents are emboldened.

TABLE 2 Differences in individual and family-of-origin characteristics across fields: means of selected variables

	Education	Arts/ humanities	Business	Social sciences	Law	Natural sciences	Technology	Medicine/ health sciences	Other field
Female	0.83	0.78	0.52	0.75	0.55	0.53	0.22	0.71	0.44
Swedish speaker	0.08	0.06	0.12	0.10	0.07	0.07	0.05	0.08	0.03
First language grade [a]	4.89	5.38	5.00	5.25	5.27	5.04	4.97	5.20	4.97
A-level math grade [a]	3.68	4.19	4.50	4.34	4.39	4.84	5.21	4.88	4.27
B-level math grade [a]	3.99	4.45	4.80	4.36	4.79	4.65	4.85	4.64	4.44
Mother's education level [b]	3.20	3.56	3.54	3.41	3.89	3.58	3.74	3.71	3.84
Mother's socioeconomic status									
Farmer/entrepreneur	0.11	0.11	0.14	0.11	0.11	0.13	0.13	0.16	0.16
High-/low-ranking official	0.71	0.74	0.74	0.74	0.82	0.74	0.74	0.72	0.73
Worker/other	0.18	0.15	0.13	0.15	0.08	0.13	0.13	0.12	0.11
Father's education level [b]	3.27	3.68	3.99	3.75	4.14	3.82	4.14	3.92	3.86
Father's socioeconomic status									
Farmer/entrepreneur	0.20	0.17	0.20	0.18	0.17	0.16	0.16	0.21	0.24
High-/low-ranking official	0.51	0.59	0.63	0.57	0.66	0.62	0.64	0.61	0.55
Worker/other	0.29	0.24	0.18	0.25	0.17	0.22	0.20	0.18	0.21
High school region type									
Metropolitan region	0.14	0.23	0.32	0.25	0.42	0.17	0.25	0.26	0.26
University region	0.24	0.25	0.23	0.27	0.31	0.28	0.25	0.26	0.16
Small region	0.61	0.52	0.44	0.48	0.27	0.55	0.50	0.47	0.58
N	546	685	422	376	160	455	794	291	140

Notes: [a] Grades from the Finnish matriculation examination. For calculating the mean grades, numerical values are assigned as follows: I = 1; A = 2; B = 3; C = 4; M = 5; E/L = 6. [b] Mother’s and father’s education levels take values from 1 (primary school) to 7 (doctoral degree).

'other field' (3%).[40] In addition, Table 2 shows that the students' grades from the matriculation examination are somewhat correlated with their field-of-study choices; for example, the average first language grade is particularly high among arts/humanities students, whereas students in technology have a relatively high A-level math grade. As for the family background variables, one notable finding from Table 2 is that the mean level of parental education is clearly lower in education than in any other field, suggesting that students with low-educated parents are relatively likely to end up studying in this field. Considerable field-of-study differences also exist in the type of region that a student originated from. In particular, students from the Helsinki metropolitan area are clearly overrepresented in business and law but underrepresented in education and natural sciences.

4 Methodology

Although the descriptive evidence of Section 3 already provides an indication of a 'parental effect' in the field-of-study choice, the magnitude of this effect cannot be directly inferred from this evidence. Thus, to estimate the impact of having a parent with a particular degree on the likelihood of choosing that degree, a variety of commonly used random utility models are applied. For the analysis, the data are organised in the following way: the 9 field-of-study alternatives are assigned for each individual in the sample, and an indicator for whether a field was chosen or not is generated; this indicator serves as the dependent variable. Thus, with 9 choice alternatives and 3,869 individuals, the resulting data includes in total 34,821 choice observations. In the random utility framework, individual i's field-of-study choice is assumed to be determined by the maximisation of the latent utility u_{ij}^{*} over the field-of-study alternatives $j = 1,2,\dots,9$. The baseline results are obtained by assuming that the latent utilities are of the following conditional logit form:

$$u_{ij}^{*} = \alpha_j + \beta d_{ij} + \gamma_j' x_i + \vartheta_{ij}, \quad (1)$$

where α_j is a field-specific constant term; d_{ij} is a *parent's field* indicator, i.e., an indicator for whether one of i's parents graduated from field j; x_i is a vector of individual-level control variables; and ϑ_{ij} is a standard logistic error term. In model (1), parameter β depicts the change in the log-odds of choosing field j in response to having a parent with a university degree in field j. More generally, β can be interpreted as a measure of the strength of the

40 Evidently, the differences by language in part arise from field-of-study differences in the availability of Swedish-speaking university education. For instance, business education in the Swedish language is relatively abundantly provided in two Swedish-speaking universities, whereas a greater scarcity of Swedish-speaking education exists in other fields.

intergenerational transmission of field of study;[41] here, the term *same-field parental effect* is also used to make a distinction to a *cross-field parental effect*, i.e., the effect of having a parent with field j on choosing another field. When reporting the conditional logit results – obtained by maximum likelihood estimation – the interpretation is based on the exponentiated coefficient estimate $exp(\hat{\beta})$, which depicts the odds ratio $P_{ij}^{1}/(1-P_{ij}^{1})/P_{ij}^{0}/(1-P_{ij}^{0})$ associated with having a parent with a degree in field j (where P_{ij}^{1} and P_{ij}^{0} are the choice probabilities for the cases $d_{ij}=1$ and $d_{ij}=0$, respectively).

The baseline model (1) has certain practical advantages over more complicated models. First, the model enables an efficient estimation of the parental effect: as this effect is only described by a single parameter, a maximum amount of variation in the data is used for its identification. Second, the model enables a quick and simple estimation, as the probability of selecting field j in this model has a closed-form solution

$$P_{ij}^{CL} = \frac{e^{u_{ij}^{*}}}{\sum_{l=1}^{9} e^{u_{il}^{*}}}, \tag{2}$$

which can be used to construct the log-likelihood function.[42] However, these advantages are achieved at the expense of making certain restrictive assumptions regarding the role of individual- and alternative-level heterogeneity in the field-of-study choice: having a parent with field j is assumed to have a similar effect on the odds of choosing field j for individuals with similar characteristics x_i and irrespective of the field-of-study alternative in question; in addition, the cross-field parental effects are restricted to zero. Clearly, these assumptions are strong, due to which a number of more flexible models are estimated during the analysis as robustness checks. The first supplemental analysis is conducted using a *mixed logit* approach, in which the same-field parental effect is allowed to vary randomly across individuals. Thus, in this approach, the latent utility from choosing field j is rewritten as

$$u_{ij}^{*} = \alpha_j + \beta_i d_{ij} + \gamma_j' x_i + \vartheta_{ij}, \tag{3}$$

where the idiosyncratic parameter β_i is assumed to be normally distributed with mean parameter $E(\beta_i) \equiv \mu$ and standard deviation $Std(\beta_i) \equiv \sigma$.[43] A non-zero estimate for σ indicates that individuals with similar characteristics x_i

---

41 In the literature, β is sometimes referred to as the 'immobility parameter' (Van de Werfhorst et al. 2001).

42 In practise, the conditional logit estimations are conducted using STATA's package 'asclogit'.

43 Alternatively, one could induce more flexibility by allowing the error terms to be correlated across the choice alternatives, e.g., by using the multinomial probit approach (see Train, 2003, ch. 5). However, in this type of model, the number of correlation terms easily becomes infeasible for estimation when the number of choice alternatives is large. The strategy of using random coefficients to allow for more flexibility is therefore preferred in this study.

respond differently to changes in d_{ij} and that the mixed logit model (3) is preferable over model (1). As the idiosyncratic effect is not observed by the econometrician, the choice probabilities for the mixed logit model must be solved by integrating the conditional logit probability – given by equation (2) – over a parameter distribution as follows:

$$P_{ij}^{ML} = \int P_{ij}^{CL} \phi(\beta_i | \mu, \sigma^2) d\beta_i, \qquad (4)$$

where P_{ij}^{CL} is the conditional logit probability, and $\phi(\beta_i | \mu, \sigma^2)$ is the normal density of β_i with mean μ and variance σ^2. The method of maximum simulated likelihood is used for the estimation of the model parameters.[44]

In the second supplemental analysis, the following *multinomial logit* specification – including all of the same-field and cross-field terms for the parent's field indicators – is considered:

$$u_{ij}^* = \alpha_j + \sum_{k=1}^{9} \beta_k d_{ik} + \gamma_j' x_i + \vartheta_{ij}, \qquad (5)$$

where d_{ik} is an indicator for whether one of i's parents graduated from field k; thus, β_k depicts a same-field parental effect when $k = j$ and a cross-field parental effect when $k \neq j$. The advantage of model (5) over models (1) and (3) is that it accounts for the fact that the field-of-study alternatives are not homogenous goods and that a particular field may be more substitutable with some fields than with some others. When estimating the multinomial logit model, the same-field and cross-field terms for the base category (education) must be normalised to zero, which makes the interpretation of the exponentiated coefficients $exp(\widehat{\beta_k})$ slightly more complicated; these coefficients depict relative changes in the odds for choosing a field versus the base category. Therefore, marginal effects $\partial P_{ij} / \partial d_{ik}$ – evaluated at the means of the explanatory variables – are also reported in Section 5 in the case of the multinomial logit specification. For the sake of comparison, a restricted version of model (5), in which all of the cross-field coefficients are normalised to zero, is also estimated; this model is similar to the one used by Van de Werfhorst et al. (2001) in their analysis of intergenerational transmission.

After having discussed the assumptions regarding the functional form of u_{ij}^*, a relevant question follows: what is the role of the control variables included in vector x_i in the estimation? The primary purpose of these variables is to capture correlations in the error terms ϑ_{ij} across the choice alternatives and thus to help in making the independence of irrelevant alternatives assumption (IIA) involved in logistic regressions more warranted (see Train, 2003).[45]

---

44 In practise, the mixed logit estimation is conducted using STATA's user-written command 'mixlogit' (see Hole 2007). In the estimation, 200 Halton draws are used to simulate the choice probabilities.

45 In effect, the mixed logit model – unlike the conditional and multinomial logit models – allows for a certain degree of correlation between the utilities arising from unobservables: if a particular unobserved individual-level characteristic affects one's tendency to choose a 'parental field', a correlation arising from this characteristic is

Furthermore, the control variables may be used – with certain limitations – for assessing the mechanisms through which the intergenerational transmission of field of study occurs. In the current control variable strategy, the models are augmented with the following variables: a female indicator, a Swedish speaker indicator, high school graduation year, matriculation examination grades in first language and math, mother's and father's education levels and vocational statuses and the type of pre-university residential region.[46] Of these variables, the first three can be thought as rather neutral controls whose purpose is only to level off differences in the basic demographic characteristics and the timing of the field-of-study choice.[47] The first language and math grades again have a crucial role in justifying the IIA assumption: the grades are likely to effectively capture correlations across ϑ_{ij} because students with similar first language and math grades are likely to be relatively similar with respect to linguistic and mathematical ability and preferences across different types of subjects (e.g., mathematical and non-mathematical ones). A justified concern involved in the use of the grades as control variables is that the grades – like any other educational outcomes – are unlikely to be fully determined by individuals' innate characteristics and thus could be related to the post-birth mechanisms behind the intergenerational transmission of field; that is, the grades are potentially endogenous to the variables of interest d_{ik} and, therefore, potentially harmful in terms of causal interpretation (see, e.g., De Coulon et al. 2009). However, as the focus of this paper is not in an exact identification of a causal parental effect, and as parents' ability to affect their children's school performance through the choice of educational field is likely to be far from perfect,[48] controlling for the grades is considered as a preferred strategy. The remaining variables – mother's and father's other characteristics and the type of pre-university region – are again likely to be highly associated with the consequences of parents' field-of-study choices because of which estimates conditioning on these 'socioeconomic' variables should be interpreted with caution. However, examining the impact of the socioeconomic variables may

---

accounted for. This feature also implies that the choice probabilities of the mixed logit model do not have the IIA property (see Train 2003).

46 The detailed descriptions of all variables employed in the analysis are included in Appendix 1.

47 Age at the year of high school graduation is not included in the demographic controls, as this variable varies very little in the sample: 98% of the students were either 19 or 20 years old that year.

48 One scenario creating a causal link between parents' field-of-study choices and high school grades could be that parents graduated from prestigious fields, such as business, law or medicine, are more able than other parents to send their children to prestigious schools prior to university. However, in the context of Finland's educational system, the existence of this link may be less obvious than in some other countries: education at all levels of the educational system is almost entirely provided in public schools with free-of-charge education and relatively homogeneous resources; therefore, parents' means to affect their children's school performance through school choice are very limited. Furthermore, the existing evidence on the determination of Finnish students' matriculation examination grades suggest that the grades are not significantly linked to school resources but rather to family background and previous school performance (Häkkinen et al. 2003).

still be interesting: if the parental effect estimates are highly sensitive to the inclusion of these variables (with other individual-level characteristics controlled for), this sensitivity could be interpreted as a sign of an intergenerational transmission occurring through family socioeconomic factors.

5 Results

In this section, results from the estimated random utility models are reported. The odds ratios from different variants of the baseline model (1) are shown in Table 3. The simplest model specification – including only 8 field-specific constants and the parent's field indicator (Model I, Column 1) – suggests that having a parent from a particular field is associated with a 2.19 times higher odds of choosing that field. A comparison of the odds ratio estimate across the columns of Table 3 reveals that this estimate is rather robust to the inclusion of individual-level control variables. A notable change in the estimate only occurs when the students' matriculation examination grades are included in the model; in these specifications (Columns 3 and 5), the estimated odds ratio is approximately 2, indicating a slightly smaller degree of intergenerational transmission.[49] In the remaining versions of the baseline model (Models II, III and IV in Table 3), differences in the parental effect across men and women and differences in mother's and father's effects are studied. These models provide two important findings. First, the odds ratios for the parent's field, mother's field and father's field indicators are systematically larger for men than for women, suggesting that men are more likely to imitate their parents' field-of-study choices; however, a statistically significant gender difference is only measured when the parent's field indicator is used as the explanatory variable (Model II), whereas the gender differences in the mother's and father' effects are not significant (Model IV). Second, although the results indicate a significant transmission of field of study from both parents – the odds ratios are significantly larger than one for both parental indicators – the father's effect appears to dominate the mother's effect in the case of both men and women. The difference in the father's and mother's effects is sizeable: in each specification and for each gender, the odds ratio for the father's field indicator is at least 48% larger than that for the mother's field indicator; these differences are also statistically significant.

The second set of results is obtained by estimating the mixed logit model (3), which allows for individual heterogeneity in the parental effect arising from unobservables. Table 4 shows that the estimate for the standard deviation of the parental effect is large and significant, suggesting that the effect varies

49 When reporting the results, the focus is solely on the estimates obtained for the variable of interest, whereas the discussion of the control variable estimates is omitted from the paper. However, to provide a better sense of the structure of the estimated models, the full set of coefficient estimates from one of the conditional logit specifications (Table 3, Model I, all controls included) is shown in Appendix 2.

TABLE 3 Odds ratio for graduating from a field associated with having a parent with a degree in that field: results from conditional logit models (sample size: 3,869 individuals, 9 choice alternatives)

Model/variable	Odds ratio				
Model I					
Parent's field	2.19 ***	2.23 ***	2.00 ***	2.19 ***	1.96 ***
	(0.13)	(0.14)	(0.13)	(0.14)	(0.14)
Model II					
Parent's field * Female		2.03 ***	1.75 ***	1.94 ***	1.68 ***
		(0.17)	(0.15)	(0.17)	(0.15)
Parent's field * Male		2.49 ***	2.35 ***	2.56 ***	2.37 ***
		(0.22)	(0.23)	(0.24)	(0.24)
Gender diff. significant		No	Yes	Yes	Yes
Model III					
Mother's field	1.39 ***	1.48 ***	1.37 ***	1.49 ***	1.33 ***
	(0.12)	(0.13)	(0.13)	(0.14)	(0.13)
Father's field	2.37 ***	2.38 ***	2.14 ***	2.32 ***	2.14 ***
	(0.16)	(0.17)	(0.16)	(0.18)	(0.17)
Model IV					
Mother's field * Female		1.43 ***	1.26 **	1.41 ***	1.21
		(0.16)	(0.15)	(0.17)	(0.15)
Mother's field * Male		1.55 ***	1.55 ***	1.61 ***	1.52 ***
		(0.22)	(0.23)	(0.24)	(0.24)
Gender diff. significant		No	No	No	No
Father's field * Female		2.26 ***	1.98 ***	2.11 ***	1.93 ***
		(0.23)	(0.21)	(0.22)	(0.21)
Father's field * Male		2.49 ***	2.30 ***	2.54 ***	2.36 ***
		(0.26)	(0.25)	(0.27)	(0.27)
Gender diff. significant		No	No	No	No
Control variables:					
Field-specific constants	Yes	Yes	Yes	Yes	Yes
Demographic variables	No	Yes	Yes	Yes	Yes
Matriculation grades	No	No	Yes	No	Yes
Socioeconomic variables	No	No	No	Yes	Yes

Notes: Standard errors are in parentheses. * ($p<0.1$), ** ($p<0.05$) and *** ($p<0.01$) indicate whether an odds ratio is significantly different from one. The set of demographic variables includes a woman indicator, a Swedish speaker indicator and high school graduation year. The set of socioeconomic variables includes mother's and father's vocational statuses, mother's and father's education levels and the type of pre-university region.

considerably across individuals with similar observed characteristics. The estimate for the average parental effect is still significantly positive and only slightly smaller than the previous conditional logit estimate: the implied average odds ratio for choosing a field is 1.81 and 1.84 with and without conditioning on family socioeconomic variables, respectively. However, because of the implied variation in the effect, the mixed logit model also predicts that in 21–22% of the cases, having a parent from a particular field decreases the likelihood of choosing that field instead of increasing it.

TABLE 4 Random variation in the parental effect on field-of-study choice: results from a mixed logit model (sample size: 3,869 individuals, 9 choice alternatives)

Parameter for the parent's field indicator	Estimate			
Mean coefficient	0.614	***	0.596	***
	(0.093)		(0.100)	
Standard deviation	0.766	**	0.761	**
	(0.303)		(0.330)	
Odds ratio at mean coefficient	1.849	***	1.814	***
	(0.173)		(0.181)	
Pr(coefficient > 0)	0.789		0.783	
Socioeconomic variables controlled for	No		Yes	

Notes: The coefficient for the parent's field indicator is assumed to be normally distributed. A significant estimate is indicated by * (p<0.1), ** (p<0.05) or *** (p<0.01). The basic individual-level control variables include a woman indicator, a Swedish speaker indicator, high school graduation year and matriculation examination grades in first language and math. The set of socioeconomic variables includes mother's and father's vocational statuses, mother's and father's education levels and the type of pre-university region.

In the remaining estimations, heterogeneity in the parental effect across the field-of-study alternatives is examined. At first, the estimation is conducted using a version of model (5), in which each cross-field coefficient for the parent's field indicator (e.g., the coefficient for having a parent from education in the utility of choosing art/humanities) is restricted to zero, yielding the estimates in Table 5. The left column shows the odds ratio estimates obtained by conditioning on the basic demographic variables and matriculation examination grades. All of these estimates exceed one, suggesting that each field is, to some extent, transmitted from parents to children. However, the magnitude of the parental effect appears to vary across the fields. In particular, as already indicated by the descriptive evidence of Section 3, the degree of intergenerational transmission appears to be clearly the strongest in the case of law: having a parent with a law degree is associated with 6.93 times higher odds of graduating from law. Sizeable estimates are also obtained for education, business and medicine/health sciences, each of the estimates indicating more than twice higher odds of choosing the field for those having by a parent from the field. In the remaining fields, the parental effect appears to be smaller, and for two field-of-study categories (natural sciences and 'other field'), the odds ratio is not significantly different from one. According to the calculated Wald tests, many of the differences between the field-of-study-specific estimates in the left column are also statistically significant; in particular, the odds ratio for law is significantly different from all other odds ratios. A comparison of the results in the left and right columns of Table 5 shows that the field-of-study-specific estimates respond quite differently to the inclusion of socioeconomic variables in the model; for example, the odds ratios for art/humanities, law and medicine/health sciences reduce considerably, whereas those for education and natural sciences increase. In the right-column specification, less variation in the odds ratio exists across the fields, and the amount of significant differences is also smaller.

TABLE 5 Odds ratio for graduating from a field associated with having a parent with a degree in that field: results from conditional logit models with heterogeneous same-field coefficients (sample size: 3,869 individuals, 9 choice alternatives)

Field of study	Odds ratio					
Education	2.08	***	l, n	2.47	***	a, n
	(0.37)			(0.56)		
Arts/humanities	1.67	***	b, l, m	1.37	*	b, e, l, m
	(0.26)			(0.25)		
Business	2.90	***	a, l, n, s, t	2.77	***	a, n
	(0.52)			(0.56)		
Social sciences	1.63	**	b, l, m	1.71	**	l
	(0.33)			(0.39)		
Law	6.93	***	a, b, e, m, n, o, s, t	4.63	***	a, m, n, o, s, t
	(1.95)			(1.45)		
Natural sciences	1.29		b, e, l, m	1.43	*	b, e, l
	(0.24)			(0.30)		
Technology	1.89	***	b, l	1.95	***	l
	(0.31)			(0.37)		
Medicine/health sciences	2.63	***	a, l, n, s	2.40	***	a, l
	(0.50)			(0.55)		
Other field	1.56		l	1.45		l
	(0.60)			(0.60)		
Socioeconomic variables controlled for	No			Yes		

Notes: Standard errors are in parentheses. * (p<0.1), ** (p<0.05) and *** (p<0.01) indicate whether an odds ratio is significantly different from one. A significant difference (p < .1) to the estimate of another field is indicated by [a] (arts/humanities), [b] (business), [e] (education), [l] (law), [m] (medicine/health sciences), [n] (natural sciences), [o] (other field), [s] (social sciences) or [t] (technology). The basic individual-level control variables include a woman indicator, a Swedish speaker indicator, high school graduation year and matriculation examination grades in first language and math. The set of socioeconomic variables includes mother's and father's vocational statuses, mother's and father's education levels and the type of pre-university region.

Finally, the full version of model (5) is estimated, yielding estimates for both the same-field and cross-field effects from having a parent with a particular university degree. Table 6 shows the estimated odds ratios and marginal effects from a model controlling for the basic demographics and matriculation examination grades. The implied same-field parental effects – shown at the diagonal – are very similar to those in Table 5: most of the odds ratios suggest that having a parent with a particular field significantly increases the odds of choosing that field versus the base category (education). All of the estimated same-field marginal effects are also positive; however, the standard errors for these estimates are sizeable, due to which only one of the estimates – that for arts/humanities – is statistically significant. A comparison of the marginal effects to the predicted choice probabilities (reported at the last row of Table 6) reveals that the suggested magnitudes of the same-field effects are also very similar to the restricted model. Once again, the implied relative size of the parental effect is the largest in the case of law: the marginal effect obtained for this field (20.7 percentage points) implies that the probability of graduating from law is approximately 5 times higher for a student having a parent with a law degree than for the average student. For business and medicine/health sciences, the 'parental multiplier' is more than two, whereas smaller effects are measured for the rest of the fields. Most of the cross-field marginal effects are negative, suggesting that, in most cases, having a parent from a particular field is associated with a lower demand for another field. However, evidence of certain positive cross-field parental effects is also obtained. In particular, choosing law rather than education appears to be positively associated with having a parent from business or medicine/health sciences; the estimated marginal effects for the corresponding cross-field terms are also large and positive but not statistically significant.

The multinomial logit estimates conditioning also on socioeconomic variables are reported in Table 7. The impact of the additional controls is similar as in the previous conditional logit model (Table 5): while all of the same-field marginal effects remain positive, the suggested degree of intergenerational transmission reduces in some of the fields (arts/humanities, law and technology) and increases in some others (education and medicine/health sciences). The estimates of the parental effects on choosing law are, once again, particularly sensitive to the socioeconomic controls: the same-field marginal effect for this field in Table 7 is only half of that in Table 6 (10.8 percentage points), and the marginal effect of having a parent from business or medicine/health sciences on choosing law are clearly smaller than previously.

6 Summary and concluding remarks

Numerous studies from different countries have investigated transmission of educational attainment from parents to children, while there is scarcity of direct evidence on the transmission of educational fields. This paper has contributed to the filling of this gap by providing evidence of the intergenerational transmission of field at the highest level of Finland's

TABLE 6 Odds ratios and marginal effects (in italics) for graduating from a field associated with having a parent with a degree in a particular field: results from a multinomial logit model (sample size: 3,869 individuals, 9 choice alternatives)

	Child's field																	
Parent's field	Education	Arts/ humanities		Business		Social sciences		Law		Natural sciences		Technology		Medicine/ health sciences		Other field		
Education	**ref.**	0.48	***	0.37	***	0.50	***	0.35	**	0.55	**	0.50	***	0.46	**	0.75		
	0.099	*-0.022*		*-0.048*		*-0.009*		*-0.018*		*0.002*		*-0.006*		*-0.014*		*0.017*		
Arts/humanities	ref.	**2.38**	***	1.27		2.00	**	1.88	*	1.28		0.95		1.70		2.50	***	
	-0.049	***0.084***	*	*-0.032*		*0.029*		*0.008*		*-0.033*		*-0.033*		*0.004*		*0.023*		
Business	ref.	2.26	**	**5.02**	***	1.49		3.86	***	1.05		2.19	**	1.24		1.24		
	-0.067	*0.011*		***0.191***		*-0.038*		*0.039*		*-0.080*		*0.002*		*-0.040*		*-0.019*		
Social sciences	ref.	1.42		1.49		**1.93**	**	1.47		0.77		1.06		1.54		1.21		
	-0.030	*0.015*		*0.020*		***0.057***		*0.006*		*-0.065*		*-0.015*		*0.016*		*-0.003*		
Law	ref.	1.64		2.29		2.13		**11.23**	***	0.96		1.61		1.60		2.45		
	-0.065	*-0.040*		*0.013*		*0.002*		***0.207***		*-0.085*		*-0.018*		*-0.022*		*0.008*		
Natural sciences	ref.	0.66		0.85		0.73		0.66		**1.07**		0.96		0.51	*	1.12		
	0.025	*-0.037*		*0.003*		*-0.015*		*-0.010*		***0.043***		*0.012*		*-0.038*		*0.016*		
Technology	ref.	1.61		1.59		1.04		1.53		0.94		**2.33**	***	1.45		0.79		
	-0.033	*0.036*		*0.026*		*-0.029*		*0.007*		*-0.047*		***0.054***		*0.007*		*-0.020*		
Medicine/health sciences	ref.	1.30		1.41		1.16		3.03	***	1.77	*	1.55		**3.96**	***	1.46		
	-0.051	*-0.042*		*-0.024*		*-0.039*		*0.038*		*0.009*		*-0.006*		***0.121***		*-0.006*		
Other field	ref.	0.96		0.79		0.56		0.62		0.80		0.82		1.16		**1.33**		
	0.018	*0.019*		*-0.014*		*-0.045*		*-0.015*		*-0.013*		*-0.004*		*0.031*		***0.023***		
Predicted choice probability	0.123	0.185		0.148		0.124		0.050		0.153		0.077		0.094		0.045		

Notes: The marginal effects are evaluated at the means of the explanatory variables. The same-field odds ratios and marginal effects are emboldened. A significant estimate is indicated by * (p<0.1), ** (p<0.05) or *** (p<0.01). The set of control variables includes a woman indicator, a Swedish speaker indicator, high school graduation year and matriculation examination grades in first language and math.

TABLE 7 Odds ratios and marginal effects (in italics) for graduating from a field associated with having a parent with a degree in a particular field: results from a multinomial logit model controlling for socioeconomic variables (sample size: 3,869 individuals, 9 choice alternatives)

	Child's field								
Parent's field	Education	Arts/ humanities	Business	Social sciences	Law	Natural sciences	Technology	Medicine/ health sciences	Other field
Education	**ref.**	0.33 ***	0.29 ***	0.52 *	0.26 **	0.37	0.34 ***	0.61	0.65
	0.134	*-0.057*	*-0.059*	*0.015*	*-0.020*	*-0.033*	*-0.022*	*0.028*	*0.014*
Arts/humanities	ref.	**1.37**	0.95	1.63	0.93	0.86	0.56	1.60	1.85
	-0.017	***0.034***	*-0.027*	*0.048*	*-0.009*	*-0.042*	*-0.043*	*0.035*	*0.021*
Business	ref.	1.55	**3.49** ***	1.14	2.06	0.72	1.26	1.22	0.89
	-0.043	*0.006*	***0.184***	*-0.031*	*0.016*	*-0.085*	*-0.013*	*-0.019*	*-0.016*
Social sciences	ref.	0.84	1.11	**1.47**	0.64	0.56	0.61	1.41	0.94
	0.005	*-0.024*	*0.023*	***0.065***	*-0.015*	*-0.067*	*-0.030*	*0.043*	*-0.001*
Law	ref.	1.09	1.68	1.52	**4.89** ***	0.76	0.93	1.47	1.72
	-0.034	*-0.039*	*0.033*	*0.013*	***0.108***	*-0.072*	*-0.026*	*0.007*	*0.009*
Natural sciences	ref.	0.43 **	0.70	0.59	0.36 *	**0.78**	0.60	0.50	0.91
	0.066	*-0.066*	*0.012*	*-0.011*	*-0.020*	***0.033***	*-0.006*	*-0.022*	*0.015*
Technology	ref.	0.93	1.01	0.66	0.56	0.68	**1.29**	1.21	0.44
	0.014	*0.008*	*0.019*	*-0.033*	*-0.017*	*-0.038*	***0.034***	*0.032*	*-0.020*
Medicine/health sciences	ref.	0.79	1.09	0.82	1.31	1.26	0.90	**3.47** **	1.34
	-0.023	*-0.069*	*-0.016*	*-0.042*	*0.003*	*0.004*	*-0.021*	***0.162***	*0.003*
Other field	ref.	0.70	0.69	0.49	0.41	0.65 ***	0.55	1.32	**1.17**
	0.040	*-0.013*	*-0.012*	*-0.044*	*-0.020*	*-0.021*	*-0.021*	*0.071*	***0.020***
Predicted choice probability	0.124	0.190	0.147	0.124	0.044	0.158	0.080	0.095	0.037

Notes: The marginal effects are evaluated at the means of the explanatory variables. The same-field odds ratios and marginal effects are emboldened. A significant estimate is indicated by * (p<0.1), ** (p<0.05) or *** (p<0.01). The set of individual-level control variables includes a woman indicator, a Swedish speaker indicator, high school graduation year, matriculation examination grades in first language and math, mother's and father's vocational statuses, mother's and father's education levels and the type of pre-university region.

educational system. The obtained results suggest that a significant amount of intergenerational transmission occurs: even according to the most cautious estimate, having a parent with a university degree from a particular field is, on average, associated with a 1.8 times higher odds of graduating from that field among individuals with similar observable characteristics. However, the parental effect on the field-of-study choice appears to be heterogeneous in many ways. In particular, the results suggest that some fields are transmitted from parents to children more frequently than others, which is consistent with the previous findings of Van de Werfhorst et al. (2001) from the Netherlands. The intergenerational transmission appears to be clearly the strongest in the case of law: a student having a parent with a law degree was found to have a more than 5 times higher probability of graduating from law than an observably similar student without that type of parent. Relatively strong intergenerational transmission was also observed in the case of education, business and medicine/health sciences – the estimates for these fields were mainly above the estimated average effect – whereas the remaining fields appear to be transmitted to a lesser extent. The parental effect was also found to differ across genders – the estimated effects were larger for men than for women – and even among individuals with similar observable characteristics. In fact, the results from the mixed logit analysis suggest that a sizeable proportion of individuals place a negative weight on fields chosen by their parents. Furthermore, the results indicate that individuals are more likely to imitate their father's field-of-study choice than the choice of their mother; this conclusion is similar to the one obtained by Dryler (1998) for Swedish upper secondary students.

The sensitivity analyses reported in this paper yielded certain interesting observations. Namely, whereas the estimate of the average parental effect was found to be rather insensitive to different sets of control variables, conclusions regarding heterogeneity in the effect across the fields altered to some extent as a consequence of including socioeconomic controls in the random utility models. One potential explanation for this finding is that the impact of childhood socioeconomic environment on the intergenerational transmission differs across the fields. For example, based on the observed changes in the estimates, it appears that a reduction in socioeconomic differences across students' families could particularly decrease the degree to which the law field is transmitted while increasing the transmission of the education field. However, these conclusions are to be treated with caution, as the current data set does not enable distinguishing between causal and selection mechanisms in a reliable manner. In future research, it would be interesting to assess the extent to which the transmission of postsecondary fields from parents and children could actually be affected through policies, for example, by using structural changes in higher education systems as natural experiments.

REFERENCES

Arcidiacono, P. 2004. Ability sorting and the returns to college major. *Journal of Econometrics* 121*:* 343–375.

Beffy, M., D. Fougère, and A. Maurel. 2012. Choosing the field of study in postsecondary education: Do expected earnings matter? *Review of Economics and Statistics* 94: 334–347.

Behrmann, J.R., and M.R. Rosenzweig. 2002. Does increasing women's schooling raise the schooling of the next generation? *American Economic Review* 92: 323–334.

Berger, M. 1988. Predicted future earnings and choice of college major. *Industrial and Labour Relations Review* 41: 418–429.

Björklund, A., M. Lindahl, and E. Plug. 2006. The origins of intergenerational associations: Lessons from Swedish adoption data. *Quarterly Journal of Economics* 121: 999–1028.

Björklund, A., and K. Salvanes. 2011. Education and family background: Mechanisms and policies. In: Hanushek, E., Machin, S. & Woessmann, L. (Eds.), *Handbook of the Economics of Education*, Vol. 3, pp. 201–247, Elsevier.

Boudarbat, B. 2008. Field of study choice by community college students in Canada, *Economics of Education Review* 27: 79–93.

Bratti, M. 2006. Social class and undergraduate degree subject in the UK. IZA Discussion Paper no. 1979.

Checchi, D. 2006. The economics of education. Human capital, family background and inequality. Cambridge University Press, New York.

De Coulon, A., E. Meschi, and A. Vignoles. 2011. Parents' skills and children's cognitive and non-cognitive outcomes. *Education Economics* 19: 451–474.

De Haan, M. 2011. The effect of parents' schooling on child's schooling: A nonparametric bound analysis. *Journal of Labor Economics* 29: 859–892.

Dryler, H. 1998. Parental models, gender and educational choice. *British Journal of Sociology* 49: 375–398.

Goyette, K.A., and A.L. Mullen. 2006. Who studies the arts and sciences? Social background and the choice and consequences of undergraduate field of study. *Journal of Higher Education* 77: 497–538.

Haveman, R., and B. Wolfe. 1995. The determinants of children's attainments: A review of methods and findings. *Journal of Economic Literature* 33: 1829–1878.

Hole, A.R. 2007. Fitting mixed logit models by using maximum simulated likelihood. Stata Journal 7: 388–401.

Häkkinen, I., T. Kirjavainen, and R. Uusitalo. 2003. School resources and student achievement revisited: New evidence from panel data. *Economics of Education Review* 22: 329–335.

Leppel, K., M.L. Williams, and C. Waldauer. 2001. The impact of parental occupation and socioeconomic status on choice of college major. *Journal of Family and Economic Issues* 22: 373–394.

Montmarquette, C., K. Cannings, and S. Masheredjian. 2002. How do young people choose college majors? *Economics of Education Review* 21: 543–556.

Plug, E. 2004. Estimating the effect of mother's schooling on children's schooling using a sample of adoptees. *American Economic Review* 94: 358–368.

Sacerdote, B. 2007. How large are the effects from changes in family environment? A study of Korean American adoptees. *Quarterly Journal of Economics* 122: 119–157.

Train, K. 2003. Discrete choice methods with simulation. New York: Cambridge University Press.

Van de Werfhorst, H.G., N.D. De Graaf, and G. Kraaykamp. 2001. Intergenerational resemblance in field of study in the Netherlands. *European Sociological Review* 17: 275–293.

Wolniak, G.C., T.A. Seifert, E.J. Reed, and E.T. Pascarella. 2008. College majors and social mobility. *Research in Social Stratification and Mobility* 26: 123–139.

APPENDIX 1 List of variables used in the analysis

- *Field of study.* The field of the first completed university degree (bachelor's or master's) after graduation from high school. This variable includes 9 categories: education; arts/humanities; business; social sciences; law; natural sciences; technology; medicine/health sciences; other field (includes agriculture, forestry, military sciences and sports sciences).
- *Parent's field.* For a particular individual and field, this variable takes a value of 1 if at least one of the individual's parents has a university degree (bachelor's, master's or PhD) in the field, and 0 otherwise.
- *Mother's/father's field.* For a particular individual and field, this variable takes a value of 1 if the individual's mother/father has a university degree (bachelor's, master's or PhD) in the field, and 0 otherwise.
- *Female*. This variable takes a value of 1 if a woman and 0 otherwise.
- *Swedish speaker*. This variable takes a value of 1 if the individual's mother tongue is Swedish and 0 otherwise.
- *High school graduation year.* This variable includes 6 categories (years from 1991 to 1996).
- *First language grade*. The highest grade obtained in the first language test of the Finnish matriculation examination. The variable includes five categories: improbatur (I), approbatur (A) or lubenter approbatur (B); cum laude approbatur (C); magna cum laude approbatur (M); eximia cum laude approbatur (E) or laudatur (L); no grade available.
- *Math grade*. The highest grade obtained in the mathematics test of the Finnish matriculation examination. This variable includes 11 categories: B-level improbatur (I) or approbatur (A); B-level lubenter approbatur (B); B-level cum laude approbatur (C); B-level magna cum laude approbatur (M); B-level eximia cum laude approbatur (E) or laudatur (L); A-level improbatur (I) or approbatur (A); A-level lubenter approbatur (B); A-level cum laude

approbatur (C); A-level magna cum laude approbatur (M); A-level eximia cum laude approbatur (E) or laudatur (L); no grade available.

- *Mother's/Father's education level*. The level of mother's/father's highest educational qualification. These variables include 7 categories: primary school or no classification available; high school diploma; vocational diploma (at the secondary level); lowest tertiary-level qualification; bachelor's degree; master's degree; doctoral degree.
- *Mother's/Father's vocational status*. The latest observed type of the individual's mother's/father's vocation (observations in 1970, 1980, or 1990). These variables include 5 categories: farmer; entrepreneur; high-ranking official; low-ranking official; worker or no classification available.
- *Type of pre-university region*. The type of the NUTS-4 region in which the individual resided one year before the year of high school graduation. This variable includes 6 categories: sparsely populated region; rural region; industrial centre; regional centre; university region; metropolitan region.

APPENDIX 2

TABLE A1 Determinants of the field-of-study choice: odds ratios from a conditional logit model (sample size: 3,869 individuals, 9 choice alternatives)

	Edu	Arts/ Hum	Bus	Soc	Law	Nat	Tec	Med/ HS	Other
Parent's field	**1.96**	**1.96**	**1.96**	**1.96**	**1.96**	**1.96**	**1.96**	**1.96**	**1.96**
Female	ref.	**0.62**	**0.29**	**0.60**	**0.27**	**0.39**	**0.12**	0.86	**0.21**
Swedish speaker	ref.	0.71	1.38	1.29	0.75	1.08	0.80	1.16	**0.36**
High school graduation year (1996)									
1991	ref.	0.82	0.96	0.74	0.57	0.77	0.90	1.35	0.84
1992	ref.	0.99	1.10	0.89	0.88	1.02	0.91	**1.70**	0.73
1993	ref.	0.98	**0.51**	0.76	0.71	0.64	**0.57**	0.91	0.77
1994	ref.	1.00	0.77	0.80	0.61	0.70	**0.68**	0.94	0.76
1995	ref.	0.97	1.17	0.93	0.69	1.11	0.89	**1.57**	0.77
First language grade (E or L)									
Not available	ref.	**0.33**	**3.40**	0.84	0.88	**4.84**	**7.83**	0.66	**2.89**
I, A or B	ref.	**0.11**	1.03	**0.35**	**0.16**	1.32	**2.87**	1.05	1.26
C	ref.	**0.28**	0.99	**0.35**	**0.42**	1.33	1.38	0.92	0.86
M	ref.	**0.35**	0.88	**0.47**	0.74	0.88	1.06	**0.59**	1.29
Math grade (E or L, A-level)									
Not available	ref.	1.29	**0.07**	**0.45**	**0.44**	**0.03**	**0.01**	**0.04**	**0.15**
I or A, B-level	ref.	**0.55**	**0.07**	**0.29**	**0.17**	**0.02**	**0.00**	**0.01**	**0.07**
B, B-level	ref.	**0.45**	**0.08**	**0.31**	**0.30**	**0.02**	**0.00**	**0.02**	**0.18**
C, B-level	ref.	**0.49**	**0.09**	**0.27**	**0.14**	**0.01**	**0.00**	**0.03**	**0.25**
M, B-level	ref.	**0.48**	**0.21**	**0.31**	**0.44**	**0.03**	**0.01**	**0.02**	**0.12**
E or L, B-level	ref.	0.72	**0.37**	**0.44**	0.66	**0.08**	**0.01**	**0.06**	**0.30**
I or A, A-level	ref.	0.42	**0.17**	**0.34**	0.37	**0.06**	**0.02**	**0.08**	**0.29**
B, A-level	ref.	0.50	**0.21**	**0.35**	0.50	**0.10**	**0.03**	**0.07**	0.51
C, A-level	ref.	0.56	**0.45**	**0.29**	0.59	**0.19**	**0.10**	**0.15**	0.55
M, A-level	ref.	0.70	0.64	0.66	1.10	**0.48**	**0.32**	**0.34**	0.99
Mother's vocational status (Worker or not available)									
Farmer	ref.	**2.13**	1.47	1.08	1.12	**2.64**	**3.48**	1.96	**3.64**
Entrepreneur	ref.	0.80	**1.97**	1.42	3.01	1.74	1.45	1.80	0.32
High-ranking official	ref.	0.91	1.47	0.85	1.79	1.25	1.11	1.24	0.74
Low-ranking official	ref.	0.89	0.90	1.12	1.62	1.21	1.03	0.92	1.03
Mother's educational level (Primary level or not available)									
High school diploma	ref.	1.07	**0.55**	**0.46**	1.24	0.69	**0.50**	0.89	1.04
Vocational diploma	ref.	0.97	**0.64**	0.76	0.78	1.02	0.98	0.98	1.38
Lowest tertiary level	ref.	1.21	0.85	0.76	1.02	1.14	1.16	1.38	**2.73**
Bachelor's degree	ref.	1.44	0.76	0.89	1.03	1.35	1.21	0.79	**4.01**
Master's degree	ref.	**1.42**	0.63	1.06	1.55	1.07	1.21	1.48	2.93
Doctoral degree	ref.	0.62	**0.25**	1.40	1.03	0.81	0.45	**0.55**	**3.87**

TABLE A1 (continued)

	Edu	Arts/ Hum	Bus	Soc	Law	Nat	Tec	Med/ HS	Other
Father's vocational status (Worker or not available)									
Farmer	ref.	0.67	1.06	0.95	0.77	0.92	0.81	1.34	1.87
Entrepreneur	ref.	0.93	1.05	0.93	0.99	0.67	0.75	1.02	**0.52**
High-ranking official	ref.	1.00	1.03	0.92	0.87	1.02	0.74	1.11	0.99
Low-ranking official	ref.	1.10	0.93	0.84	0.95	0.85	0.85	1.15	**0.45**
Father's educational level (Primary level or not available)									
High school diploma	ref.	1.12	1.07	1.03	1.20	1.09	1.37	0.92	1.98
Vocational diploma	ref.	1.06	1.08	1.30	0.79	1.02	1.02	1.11	1.41
Lowest tertiary level	ref.	1.11	**2.13**	**1.76**	**1.80**	1.43	1.38	0.98	1.78
Bachelor's degree	ref.	0.96	1.23	1.19	0.81	0.98	1.57	0.81	1.14
Master's degree	ref.	1.21	1.49	1.35	1.45	0.91	**1.47**	0.97	1.42
Doctoral degree	ref.	1.42	1.32	1.41	1.10	1.04	**2.03**	1.10	1.47
Type of pre-university region (Sparsely populated region)									
Rural region	ref.	0.73	**0.52**	**0.71**	0.52	1.09	0.74	**0.67**	**0.42**
Industrial centre	ref.	0.63	**0.40**	**0.62**	**0.20**	0.88	**0.65**	**0.47**	**0.59**
Regional centre	ref.	**0.55**	**0.37**	**0.55**	**0.26**	0.86	**0.58**	**0.55**	0.46
University region	ref.	0.69	**0.49**	**0.36**	**0.22**	0.71	**0.50**	**0.41**	**0.49**
Metropolitan region	ref.	0.77	**0.28**	**0.53**	**0.17**	0.67	**0.46**	**0.44**	0.45
Constant	ref.	**7.38**	**15.87**	**8.28**	4.58	**16.03**	**95.42**	**9.22**	**2.06**

Notes: The references categories of the caterogorical variables are in parentheses. Significant odds ratios ($p<.1$) are emboldened. Abbreviations: Edu = education; Hum = humanities; Bus = business; Soc = social sciences; Nat = natural sciences; Tec = technology; Med/HS = medicine and health sciences.

CHAPTER 4: ARE THERE RETURNS FROM UNIVERSITY LOCATION IN A STATE-FUNDED UNIVERSITY SYSTEM?[50]

A location in an economically active, high-amenity region could in many ways be a significant advantage for a university and its students and thus could also be positively linked to students' subsequent earnings. Based on this hypothesis, the present study empirically examines the effect of university location choice on earnings in Finland, focusing on the following question: To what extent does the choice of university location explain the observed positive early-career earnings premium for students graduating from the Helsinki metropolitan area rather than from one of the nine other university cities? The results suggest that no positive average earnings premium exists for metro area graduates after differences in students' pre-university characteristics are taken into account. However, the metro area university premium is found to be, to some extent, heterogeneous across fields and regional labour markets. The findings also indicate the importance of accounting for the selective nature of individuals' migration behaviour when conditioning on post-university region in the estimation of university location effects.

Keywords: earnings, university choice, regional labour markets, selective migration

---

50 This article has been published as: Suhonen, T. 2013. Are there returns from university location in a state-funded university system? *Regional Science and Urban Economics* 43(3), 465–478. Financial support from the Academy of Finland is gratefully acknowledged (grant number 127049). I also wish to thank Katariina Nilsson-Hakkala, Hannu Tervo, Olle Westerlund, the two anonymous referees and the participants at the ERSA 2008 and EEA-ESEM 2009 congresses for providing helpful comments.

1 Introduction

A large body of literature has studied the causal effect of university choice on students' earnings after graduation. In particular, numerous studies from the U.S. have investigated whether attending a prestigious university or college increases a student's subsequent earnings, typically concluding that a positive link exists (e.g., Behrman et al., 1996; Brewer et al., 1999; Monks, 2000; Dale and Krueger, 2002; Black and Smith, 2004; 2006; Long, 2008; Hoekstra, 2009; Hershbein 2011).[51] However, very few studies have explicitly examined whether earnings are affected by the geographical location of the university. As students in many countries select the cities in which they attend a university from a large number of heterogeneous alternatives, the locational aspect of university choice could be important. In particular, one may suspect that being located in an economically active, high-amenity region could, in many ways, be a significant advantage for a university and its students and thus could also be positively linked to students' subsequent earnings. In this paper, the relationship between university location choice and students' early-career earnings is studied in Finland. Given that Finland is a relatively vast and scarcely populated country with a geographically dispersed public university system, the topic may be of practical relevance to many: Finnish university applicants may wonder whether it is worthwhile to apply to and attend a distant university instead of one located nearby, whereas policy makers may be concerned with whether the system generates earnings inequality on the basis of students' locational choices.

From a theoretical perspective, several mechanisms could cause early-career earnings differences between graduates from different university cities conditional on individuals' pre-university characteristics. In this paper, we discuss some of these mechanisms, focusing on the most obvious ones. One particularly important mechanism could be regional variation in university quality: graduates from regions that have prestigious universities – such as those with good reputation, high-quality resources and high entrance requirements – could earn more than other graduates, for example, because of a higher level of human capital acquired by these graduates during their studies.[52] In effect, one may suspect that quality differences between

---

51 In recent years, some evidence on the earnings effects of university choice has also been presented from European countries, including the U.K. (Chevalier and Conlon, 2003), Italy (Brunello and Cappellari, 2008; Triventi and Trivellato, 2012) and Sweden (Lindahl and Regnér, 2005; Eliasson, 2006; Holmlund, 2009). While most prior studies find significant returns from college/university choice, some contradicting evidence also exists. In particular, Dale and Krueger (2002) find no premium for attending a highly selective college after controlling for students' application behaviour and admission records.

52 Studying at a prestigious university could enhance one's human capital accumulation, e.g., because of high-quality instruction or peer effects (Sacerdore, 2001; Zimmerman, 2003; Dale and Krueger, 2002). A positive link between university choice and labour market outcomes could also arise because employers might interpret graduation from a prestigious university as a signal of high worker

universities within a university system are partly determined by their locational differences. This hypothesis – serving as a working hypothesis for the current research – arises because students and staff members may base their choice of university partly on universities' locational characteristics such as the quality of the local labour market and other location-specific amenities.[53] In particular, these skilled individuals could be drawn to large cities because of good consumption possibilities and other urban amenities or, alternatively, because of human capital externalities that are expected to result from living and working in a densely populated labour market (see Glaeser et al., 2001; Glaeser and Maré, 2001; Storper and Scott, 2009). This behaviour promotes competition on student places and jobs at universities located in attractive city regions, resulting in ability sorting among students and staff and hence in differences in the quality of universities' human resources across regions.[54]

Aside from university quality,[55] also other locational features could be important regarding students' subsequent earnings. In particular, opportunities to work in the local labour market during one's studies might matter: students in large and economically active city regions might be more able than students in smaller cities to acquire relevant work experience prior to graduation, which could enhance these students' field-of-study-specific human capital, networks and/or productivity signals and thus help them to receive higher early-career earnings (Häkkinen, 2006). Another important earnings effect may result from the limited mobility of university graduates across regions: after graduation, individuals might be unwilling to migrate far from their region of graduation, even in the presence of economic gains from doing so, and thus be systematically sorted into high- and low-earnings regions based on their university location (Lindahl and Regnér, 2005; Brunello and Cappellari, 2008). In particular, students from large university cities may be more inclined than

---

productivity and use university choice as a screening device when recruiting new workers (Hershbein 2011; Lang and Siniver, 2011). Alternatively, students in prestigious universities could be more able than other students to establish networks that help in their subsequent careers (Lang and Siniver, 2011).

53 Different pieces of empirical evidence support the view that locational characteristics matter. For instance, the questionnaire results of Keskinen et al. (2008) suggest that university location is an important factor in the choice of university among Finnish psychology students. Furthermore, recent evidence on the interstate migration flows of college students in the U.S. suggests that apart from migration distance and college characteristics, different geographical variables affect students' choice of college state (Cooke and Boyle, 2011).

54 The full picture of the theoretical linkages between university quality and regional differences is, naturally, more multifaceted. In particular, studies have acknowledged that universities may have both short-term and long-term effects on local regional development (Faggian and McCann, 2009), and thus a reverse causal link between university quality and locational characteristics could also exist.

55 When using the term 'university quality', it must be acknowledged that this term admits several definitions, e.g., those derived from reputation, inputs, outputs, process, content and value-added (Adams, 1997). Here, as in many previous studies, the terms 'quality' and 'selectivity' are used almost interchangeably: a university's ability to acquire high-ability students and staff may depend on the attractiveness of its location, and hence, e.g., the level of peer effects and the quality of teaching may be higher in more attractive locations.

other students to reside in these cities after graduation and are thus more likely to benefit from an 'urban wage premium' (see Glaeser and Maré, 2001). Studies examining returns to university quality have been rather silent about these 'other' mechanisms behind university location effects despite their practical implications for the analysis: as high-quality universities are usually located in attractive city regions, one may end up overestimating the actual effect of attending a high-quality university if one is unable to adjust the estimates for locational differences between universities.[56]

Based on the general features of Finland's university system, competing conjectures regarding the existence of returns from university location choice in this country may be advanced. The political tradition in Finland has favoured equality among the universities, and the state-funded university system has evidently strived towards an equal distribution of educational resources, for example, by providing sufficient core funding for the universities to cover their educational expenses (see Ministry of Education, 2005). Thus, if the system has succeeded in preventing a stratification of universities by educational quality, university choice may have been rather insignificant with respect to students' outcomes.[57] However, the Finnish university system is also geographically rather dispersed, and government decisions have placed universities in considerably heterogeneous locations, ranging from small and remote cities of less than 100,000 inhabitants (such as Rovaniemi and Joensuu) to a metropolitan area of one million inhabitants (Helsinki). Thus, one may suspect that returns from university location choice have arisen because the cities differ in their ability to attract talented individuals and provide jobs for students during their studies and after graduation. The magnitude and persistence of these effects could have been further boosted by the relatively low interregional mobility of university graduates in Finland.[58] Against this background, the predominance of the Helsinki metropolitan area over the other nine university cities in the provision of academic jobs and different urban amenities raises an interesting question: Are Finnish students better off attending a university in Helsinki? Given that nearly all university disciplines can be studied both inside and

---

56 The usual approach in previous studies has been to only control for students' pre-university characteristics when estimating returns to university choice. Some studies have, however, discussed and attempted to control for regional earnings differences by simply conditioning on students' post-university region (see Lindahl and Regnér, 2005; Brunello and Cappellari, 2008; Holmlund, 2009).

57 Previous findings on the effect of university choice from highly similar European university systems have been rather mixed. Whereas Lindahl and Regnér (2005) find a significantly positive premium for graduating from an 'old' university in Sweden, more recent Swedish studies by Eliasson (2006) and Holmlund (2009) suggest that the relationship between university choice and earnings is rather weak. Furthermore, the findings of Brunello and Cappellari (2008) from Italy suggest that graduation from a certain type of university – in particular, private universities or universities located in the Northern Italy – is positively associated with labour market outcomes. However, another recent Italian study by Triventi and Trivellato (2012) does not find a significant premium for attending a private university.

58 A recent study by Haapanen and Tervo (2012) has shown that Finnish university graduates are indeed rather immobile, as most of them do not move from their region-of-graduation within 10 years after graduation.

outside Helsinki, many university applicants may find this question relevant. In the empirical analysis presented below, we therefore focus on examining the effect of graduating from a 'metro area university', that is, a university located in Helsinki, on students' early-career earnings.

The study utilises administrative data on a sample of Finnish university graduates from 1994–2000 while paying close attention to two problems involved in estimating the *metro area university premium*. The first problem is the endogeneity of university location: if students sorting into the metro area universities are of higher average ability than other students, they could receive higher average earnings after graduation regardless of the university location chosen. Another challenge arises from the above-mentioned fact that the choices of university location and post-university residential location may be highly correlated because of limited interregional mobility. Therefore, to assess the earnings effects related directly to university location choice – such as that arising from university quality – one should be able to isolate the indirect 'post-university region effect' from the total effect. Different approaches are employed to address these issues. The selection problem[59] is attempted solve either by controlling for a rich set of students' pre-university characteristics such as their matriculation examination grades and parents' socioeconomic status or by using time variation in the *supply of metro area student places*[60] as an instrumental variable (IV) for graduating from a metro area university. The aim of separating the indirect post-university region effect from the total effect is pursued either by simply conditioning on regional variables in a linear model or by estimating Heckman-type sample selection models for the groups of 'residents in metro area' and 'residents in other regions', i.e., individuals who resided either inside or outside the Helsinki metropolitan area after graduation. The latter approach is based on the idea presented, e.g., by Heckman et al. (1996), Dahl (2002) and McHenry (2011) that the selective nature of individuals' locational choices after completing their education must be accounted for to be able to study returns to education across regional labour markets.

The paper is organised as follows. A brief description of the Finnish university system and a comparison of Finnish university cities are included in Section 2. In Section 3, the data are discussed, and descriptive evidence of the differences between metro area graduates and other graduates is presented. Section 4 discusses the methodology used for the estimation of university location effects. The results are reported in Section 5, and Section 6 presents a summary and concluding remarks.

---

59 Here, the term 'selection' refers to both students' self-selection and institutional selection: in the selection process, students decide which universities to apply to, whereas universities accept and reject students on the basis of different criteria. Furthermore, students must also decide whether to accept an offer if admitted.

60 The 'supply of metro area student places' refers to the relative number of student places available in the metro area universities. In the estimation, the Helsinki metro area's proportion of first-year students in an individual's field of study is used to approximate this supply (see Appendix B).

2 The Finnish university system

The Finnish university system is a publicly financed and administered system responsible for the provision of academic research and education at the bachelor's, master's and PhD levels in Finland.[61] In the 1990s, when the university graduates in the study sample completed their studies, the university system consisted of 16 universities (10 multidisciplinary universities, 3 business schools and 3 universities of technology), 4 art academies and the Finnish National Defence University. All of these institutions continue to exist today, but some of the institutions have merged. Four of the universities are located in the Helsinki metropolitan area, and the remaining 12 universities are located in 9 smaller city regions: Turku, Tampere, Lappeenranta, Kuopio, Joensuu, Jyväskylä, Vaasa, Oulu and Rovaniemi.[62] All of the art academies and the National Defence University are in Helsinki. Admission to the institutions is generally based on applicants' high school grades (particularly grades from the matriculation examination) and a field-of-study-specific entrance examination. Studying is free of charge and publicly subsidised by the provision of study grants, housing allowances and state-guaranteed loans.[63] Thus, virtually no financial barriers are involved in university participation or the choice of institution. Nearly all fields of study are available both inside and outside the Helsinki metropolitan area; in the 12-category classification of fields used in the current analysis, only two exceptions exist: sports sciences are only available in Jyväskylä, whereas military sciences may be studied only in Helsinki. Thus, almost regardless of the field chosen, a Finnish university applicant faces the choice between a 'metro area university' and 'another university', which justifies the binary viewpoint of this paper.

The universities are under the supervision of the Ministry of Education and to a major extent publicly financed: approximately 65% of the universities' revenues consist of non-competitive core funding from the state, while the remaining 35% are competitive external funding from various public and private organisations such as the Academy of Finland, the Finnish Funding Agency for Technology and Innovation (TEKES), the EU and different foundations (Ministry of Education, 2005). Arguably, the state-funded system

---

61 The Finnish higher education system also consists of another sector: polytechnics. This sector rewards professionally oriented degrees that are usually paralleled by the bachelor's degree. The polytechnics were excluded from the current analysis, as these schools were not officially established until 1996; given the span of the current data, this fact limits the analysis of labour markets for polytechnic graduates.

62 In addition, smaller university consortiums are currently located in six other cities. However, as these consortiums were only officially established in 2004, the individuals in the current sample almost entirely come from the 'old' university regions.

63 The generous social security benefits are believed to be important in enabling students to study full time and to live separately from their family-of-origin regardless of socioeconomic background. In Finland, university students have indeed been rather independent: according to the questionnaire by Lempinen and Tiilikainen (2001), only 6% of the students lived with their parents in 2000.

has supported an equal distribution of educational resources across universities: universities' teaching expenses have been almost entirely determined by the core funding, whereas teacher salaries in all universities have been paid according to a common payroll system; thus, one may argue that the universities have neither competed for the financial resources used for undergraduate education nor have been able to engage in a serious wage competition for skilled professors.[64] These features – which are very common among European public university systems – are generally believed to prevent the stratification of universities by quality (Villalonga, 2011).

Table 1 provides an overview of the university cities' characteristics. The information obtained from the ALTIKA database (Statistics Finland, 2012a) regarding regional characteristics shows that the socio-economic environment differs considerably across locations. Helsinki particularly stands out by being roughly as large in population size as the other nine university cities combined, by having the highest average education, income and housing price levels and by offering more jobs in skill-intensive sectors and cultural activities than the other cities. Thus, based on the earlier discussion in Section 1, one might suspect that by offering the most urbanised environment, Helsinki provides significant advantages – such as human capital externalities and networks – for its university students and staff members, which could also contribute to students' outcomes.

Table 1 also shows information from the KOTA database (Ministry of Education, 2012) regarding universities' characteristics in different cities. The information on the admissions ratios suggests that certain differences in the popularity of the university cities have existed. In particular, the relatively small number of applicants in some of the smallest cities – Lappeenranta, Kuopio and Vaasa – could reflect problems in the ability of these cities to attract students. The admissions ratios do not, however, support the view that students would systematically prefer Helsinki to all other alternatives, as this ratio is only the fourth highest in Helsinki. A remarkable finding from Table 1 is that despite the general heterogeneity of the university cities, no large variation in the universities' average staff salaries exists. For example, university employees in Helsinki and Kuopio earned an approximately equal average salary despite that these two cities differ greatly in location, size, cost of living, etc. This finding supports the view that the state-funded system has treated universities rather equally.

Table 1 also shows that the number of teachers per 100 students varies to some extent – from 4.8 (Tampere) to 8.6 (Kuopio) – across the cities. However, these figures are obviously not very informative about differences in the

---

64 Although universities also receive a considerable amount of external funding, these funds are mainly dedicated to research and not education. For example, statistics show that over the last two decades, more than 90% of universities' teaching staffs were salaried using the core funding, whereas more than 70% of the research staff received their salaries from external sources (Ministry of Education, 2012). Moreover, although some differences in average professor salaries have existed between universities, these differences have hardly been large enough to create competitive advantages for universities (see Professoriliitto, 2011).

TABLE 1 Characteristics of the Finnish university cities

	Helsinki	Turku	Tampere	Lappeenr.	Kuopio	Joensuu	Jyväskylä	Vaasa	Oulu	Rovaniemi
Regional characteristics										
Inhabitants	964,953	173,686	197,774	70,918	92,915	71,257	118,380	57,014	126,569	56,991
Inhabitants/km^2	1253	707	377	49	58	30	101	302	90	8
Highly educated per 100 inhabit.	17	12	13	8	10	9	12	12	14	9
R&D establishments	125	25	24	2	11	2	8	2	16	0
Data processing establishments	1925	210	338	63	83	47	142	69	169	40
Cultural establishments	1601	175	265	43	70	48	119	62	103	66
Average annual income (€)	23,507	17,748	18,495	17,007	16,910	15,659	17,397	18,212	18,859	16,394
Average housing price (€/m^2)	2255	1310	1488	1231	1303	1213	1359	1200	1386	995
University characteristics										
Universities (incl. art academies)	8	3	2	1	1	1	1	1	1	1
Undergraduate students	48,944	18,563	19,087	3642	3436	5502	10,389	3437	11,132	2998
Applicants per admitted student	4.2	3.6	4.8	2.3	3.0	3.8	4.8	2.5	3.4	4.4
Teachers per 100 students	5.7	6.7	4.8	5.7	8.9	6.7	6.2	4.8	7.8	6.1
Salary costs per staff member (€)	34,037	36,544	34,764	33,377	33,635	36,441	35,049	32,969	35,373	33,073

Notes: In calculating the figures, the cities were defined according to municipalities: Helsinki consists of four municipalities, Helsinki, Espoo, Vantaa and Kauniainen, whereas the rest of the university cities consist of only one municipality. The regional characteristics were obtained from the ALTIKA database (Statistics Finland, 2012a); these figures are from 2001, except that the average housing prices are from 2003. The university characteristics were obtained from the KOTA database (Ministry of Education, 2012); these figures were calculated by taking averages of yearly values from 1997–2000.

availability of educational resources because field-of-study options vary considerably across the cities. A more accurate picture is provided in Figure 1, showing the relative differences in the teacher-student ratio between Helsinki and the other cities in eight fields of study between 1991 and 2000. These figures suggest that despite the state-funded university system, certain differences in educational resources existed between metropolitan and non-metropolitan universities. In fact, for most of the years and fields, the teacher-student ratio was higher outside than inside the metro area. The most notable difference existed in humanities, where the ratio was systematically approximately 40% lower in Helsinki than elsewhere. The variation across fields in the gap between Helsinki and the other cities is likely to be partly explained by the fact that universities have been able to independently allocate their resources across faculties; therefore, even if educational resources were allocated rather uniformly across universities, a particular discipline may have not been treated similarly in all universities. We return to these differences in educational resources in Section 5.2 when discussing potential explanations for field-of-study-specific findings.

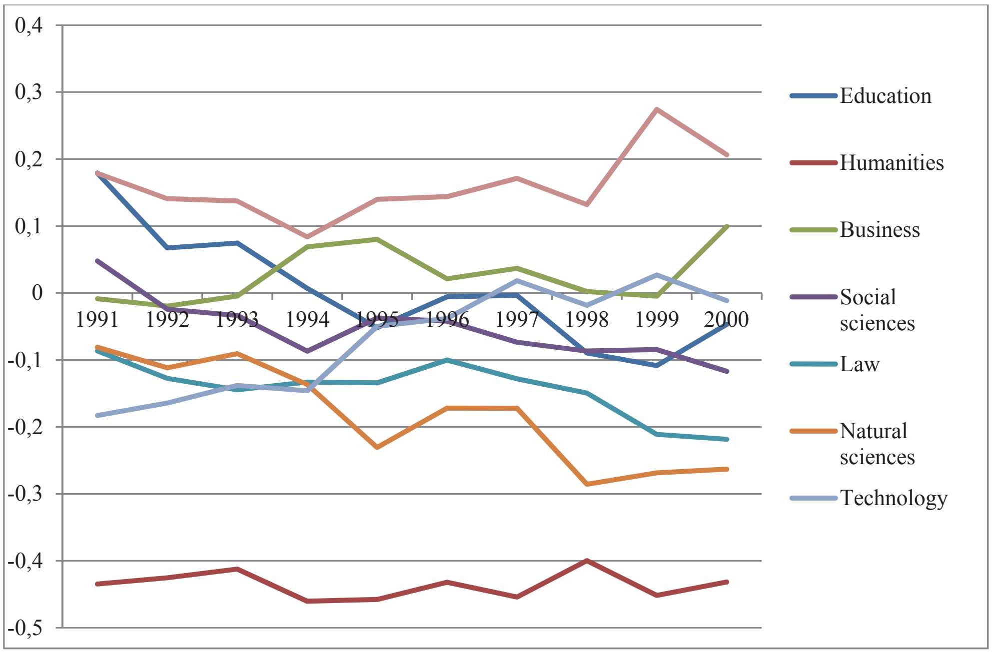


FIGURE 1 Relative difference in the teacher-student ratio between metro area universities and other universities in eight fields of study. Source: Ministry of Education (2012).

3 Data and descriptive analysis

The data set of this study is based on a 7% random sample drawn from the population of Finland in 2001.[65] These data originate from the registers of Statistics Finland and include rich panel information, e.g., on individuals' labour market outcomes, residential region, job characteristics, educational qualifications and family background from the period between 1970 and 2006. Only a small portion of the full data set is used for the purposes of this study: the examined subsample includes individuals who graduated from one of the Finnish universities between 1994 and 2000. Bachelor's and master's level graduates from all fields of education except sports and military sciences are included in the analysis.[66] Furthermore, the analysis is restricted to those who participated in the Finnish matriculation examination in 1985, 1987–1989 or 1991 or later, as high school graduates' matriculation grades in math and first language are observed only from these years in the data.[67] Additionally, observations with missing information regarding post-university residential location or high school graduation year are excluded from the sample.

With regard to the analysis of university location effects, the data set is strong because it allows for the examination of individuals' labour market performance for several years after graduation, whereas many previous papers examining the effects of university/college choice have relied on information from a single cross-section year. The dependent earnings variable used in the analysis is constructed by calculating the sum of taxable income over six years after the year of graduation; for instance, for an individual who graduated in 2000, the earnings are calculated from the years between 2001 and 2006.[68] The use of these six-year earnings may be preferred to one-year earnings, as no observations must be excluded because of short-run distortions in labour

---

65 The restriction on the sample size (7%) is based on the privacy protection regulations of the Finnish Statistics Act.

66 Graduates from sports and military sciences are excluded because, as noted in Section 2, university location does not vary within these groups. Therefore, the treatment of choosing a metro area university does not exist for students applying for these fields. Until 2005, completing a bachelor's degree before the master's degree was optional, and university students rarely selected this option; this fact explains the small proportion of graduates with a bachelor's degree (7%) in the study sample.

67 Finnish high school students participate in the matriculation examination during their final year of high school. The examination includes numerous standardised tests, some of which are mandatory and some are not. Passing the first language test is mandatory for the completion of the examination; due to this requirement, the current sample is in practice formed by excluding individuals whose first language grade is missing from the sample of university graduates. However, those with missing grades for mathematics are not excluded because the math test is not compulsory: dropping these observations would result in a selection problem. In the sample used, the students without a math grade only comprise 14%.

68 Altogether, the earnings information for the sample was gathered from the period 1995–2006. In Finland's economic history, this period marked an era of strong economic boom after a deep recession. During the period, the average annual GDP growth was 3.8%, and the unemployment rate decreased every year, being 15.4% in 1995 but only 7.7% in 2006 (Statistics Finland, 2012b).

market participation; therefore, the results are not confounded by self-selection in participation. The data set still has one potentially important limitation: university drop-outs are not observed. This limitation must be kept in mind when interpreting the results: if university location choice affects the probability of dropping out, estimates could suffer from selection bias when interpreted as the effects of university attendance.[69]

After all necessary restrictions, the estimation sample includes 2658 individuals.[70] Of these, 890 individuals (33.5%) graduated from the Helsinki metropolitan area, and the remaining 1768 individuals (66.5%) graduated elsewhere. Table 2 presents descriptive evidence of the differences between metro area graduates and other graduates. A clear gap in earnings exists between the groups: metro area graduates earned on average 4053 € (13.6%) more per year during the six-year period after graduation. The relative gap is somewhat larger for men (13.7%) than for women (9.5%). One obvious explanation for the earnings gap arises from field-of-study differences between universities: compared to his or her counterpart from outside the metro area, the average metro area student more frequently graduated from some of the most lucrative fields (business, law and technology) and more rarely from some of the less lucrative ones (education and natural sciences). However, even when examining the fields separately, sizeable differences are found between metro area graduates and other graduates: the earnings gap is significantly positive (at the 10% level) for graduates in business (16.9%), social sciences (22.1%), law (20.6%) and agriculture and forestry (45.1%) while significantly negative (–10.5%) for humanities graduates; the remaining differences are clearly smaller and insignificant.

Given that earnings differences exist between metro area graduates and other graduates, an obvious question follows: Do the raw earnings gaps reflect genuine earnings effects from university location choice or merely compositional differences between the groups? As can be seen from Table 2, apart from earnings and fields of study, metro area graduates and other graduates also differ with respect to several other characteristics. In particular, the average metro area graduate has significantly higher matriculation examination grades in math and first language, suggesting that these graduates

69 The earliest available statistics on the discontinuation of education in Finland show that between September 2002 and September 2003, 5.0% of university students ceased their studies. This proportion was slightly smaller for students in Helsinki (4.8%) than for those in other cities (5.1%). (Statistics Finland, 2012b) Based on this information, it is though difficult to say whether university location choice is associated with dropping out, mainly because fields of study and students' characteristics vary across universities.

70 The original sample of university graduates from 1994–2000 includes 4846 observations. By excluding individuals who did not participate in the matriculation examination in the given years (1985, 1987–1989 or 1991 or later), the sample size is reduced to 2752. Additionally, 48 observations are excluded because of missing information on post-university residential location, and 43 observations are excluded because of the inability to determine the high school graduation year (this information is required for the IV strategy discussed in Section 4). In addition, three observations with zero earnings are dropped as the natural logarithm of earnings is used in the regression analysis.

were already mathematically and linguistically more skilled before entering university studies.[71] Thus, in line with predictions, students appear to have sorted into metro area universities on the basis of ability, which again indicates that part of the positive metro area university premium could arise from effects (such as peer effects) associated with attending a highly selective university. Another notable difference between the groups is that metro area graduates already worked significantly more than other graduates before graduating from a university. Thus, as argued in Section 1, regional differences in job opportunities offer another potential explanation for the higher average earnings of metro area graduates. Alternatively, the differences in average grades, early work experience and other characteristics could indicate that the positive metro area university premium is merely a result of selection bias.

Table 2 shows that high school location is another important determinant of university choice in Finland; because of the limited interregional mobility of high school graduates, individuals from the Southern Finland are significantly more likely to end up in a metropolitan university than a non-metropolitan university, whereas the opposite applies for individuals from the west, east and north. Altogether, the degree of immobility among Finnish university students crystallises in the following three figures calculated from the data: 1) 37% of students graduated from a university in the same NUTS-3 region where they completed high school; 2) 49% of students resided in the same NUTS-3 region where they graduated from a university six years after graduation; and 3) 47% of students resided in the same NUTS-3 region where they graduated from high school six years after graduating from a university. In addition, the data show that 30% of those who did not graduate from a university in their pre-university region returned to their pre-university region and resided there six years after graduation.

In the light of the above-discussed evidence, the limited mobility of university graduates could be an important mechanism generating university location effects in Finland. In particular, the positive metro area university premium could reflect the fact that graduates from Helsinki have a higher probability of ultimately living and working in Helsinki, which is the largest and most diverse labour market area in Finland. To illustrate this phenomenon, average labour market outcomes are shown separately for two groups of graduates in Table 3: 1) 'residents in metro area', i.e., graduates who resided in the Helsinki metropolitan area continuously between the third and sixth year after graduation and 2) 'residents in other regions', i.e., graduates who resided outside the metro area for the corresponding period.[72] The average labour

71 Naturally, the differences in the average math and first language grades between Helsinki and other cities partly reflect differences in field-of-study options, as students self-select into different fields on the basis of the grades. However, a closer analysis reveals that the average grades are also higher for metro area graduates within most of the field-of-study groups.

72 According to the data, a large proportion of individuals change their residential location during the first and second years after graduation, but after this, the migration frequency declines. Therefore, when analysing labour market outcomes separately for different regions, it is reasonable to define the sub-samples based on

TABLE 2 Means of selected variables for metro area graduates and other graduates

	Metro area graduates	Other graduates	
Average annual earnings (€)	33,880	29,828	***
For men	41,253	36,295	***
For women	27,862	25,436	***
Woman	0.55	0.60	**
Age at graduation	27.10	26.69	***
Swedish speaker	0.08	0.05	***
Matriculation grade in first language [a]	4.32	3.99	***
Matriculation grade in math [a]	3.91	3.50	***
Math test taken at the A-level	0.70	0.59	***
No math grade	0.14	0.15	
Months worked before graduation [b]	20.89	14.80	***
Bachelor's degree	0.05	0.08	***
High school location [c]			
South	0.64	0.20	***
West	0.19	0.44	***
East	0.09	0.15	***
North	0.06	0.19	***
Abroad	0.01	0.01	
Field of study			
Education	0.07	0.21	***
Arts	0.07	0.01	***
Humanities	0.13	0.14	
Business	0.17	0.11	***
Social sciences	0.07	0.12	***
Law	0.06	0.03	***
Natural sciences	0.08	0.10	*
Technology	0.24	0.18	***
Agriculture and forestry	0.05	0.01	***
Medicine and health sciences	0.07	0.07	
N	890	1768	

Notes: A significant difference in group means is indicated by * ($p < .1$), ** ($p < .05$) or *** ($p < .01$). [a] For calculating the average matriculation grades, numeric values for the grades were assigned as follows: A = 1; B = 2; C = 3; M = 4; E or L = 5. [b] The months worked are summed across the four years preceding the graduation year. [c] The broad high school regions are defined according to the former administrative provinces of Finland (lääni).

market outcomes indeed differ considerably across the metropolitan and non-metropolitan regions: the average metro area resident earned significantly more, acquired more work experience and more frequently worked in a private sector job than an average resident in other regions. Clearly, the metro area labour market also offers a more diverse selection of jobs: whereas a majority (59%) of

the last four subsequent years instead of the full six-year period. The residents-in-metro-area and residents-in-other-regions subsamples are also used subsequently in this paper when applying Heckman's and Wooldridge's sample selection approaches.

residents in other regions worked either in the education or manufacturing sectors, the work sector distribution is more dispersed among residents in the metro area. Another interesting finding from Table 3 is that no large or significant metro area university premium exists in average earnings or months worked among graduates who resided in the same labour market area (Helsinki) after graduation. This finding suggests that the limited interregional mobility may indeed provide an explanation for the variation in earnings by university location.

TABLE 3 Labour market outcomes and the choice of post-university region: means of selected variables for residents in metro area and residents in other regions

	Residents in metro area	Residents in other regions	
Average annual earnings during the four-year period (€)	38,403	30,341	***
For metro area graduates	38,529	32,276	***
For other graduates	38,213	30,121	***
Months worked during the four-year period	42.2	36.5	***
For metro area graduates	41.7	40.7	
For other graduates	42.8	36.0	***
Entrepreneur	0.03	0.03	
Public sector job	0.30	0.50	***
Job industry			
Agriculture, forestry, game and fishing	0.01	0.01	**
Manufacturing and mining	0.15	0.17	
Electricity, gas and water supply	0.01	0.00	**
Construction	0.01	0.01	
Wholesale, retail trade, repair, accomm. and food services	0.10	0.06	***
Transportation, storage and telecommunications	0.04	0.01	***
Financing	0.05	0.01	***
Real estate, professional and technical services	0.20	0.09	***
Public admin., defence and extraterritorial organizations	0.09	0.07	**
Education and scientific research	0.21	0.42	***
Health and social work	0.05	0.09	***
Other social and personal services	0.05	0.02	***
No classification available	0.04	0.04	
N (total)	1153	1324	
N (metro area graduates)	690	135	
N (other graduates)	463	1189	

Notes: A significant difference in group means is indicated by * (p < .1), ** (p < .05) or *** (p < .01). The residents-in-metro-area sample includes graduates who resided in the Helsinki metropolitan area during the entire four-year period between the third and sixth year after graduation, whereas residents in other regions are those who resided elsewhere for the entire period. The individual's work sector is measured at the sixth year after graduation.

4 Methodology

In the estimation of the average early-career earnings premium for graduating from a metro area university, the following linear earnings equation is employed:

$$y_i = \alpha + \beta d_i + \gamma' x_i + \delta' u_i + \varepsilon_i, \qquad (1)$$

where y_i is the natural logarithm of earnings from the six-year period after graduation for individual i; α is an intercept; d_i is an indicator for having graduated from a metro area university instead of a university located outside the metro area; x_i is a vector of i's pre-university characteristics; u_i is a vector of i's university degree characteristics; and ε_i is the error term.[73] The baseline results are obtained by estimating model (1) by *ordinary least squares* (OLS), which relies on the conditional independence assumption (CIA): conditional on the observables included in the model, graduating from a metro area university is assumed to be independent of the counterfactual outcomes, that is, the potential earnings in the cases of $d_i = 1$ and $d_i = 0$. To enhance the credibility of this assumption, a rich set of control variables is included in the model. The non-random sorting of individuals across metropolitan and non-metropolitan universities is controlled for by the following pre-university characteristics included in vector x_i: gender, age, first language, mother's and father's education and socio-economic status, matriculation examination grades in math and first language, high school region and high school graduation year.[74] Of these control variables, the grades and high school region may be particularly important; these variables can be thought to impose the primary restrictions for i's university choice and capture a significant part of the individual heterogeneity related, e.g., to ability, motivation and the quality of pre-university education across university locations. The high-school-region dummy variables also have a special role in controlling for a potential confounding factor involved in the association between university location and subsequent location: regardless of the university city selected, i could be inclined to reside near his or her pre-university region after graduation, for example, because of family ties and networks located there, which could also have implications for his or her future earnings. As the study sample includes individuals graduating at different points of time and with different types of degrees, model (1) is further augmented with vector u_i that includes indicators for i's graduation year, field of study and degree level (bachelor's or master's degree). These variables control for the potential earnings heterogeneity arising

---

73 To assess heterogeneity in the metro area university premium across field of study, a version of the earnings equation including interactions with d_i and i's field of study is also employed during the analysis.

74 See Appendix A for detailed descriptions of the variables used in the analysis.

from time-variant factors (e.g., the business cycle and inflation) and differences in the contents of studies across universities.[75]

The validity of the conditional independence assumption involved in OLS is threatened by the fact that individuals may sort into universities partly based on unobserved earnings determinants such as ambition or motivation. Therefore, as a robustness check, equation (1) is also estimated by *two-stage least squares* (2SLS), that is, by employing the following reduced-form equation for the selection into a metro area university:

$$d_i = \mu + \pi' x_i + \rho' u_i + \tau' z_i + \omega_i, \qquad (2)$$

where vector z_i includes one or more instrumental variables (IV), i.e., variables that significantly affect the choice of metro area university but are excludable from the earnings equation; and ω_i is the error term. Apart from the linear functional form, identification in the 2SLS method relies heavily on whether the chosen instruments are relevant and valid. The current IV strategy is based on the hypothesis that temporal changes in the supply of student places in metro area universities relative to the total supply have created exogenous variation in students' locational choices. Therefore, a variable describing the Helsinki metro area's proportion of first-year students in i's field of study at the high school graduation year is included in equation (2) to generate an exclusion restriction for the earnings equation. When implementing the IV strategy, the supply of metro area student places is interacted with an indicator for whether i's high school was located in the Southern Finland because, in this manner, the explanatory power of the instrument increases. A detailed discussion of the conditions for the validity and relevance of the IV strategy is included in Appendix B.

In addition to estimating the total average metro area university premium, the effect conditional on the choice of post-university region is examined. Two alternative approaches are used in the attempt to extract the indirect post-university region effect from the total effect: a 'linear conditioning approach' and a 'sample selection approach'. In the first approach, the baseline model is simply augmented with vector r_i including 19 dummy variables for the region-of-residence six years after graduation and 6 dummy variables for the number of years resided in the Helsinki metropolitan area during the six-year period:

$$y_i = \alpha + \beta d_i + \gamma' x_i + \delta' u_i + \varphi' r_i + \varepsilon_i. \qquad (3)$$

75 An important difference exists between variables in vectors x_i and u_i. The first describe students' characteristics prior to attending university and are thus valid controls for university choice. However, the latter variables are determined along with or after university choice and are therefore potential mediating factors or 'bad controls' that may induce biases in the estimated university location effects (see, e.g., Angrist and Pischke, 2009, ch. 3.2.3.). For the sake of caution, OLS estimates conditioning only on pre-university variables are therefore also reported in Section 5.1.

The approach of simply conditioning on region-of-residence has been widely used in studies examining returns to educational choices (e.g., Card and Krueger, 1992; Lindahl and Regnér, 2005; Brunello and Cappellari, 2008), although this approach has been criticised for its strong underlying assumptions (see Heckman et al., 1996; Dahl, 2002; McHenry, 2011). Namely, the linear model (3), when estimated by conventional methods, assumes that the earnings premium for graduating from a metro area university does not vary across regions and that individuals residing in a particular region after graduation have been randomly sorted from metropolitan and non-metropolitan universities. Obviously, these assumptions are strong. In particular, individuals may choose to migrate from their region-of-graduation to another region partly on the basis of their expected gains from doing so, and thus an average graduate from a particular university city may be essentially different in different regions.[76]

The use of the 'sample selection approach' is based on the notion by Heckman et al. (1996) that the problems involved in the linear conditioning approach may be circumvented by explicitly allowing for self-selection in migration and interaction effects between region-of-school and region-of-work in the econometric model.[77] A straightforward strategy for implementing this approach would be to include a vector of interaction terms $d_i \times r_i$ in model (3) and to estimate this model by 2SLS, treating the choices of both university and post-university region as endogenous. However, this strategy unlikely yields robust estimates, mainly for two reasons: 1) in most cases, the coefficients for $d_i \times r_i$ are estimated using very few observations, and 2) one is unlikely to find legitimate instruments for all of the 51 endogenous variables. Therefore, instead of attempting to identify all of the region-of-school-region-of-work interaction effects, we focus on one particular effect: the earnings premium for graduating from a metro area university for those residing in the Helsinki metro area after graduation. Given that a large majority of metro area graduates (84% according to Table 3) and a sizeable proportion of other graduates (28%) end up residing in Helsinki, a sufficient number of observations from both the treatment and control groups is available for the estimation of the 'metro-graduate-metro-resident' interaction effect.[78] To estimate this parameter, the sample is – similar to Section 3 – divided into 'residents in metro area' and 'residents in other

---

76 In the literature, this selection problem is also known as the 'bad control' problem: as post-university region is a direct outcome of the choice of university location, including these two variables in the same regression may result in selection bias even if individuals were randomly assigned across university locations (Angrist and Pischke, 2009).

77 Previously, Dahl (2002) and McHenry (2011) have applied a correction for selective migration when estimating returns to education for different regional labour markets in the U.S.

78 Although the data would allow splitting the sample into even smaller labour-market-area-specific subsamples, these samples would inevitably be too small – especially in terms of treated observations – to be analysed separately. For example, although Tampere and Turku are the second and third largest urban regions in Finland, even the combined sample of residents in these regions includes only 21 graduates from Helsinki.

regions' on the basis of each individual's region-of-residence during the four-year period that falls between the third and sixth year after graduation from a university.[79] Given that the two groups have been formed as a result of individuals' locational choices, the earnings equation for individual i belonging to group j ('metro' or 'other') can be written as follows:

$$y_{ij} = \alpha_j + \beta_j d_i + \gamma_j' x_i + \delta_j' u_i + E\left[\varepsilon_{ij} | d_i, x_i, u_i, choose\ j\right] + \eta_{ij}, \qquad (4)$$

where η_{ij} is an independent error term, and $E\left[\varepsilon_{ij} | d_i, x_i, u_i, choose\ j\right]$ is an error component arising from fact that i's four-year earnings in a specific region are observed only if i actually resided in that region for the entire period.

In model (4), we are particularly interested in the population parameter β_{metro}, referring to the average metro area university premium among a group of individuals who were randomly selected as residents in the metro area after graduation. As sorting into group j may occur on the basis of unobservables, expectation $E\left[\varepsilon_{ij} | d_i, x_i, u_i, choose\ j\right]$ is non-zero in general, and hence the OLS and 2SLS estimates of parameter β_{metro} are potentially biased.[80] Here, the parametric approach of Heckman (1979) is applied to solve the selection problem: i is assumed to reside in the metro area for the entire four-year period only if his or her utility from this decision v_i^* is positive. This utility is assumed to be given by the following linear model:

$$v_i^* = \mu + \theta d_i + \pi' x_i + \rho' u_i + \tau' z_i + \varphi_i, \qquad (5)$$

where z_i denotes a vector of variables that are excludable from the earnings equation, and φ_i is the unobserved component of the utility. By further assuming that error terms ε_{ij} and φ_i are jointly normally distributed, the earnings equation for group j takes the following form:

$$y_{ij} = \alpha_j + \beta_j d_i + \gamma_j' x_i + \delta_j' u_i + \varphi_j \lambda_i + \eta_{ij}, \qquad (6)$$

where $\varphi_j \lambda_i \equiv E\left[\varepsilon_{ij} | d_i, x_i, u_i, choose\ j\right]$, and λ_i is the standard inverse Mills ratio obtained from the linear prediction of the selection model. Two slightly

79 'Movers', i.e., those that changed their place of residence between the metro area and other regions during the four-year period are, for simplicity, ignored in the sample selection approach. This group comprises only 178 observations, and thus the selection problem arising from ignoring this group is unlikely to be serious.

80 In general, the 2SLS estimate of β_{metro} represents the effect of graduating from a metro area university among graduates who actually resided in Helsinki after graduation, that is, a subsample parameter. The population parameter β_{metro} is only identified by 2SLS in two special cases: 1) if selection is a deterministic function of the model variables or 2) if selection is independent of the model variables (see Wooldridge, 2002, ch. 17.2). Sorting into universities and sorting into regional labour markets after graduation are essentially two different processes: the first concerns earning admission to a preferred school, whereas the latter concerns finding an optimal job and region-of-residence. As one may suspect that these processes are affected by different unobserved student characteristics, one must in general apply a simultaneous correction for the two endogeneity problems to obtain unbiased results.

different methods are employed to estimate model (6): 1) Heckman's two-step approach (Heckman, 1979) that corrects for the endogeneity of post-university region but treats university location choice as exogenous and 2) the approach discussed by Wooldridge (2002) that combines Heckman's approach and 2SLS; thus, this method allows for the endogeneity of both university location and post-university region.[81] In practise, the selectivity-correction methods require that, for each endogenous variable, there is at least one exclusion restriction, i.e., a variable that enters the reduced-form selection models but does not appear in the earnings equation (see Wooldridge, 2002, ch. 17.4). The current analysis uses two alternative variables to construct the exclusion restrictions: 1) the *supply of metro area student places* (interacted, as in the 2SLS estimation, with an indicator for high school in the Southern Finland); 2) *parent's residence near Helsinki*, i.e., an indicator for whether one of i's parents resided near Helsinki during the four-year period. These exclusion restrictions are further discussed in Appendix B.

5 Results

5.1 Estimates of the average metro area university premium

The analysis begins with the estimation of simple OLS earnings equations in which the identification of the average early-career earnings premium for graduating from a metro area university relies on the conditional independence assumption; these results are presented in the first row of Table 4. As shown in the leftmost column, without including any control variables in the earnings equation, the estimated average metro area university premium is significant and implies that, on average, metro area graduates earn 9.5% more than graduates from other cities.[82] To control for the non-random sorting of students across universities, a set of students' pre-university characteristics is included in the second specification (column 2). As a result, the OLS estimate is reduced to approximately zero (the implied earnings premium being .6%) and becomes insignificant. Thus, the results suggest that the positive earnings gap between

81 Heckman's version of the sample selection model is estimated using the standard two-step procedure, whereas Wooldridge's version is estimated in three steps: 1) the selection equation (5) is estimated by probit (without including the endogenous variable d_i in the model) to obtain $\hat{\lambda}_i$; 2) the linear regression model for graduation from a metro area university is estimated for the subsample in question (including $\hat{\lambda}_i$ as an explanatory variable) to obtain $\hat{d}_i$; 3) the earnings equation (5) is estimated with the replacement of d_i and λ_i with their predicted values. STATA's two-step robust standard errors are used for inference in the case of the Heckman estimates, whereas the standard errors for the Wooldridge estimates are obtained by conducting a 1000-round bootstrap. For a thorough discussion of the selectivity-correction methods, see Wooldridge (2002, ch. 17.4).

82 The percentage effect on earnings is obtained by subtracting one from the antilog of the regression coefficient. In reporting the results, estimates with a p-value of less than .10 are referred to as significant. More accurate significance levels (.10, .05 and .01) are reported in the tables.

metro area graduates and other graduates is entirely explained by differences in students' pre-university characteristics, whereas the choice of university location itself does not matter. From the third column, we see that the inclusion of university degree characteristics – university graduation year, field of study and degree level – as additional controls only has a marginal effect on the OLS estimate; the implied effect remains approximately zero (–0.4%). In the fourth OLS specification (column 4), the first attempt is made to extract the effect of post-university residential location from that of university location. After indicators for the region-of-residence six years after graduation and the number of years resided in the metro area are included, the average metro area university premium is negative and significant, implying an effect of –4.8%. Thus, the OLS evidence suggests that as graduating from a metro area university increases the probability of residing in more lucrative regions after graduation, the direct effect of graduating from a metro area university may even be negative.[83]

The robustness of the OLS findings to relaxing the conditional independence assumption is examined in the second row of Table 4, showing two-stage least squares (2SLS) estimates that were obtained using the supply of metro area student places (interacted with attending high school in the south) as an instrument for graduating from a metro area university. Regardless of whether post-university region is controlled for in the model, the IV estimation yields a clearly larger negative estimate for the average metro area university premium than OLS (–.209 and –.267 with and without post-university regional controls), which is in line with an upward ability bias in the OLS estimate. Alternatively, the difference between the OLS and IV estimates could arise because these estimates approximate different parameters: if the 'treatment effect' of graduating from a metro area university is heterogeneous across individuals, the OLS estimate could still relatively accurately approximate the average treatment effect (ATE), whereas the IV estimate could reflect a local average treatment effect (LATE), that is, the average effect of university location choice for a group of individuals who actually changed their university choices as a result of changes in supply. More specifically, because the supply variable is interacted with 'high school in south', the IV estimate could reflect a weighted average of the LATE's for the two high-school-location groups (see

---

83 In addition to the OLS specifications reported in Table 4, additional specifications were estimated to study the sensitivity of the OLS results to the inclusion of different sets of pre-university characteristics. When only including controls for the basic demographics (age, gender and first language) and the timing of university choice (high school graduation year), the estimated average metro area university premium is 7.6%. Thus, these basic controls explain only a small portion of the positive metro area university premium. Furthermore, when included in the model along with the basic controls, the remaining control variables have the following effects: 1) controlling for parents' education and socioeconomic status decreases the estimate to 5.6%; 2) controlling for matriculation examination grades decreases the estimate to 5.2%; 3) controlling for high school region decreases the estimate to 3.7%. Thus, based on these findings, differences in pre-university location seem to be the most important single source behind the earnings differences between metro area graduates and other graduates.

Angrist and Pischke, 2009, p. 173–175). However, as the IV estimates are neither significant nor significantly different from the OLS estimates due to their sizeable standard errors, no far-reaching conclusions can be made based on these results. Given that the IV point estimates are negative, the IV estimation at least does not provide evidence against the earlier conclusion that no positive average metro area university premium exists.[84]

TABLE 4 Effect of graduating from a metro area university on logarithmic six-year earnings: ordinary least squares (OLS) and two-stage least squares (2SLS) estimates

OLS estimate	.091	***	.006		-.004		-.049	*
	(.022)		(.025)		(.024)		(.026)	
2SLS estimate [a]					-.209		-.267	
					(.276)		(.316)	
First-stage F-statistic for the instruments					9.80	***	8.45	***
Earnings equation controlling for								
Pre-university individual characteristics			√		√		√	
University degree characteristics					√		√	
Post-university region							√	
N	2658		2658		2658		2658	

Notes: Standard errors are in parentheses. A significant estimate/test statistic is indicated by * (p < .1), ** (p < .05) or *** (p < .01). The pre-university individual characteristics include age, age squared, a woman indicator, a Swedish speaker indicator, mother's and father's education and socioeconomic status, high school region, high school graduation year and matriculation grades in math and first language. The university degree characteristics include university graduation year, a bachelor's degree indicator and field of study. The controls for post-university region include the region-of-residence six years after graduation and the number of years resided in metro area during the six-year period. [a] In the 2SLS estimation, the supply of metro area students places (interacted with high school in south) is used as an exclusion restriction for the earnings equation.

The previous estimates have suggested that the average metro area university premium is negative for individuals who resided in the same region after

---

84 Given that the heterogeneity of treatment effects might impose problems for the estimation of the average metro area university premium using linear approaches, the estimations were also conducted using propensity score matching (PSM), which relaxes the linearity assumption. In most cases, the PSM estimates of the average treatment effect – for both the treated and untreated – were very similar to the OLS estimates, suggesting that the OLS evidence is rather robust to heterogeneity. The only exception occurred when controlling for post-university region in the estimation of the propensity score; in this case, the average treatment effect for the untreated differed from the OLS estimate, being slightly positive.

graduation. To assess whether this result is driven by the selective nature of graduates' migration decisions, an alternative sample selection approach is applied: the analysis is conducted using 'residents in metro area' and 'residents in other regions', i.e., individuals who stayed either inside or outside the Helsinki metropolitan area between the third and sixth year after graduation. In this analysis, we are particularly interested in the selectivity-corrected estimate of the average metro area university premium for residents in the metro area, as this estimate can be interpreted to reflect the average effect for a randomly selected group of individuals who faced the same local labour market conditions after graduation. However, estimates for residents in other regions are also reported, in addition to the non-selectivity-corrected OLS and 2SLS estimates, to provide an overall comparison of results across different samples and methods.

The results for residents in the metro area are presented in the left column of Table 5. The baseline OLS estimate for this subsample is significantly negative and indicates that graduation from a metro area university decreases earnings by an average of 7.3% conditional on residing in the metro area afterwards. Compared to the OLS estimate, the two Heckman estimates (obtained using two alternative exclusion restrictions) are clearly smaller, –.015 and –.035, and not significantly different from zero. Thus, the Heckman results appear to indicate a particular pattern of self-selection: of graduates from non-metropolitan universities, those with relatively high average expected earnings self-select into Helsinki, and therefore OLS overestimates the negative premium for graduating from a metro area university among metro area residents. Compared to the methods relying on the conditional independence assumption (OLS and Heckman), the IV estimation yields larger negative estimates. The difference between the IV and OLS estimates is, however, notably smaller than in the previous results: without correcting for sample selection, the IV estimate indicates that graduation from a metro area university decreases earnings by 11.1% on average, whereas the two selectivity-corrected Wooldridge estimates indicate effects of –12.1% and –7.1%. The standard errors for the IV estimates are once again substantially large, and therefore none of the estimates is significantly different from zero. Thus, although the IV results appear to indicate a modest bias in the OLS and Heckman estimates, this conclusion cannot be confirmed because of inefficiency involved in the IV estimation. Furthermore, the possibility of the IV estimates reflecting local average treatment effects must be bear in mind.

The results for the residents-in-other-regions subsample (Table 5, right column) differ to some extent from those for the previous subsample: the OLS estimate is close to zero and insignificant, whereas the Heckman-corrected estimates are even considerably positive, indicating that graduation from a metro area university increases earnings by 15.8% or 22.9% on average, depending on the exclusion restriction employed. Thus, the Heckman findings suggest that metro area graduates residing outside Helsinki after graduation

TABLE 5 Effect of graduating from a metro area university on logarithmic four-year earnings for graduates who resided either inside or outside the Helsinki metropolitan area

	Residents in metro area		Residents in other regions	
Exogenous metro area residence assumed				
OLS estimate	-.075	**	.016	
	(.038)		(.051)	
2SLS estimate [a]	-.118		-.343	
	(.289)		(.561)	
First-stage F-stat. for the instruments	8.26	***	4.84	***
Endogenous metro area residence allowed for				
Heckman estimate I [a]	-.015		.147	
	(.066)		(.130)	
Heckman estimate II [b]	-.035		.206	*
	(.063)		(.120)	
Wooldridge estimate I [a]	-.129		-.158	
	(.324)		(1.001)	
First-stage F-stat. for the instruments	8.64	***	4.42	**
Wooldridge estimate II [a b]	-.080		-.128	
	(.303)		(.662)	
First-stage F-stat. for the instruments	6.11	***	3.35	**
N	1153		1324	

Notes: Standard errors are in parentheses. A significant estimate/test statistic is indicated by * ($p < .1$), ** ($p < .05$) or *** ($p < .01$). The estimation sample includes graduates who resided either inside or outside the Helsinki metropolitan area for the entire four-year period between the third and sixth year after graduation. The basic conditioning covariates in each model are the same as in the third column of Table 4. The exclusion restrictions used in the estimation are indicated by [a] (supply of metro area students places interacted with high school in south) and [b] (parent's residence near Helsinki).

benefit from their university location choice.[85] After weighting the Heckman estimates for residents in the metro area and residents in other regions by the number of metro area graduates in these subsamples – 690 and 135, respectively – the implied average metro area university premium is 1.1% or 0.4%, depending on the exclusion restriction employed. Thus, it appears that the

85 However, when interpreting the results for residents in other regions, one must bear in mind that the individuals in this group resided in different parts of Finland. Thus, the estimated positive premium may partly reflect differences in locational choices between metro area graduates and other graduates residing outside the Helsinki metropolitan area. Some indication of this bias is obtained, as the Heckman estimates for residents in other regions are slightly smaller (.118 and .185) after augmenting the earnings equation with dummy variables for the region-of-residence six years after graduation.

previous OLS estimates – indicating a negative average premium – are biased because of graduates' selective migration behaviour, whereas the actual average premium is very close to zero. The IV results for residents in other regions are to be treated with caution: although these estimates are even more negative than those for residents in the metro area, the IV strategy currently employed is obviously too weak to yield robust IV estimates for this subgroup, as indicated by the sizeable standard errors and small first-stage F-statistics.

Provided that the total average premium for graduating from a metro area university is – as suggested by the Heckman results – similar to the average premium conditional on post-university region, one may ask: How is this possible? A likely explanation for this finding is that by controlling for high school region in the estimation, the primary mechanism behind the choice of post-university region is already effectively captured: many of the university students who do not attend university in their region-of-origin might consider their university city to only be a temporary place of residence, whereas the former region could matter more in the subsequent locational choice. Thus, the earnings effect of university location choice arising through post-university region could also be insignificant. This hypothesis is supported by the previous finding in Section 3 that 30% of the graduates who migrated out of their high school region to complete a university degree returned 'home'.

5.2 Heterogeneity across fields

After studying the average effect of graduating from a metro area university, we focus on examining a potential source of heterogeneity in this effect: that arising from the field-of-study choice. This supplemental analysis may be motivated by arguing that students who reside in the same city but study different fields may face very different conditions during their studies, e.g., in terms of job opportunities and university quality. Thus, the university location effects could be heterogeneous across fields. For this analysis, interaction terms between field of study and the metro area university indicator are included in the earnings equations estimated for the full sample and the residents-in-metro-area subsample. As the samples for individual fields are relatively small, the IV estimator is unlikely to yield accurate estimates; therefore, the estimations are only carried out by OLS for the full sample and by Heckman's method for residents in the metro area.

The results reported in Table 6 suggest that the effect of graduating from a metro area university varies across fields.[86] For the full sample, half of the point estimates are positive, whereas the other half are negative. Without conditioning on post-university region (column 1), a significantly positive effect is found in business (13.0%) and social sciences (15.0%), whereas a significantly negative effect is found only in humanities (–11.8%). Compared with the

86 The estimates for the two smallest fields, arts and agriculture and forestry, are left unreported, as these estimates are based on very few observations and are therefore unlikely to be reliable.

estimates of the total metro area university premium, the full-sample estimates conditioning on post-university region (column 2) are again lower or more negative. Of these estimates, the positive one for business (10.8%) and all of the negative ones, i.e., those for education (–13.0%), humanities (–17.2%), natural sciences (–14.9%) and technology (–8.3%), are significantly different from zero. Many of the differences between the field-of-study-specific estimates are also significant; in particular, all of the positive estimates are significantly different from the negative estimate for humanities, whereas all of the negative estimates are significantly different from the positive estimates for business and social sciences. The estimates obtained for the residents-in-metro-area subsample are qualitatively in line with the full-sample estimates; only the point estimate for education has a different sign in the two sets of results. However, due to the small sample size, the standard errors for the subsample results are considerably large. As a result, only the sizeable negative estimate for humanities is significant; this estimate implies an effect of –29.4% for graduating from a metro area university for humanities students. The estimate for humanities is also significantly different from the remaining field-of-study-specific estimates, whereas, apart from this, a significant difference is found only between the estimates for natural sciences and medicine/health sciences.

Given that significant differences were found in the metro area university premium across fields, the mechanisms behind this heterogeneity may be briefly discussed. One potential explanation is offered in Figure 1: many of the gaps in the teacher-student ratio between Helsinki and the other cities are surprisingly well in line with the estimated field-of-study-specific earnings premiums. In particular, both the average gap in the teacher-student ratio and the earnings premium receive their largest negative values in humanities and natural sciences, leading one to suspect that the lack of educational resources in Helsinki could be a source of the relatively low early-career outcomes of metro area graduates in these two fields. Moreover, the relative abundance of resources in medicine and health sciences in Helsinki could offer an explanation for the positive (although insignificant) estimates obtained for this field.

An alternative explanation for the heterogeneity could be that a student's ability to benefit from the metro area labour market during studies and afterwards varies across fields. Students in business, social sciences and law – for which positive estimates were observed in Table 6 – are particularly likely to benefit from a location in Helsinki, as these students often find themselves employed in state government and large private enterprises that are highly concentrated in the capital region. Students in education, humanities and natural sciences are again the most likely to work in the education sector that is scattered across the country, and thus returns from locational choices could be less likely to arise for these students. Consequently, the effect of the limited interregional mobility of graduates could also differ across fields: compared to students in education, humanities and natural sciences, those in business, social sciences and law could benefit more from a high likelihood of being in the metro area labour market after graduation.

TABLE 6 Effect of graduating from a metro area university on logarithmic earnings by field of study

	Full-sample OLS estimate			Full-sample OLS estimate			Heckman estimate for residents in metro area		
Education	-.074		b, s	-.139	**	b, s	.011		h
	(.067)			(.068)			(.135)		
Humanities	-.126	**	b, l, m, s	-.189	***	b, l, m, s	-.349	***	b, e, l, m, n, s, t
	(.055)			(.056)			(.119)		
Business	.122	**	e, h, n, t	.102	*	e, h, n, t	.035		h
	(.054)			(.054)			(.073)		
Social sciences	.139	**	e, h, n, t	.076		e, h, n, t	.001		h
	(.070)			(.071)			(.119)		
Law	.078		h	.037		h, n	.103		h
	(.091)			(.091)			(.131)		
Natural sciences	-.098		b, s	-.161	**	b, l, m, s	-.121		h, m
	(.067)			(.069)			(.144)		
Technology	-.040		b, s	-.087	*	b, s	-.084		h
	(.044)			(.045)			(.102)		
Medicine and health sciences	.048		h	.011		h, n	.107		h, n
	(.073)			(.074)			(.165)		
Conditional on post-university region	No			Yes					
N	2658			2658			1153		

Notes: Standard errors are in parentheses. A significant field-of-study-specific estimate is indicated by * ($p < .1$), ** ($p < .05$) or *** ($p < .01$). A significant difference ($p < .1$) to the estimate of another field is indicated by [b] (business), [e] (education), [h] (humanities), [l] (law), [m] (medicine and health sciences), [n] (natural sciences), [s] (social sciences) or [t] (technology). Six-year earnings are used as the dependent variable for the full sample, whereas four-year earnings are used for the sample of metro area residents. The basic conditioning covariates in each model are the same as in the third column of Table 4. In implementing Heckman's selectivity correction, the supply of metro area student places (interacted with high school in the south) is used as an exclusion restriction for the earnings equation.

The heterogeneity in educational resources and job opportunities across regions and fields could explain a significant part of the observed field-of-study differences in the earnings effect of graduating from Helsinki. However, one might suspect that this heterogeneity still does not sufficiently explain why considerably large negative estimates were obtained for some fields. The data shows that graduates from Helsinki are relatively immobile during early careers: 77% of these graduates still resided in their region-of-graduation six years after graduation, whereas the corresponding figure for graduates from the other cities is 35%. Therefore, an additional mechanism behind the negative estimates could be that studying in a metro area university induces 'excess immobility' during the early career. Namely, if metro area graduates, having resided in an economically active, urban region for several years, were more inclined than other graduates to accept a low-paid job in return for not having to migrate away from their university city, the total return from university location choice could be negative for some of the metro area graduates, in particular for those from less lucrative fields. Theoretically, this type of behaviour could occur if metro area graduates faced a relatively high 'option value of waiting' (see Burda, 1995); that is, for metro area graduates, the costs involved in migrating out could be relatively high – for example, because of the loss of different urban amenities and acquired networks – whereas the expected long-term economic gain from relocation could be relatively low; then again, the opposite could apply for graduates from smaller urban regions.[87] This immobility, in combination with the other mechanisms discussed above, could provide a sensible explanation for the field-of-study-specific findings.[88]

6 Summary and concluding remarks

In this paper, economic returns from university choice were studied from a perspective rarely adopted in previous studies: university location. A novelty of this paper is its comprehensive discussion of both the mechanisms behind 'university location effects' and econometric problems involved in the estimation of these effects. In the empirical analysis, the existence of university location effects was studied in Finland by focusing on the early-career earnings premium for graduating from a 'metro area university', i.e., a university located in the Helsinki metropolitan area. Given that Helsinki offers a more

---

87 A traditional explanation for immobility states that in addition to direct and informational costs, psychic costs involved in human migration lower individuals' willingness to move (Sjaastad, 1962; Schwartz, 1973). Furthermore, Burda (1995) has theorised that postponing migration to another location may have a positive option value even if immediate returns to migration exist, as migration involves sunk costs and uncertainty. Studies examining regional wage differences have also suggested that individuals may be willing to accept a negative compensating wage differential or a lower return to education to be able to live close to certain location-specific amenities (Roback, 1982; Black et al., 2009). These theories yield a basis for the argument that studying in an economically active, high-amenity region may, at least temporarily, generate negative returns through immobility.

88 Similar to some earlier studies (e.g., Lindahl and Regnér, 2005; Brunello and Cappellari, 2008), a simple test was also conducted to evaluate the extent to which the estimated returns to university choice are explained by differences in educational resources by augmenting earnings equations with a variable describing the average teacher-student ratio in a student's university city and field at the time of study. The inclusion of this variable caused only negligible changes in the estimated field-of-study-specific earnings premiums, suggesting that the field-of-study heterogeneity is not related to differences in educational resources. Furthermore, the coefficient estimate for the teacher-student ratio was not significantly different from zero, and thus no evidence of a positive premium for 'teaching quality' was obtained.

advantageous environment for its students than the nine other university cities – for example, in terms of the quality of peers and professors and job opportunities – one may expect that a student from a metro area university earns more than a similar student from elsewhere after graduation. However, the obtained results do not in general support this hypothesis. Although graduates from Helsinki earn an average of 13.6% more than other graduates during the six-year period after graduation, this premium seems to be entirely explained by differences in students' pre-university characteristics and not by the choice of university location itself. The OLS estimates for the average metro area university premium were close to zero and insignificant, whereas the IV estimation provided considerably negative point estimates. However, the IV estimates were also more imprecise, which suggests – along with the findings indicating that the treatment effects are heterogeneous – that the IV results should be interpreted with caution. Thus, we may conclude that despite the locational heterogeneity of universities in Finland, no strong evidence is found that university choice is generally significant in terms of students' labour market success. This conclusion is comparable to some of the recent findings from Sweden (Eliasson, 2006; Holmlund, 2009) and Italy (Triventi and Trivellato, 2012) that have highly similar state-funded and geographically dispersed university systems.

The findings of this paper nonetheless suggest that certain subgroups may benefit from locational choices. According to the Heckman-corrected results obtained for 'residents in metro area' and 'residents in other regions', graduation from a metro area university has a sizeable positive average effect on earnings for the latter group but not for the former. Thus, the valuation of a graduate's university choice appears to depend on the regional labour market in which he or she is located after graduation. Furthermore, the results indicated heterogeneity in the university location effects across fields. For example, positive estimates for graduating from a metro area university were obtained for students from business and social sciences, whereas humanities students were found to benefit considerably from attending university outside Helsinki. According to the hypotheses discussed above, the high degree of immobility among metro area graduates during their early career in combination with certain field-of-study differences – such as those in educational resources and in a student's ability to benefit from the metro area labour market – could explain why the estimates were negative for some of the fields while positive for others.

Due to the close relationship between the choices of university location and post-university residential location, prior studies have used the approach of conditioning on post-university region in a linear model to be able to study earnings effects directly related to university choice. The problem involved in this approach arising from graduates' selective migration behaviour was discussed and corrected for in this paper, following the ideas presented in studies examining returns to education across regional labour markets (Heckman et al., 1996; Dahl, 2002; McHenry, 2011). The results suggest that one may indeed draw biased conclusions if estimates conditioning on post-university region are not adjusted for the endogeneity of locational choices: whereas the baseline OLS estimates indicated a significantly negative average premium for graduating from a metro area university conditional on post-university region, the Heckman-corrected estimates suggested that the average effect is close to zero, that is, approximately the same as without conditioning on the subsequent locational choice. Obviously, the bias in the OLS estimate arises because a large majority of metro area graduates stay in Helsinki after graduation, whereas only a self-selected group of other graduates migrate there. Thus, when conditioning on post-university region, we are essentially comparing an approximately average metro area graduate to a non-average other graduate. This example stresses the importance of being careful when conditioning on variables that are determined after the treatment of interest.

REFERENCES

Adams, D. 1997. Defining Educational Quality. Educational Planning 9(3), 3–18.

Angrist, J.D., Pischke, J.-S., 2009. Mostly harmless econometrics. Princeton University Press, New Jersey.

Behrman, J., Rosenzweig, M., Taubman, P., 1996. College choice and wages: Estimates using data on female twins. Review of Economics and Statistics 78 (4), 672−685.

Black, D., Kolesnikova, N., Taylor, L., 2009. Earnings functions when wages and prices vary by location. Journal of Labor Economics 27 (1), 21–47.

Black, D., Smith, J., 2004. How robust is the evidence on the effects of college quality? Evidence from Matching. Journal of Econometrics 121 (1−2), 99−124.

Black, D., Smith, J., 2006. Estimating the returns to college quality with multiple proxies for quality. Journal of Labor Economics 24 (3), 701–728.

Brewer, D., Eide, E., Ehrenberg, R., 1999. Does it pay to attend an elite private college? Cross-cohort evidence on the effects of college type on earnings. Journal of Human Resources 34 (1), 104–123.

Brunello, G., Cappellari, L., 2008. The labour market effects of Alma Mater: Evidence from Italy. Economics of Education Review 27(5), 564–574.

Burda, M. 1995. Migration and the option value of waiting. Economic and Social Review 27(1), 1–19.

Dahl, G.B., 2002. Mobility and the return to education: Testing a Roy model with multiple markets. Econometrica 70(6), 2367–2420.

Dale, S.B., Krueger, A., 2002. Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. Quarterly Journal of Economics 117 (4), 1491–1527.

Card, D., Krueger, A., 1992. Does school quality matter? Returns to education and the characteristics of public schools in the United States. Journal of Political Economy, 100 (1), 1–40.

Chevalier, A., Conlon, G., 2003. Does it pay to attend a prestigious university? CEP Discussion Paper No. 33. London School of Economics and Political Science.

Cooke, T.J., Boyle, P., 2011. The migration of high school graduates to college. Educational Evaluation and Policy Analysis, 33(2), 202–213.

Eliasson, K., 2006. The role of ability in estimating the returns to college choice: New Swedish evidence. Umeå Economic Studies No. 691, Umeå University.

Faggian, A., McCann, P., 2009. Universities, agglomerations and graduate human capital mobility. Tijdschrift voor Economische en Sociale Geografie 100(2), 210–223.

Glaeser, E.L., Kolko, J., Saiz, A., 2001. Consumer City. Journal of Economic Geography 1(1), 27–50.

Glaeser, E.L., Maré, D.C., 2001. Cities and Skills. Journal of Labour Economics 19(2), 316–342.

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.

Heckman, J., 1979. Sample selection bias as a specification error. Econometrica 47 (1), 153–162.

Heckman, J., Layne-Farrar, A., Todd, P., 1996. Human capital pricing equations with an application to estimating the effect of schooling quality on earnings. Review of Economics and Statistics 78 (4), 562–610.

Hershbein, B., 2011. Worker signals among new college graduates: The role of selectivity and GPA. Unpublished manuscript.

Hoekstra, M., 2009. The effect of attending the flagship state university on earnings: A discontinuity-based approach. Review of Economics and Statistics 91(4), 717–724.

Holmlund, L., 2009. The effect of college quality on earnings. Evidence from Sweden. Umeå Economic Studies No. 781, Umeå University.

Häkkinen, I., 2006. Working while enrolled in a university: Does it pay? Labour Economics 13 (2), 167–189.

Keskinen, E., Tiuraniemi, J., Liimola, A., 2008. University selection in Finland: how the decision is made. International Journal of Educational Management 22(7), 638–650.

Lang, K., Siniver, E., 2011. Why is an elite undergraduate education valuable? Evidence from Israel. NBER Working Paper No. 16730.

Lempinen, P., Tiilikainen, A., 2001. Opiskelijatutkimus 2000 (in Finnish). Publications of Otus 21/2001, retrieved from: http://www.otus.fi/Opiskelijatutkimus2000.htm.

Lindahl, L., Regnér, H., 2005. College choice and subsequent earnings: Results using Swedish siblings data. Scandinavian Journal of Economics 107 (3), 437–457.

Long, M.C., 2008. College quality and early adult outcomes. Economics of Education Review 27 (5), 588–602.

McHenry, P., 2011. The effect of school inputs on labor market returns that account for selective migration. Economics of Education Review 30(1), 39–54.

Ministry of Education, 2005. OECD thematic review of tertiary education: Country background report for Finland, Publications of the Ministry of Education, Finland 2005:38.

Ministry of Education, 2012. KOTA online database service. Available at https://kotaplus.csc.fi.

Monks, J., 2000. The returns to individual and college characteristics: Evidence from the National Longitudinal Survey of Youth. Economics of Education Review 19 (3), 279–289.

Professoriliitto, 2011. Wage report 2011 (in Finnish), retrieved from http://www.professoriliitto.fi/52/.

Roback, J., 1982. Wages, rents and the quality of life. Journal of Political Economy 90(6), 1257–1278.

Sacerdote, B., 2001. Peer effects with random assignment: Results for Dartmouth roommates. Quarterly Journal of Economics 116(2), 681–704.

Schwartz, A., 1973. Interpreting the effect of distance on migration. Journal of Political Economy 81(5), 1153–1169.

Sjaastad, L., 1962. The costs and returns of human migration. Journal of Political Economy 70(5), 80–93.

Statistics Finland, 2012a. ALTIKA database. Available at http://www.stat.fi/tup/altika.

Statistics Finland, 2012b. Official Statistics of Finland (OSF). Available at http://www.stat.fi/meta/svt.

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

Triventi, M., Trivellato, P., 2012. Does graduating from a private university make a difference? Evidence from Italy. European Journal of Education 47(2), 260–276.

Villalonga, J.R., 2011. Statification of universities and students' segregation. Unpublished manuscript, retrieved from http://dea.uib.es/digitalAssets/199/199335_w45.pdf.

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, Massachusetts.

Zimmerman, D.J., 2003. Peer effects in academic outcomes: evidence from a natural experiment. Review of Economics and Statistics 85(1), 9–23.

APPENDIX A List of variables used in the analysis

- *Six-year earnings* are constructed by summing the annual values of taxable income from the six-year period after the graduation year. This variable is used as the dependent variable for the full sample.
- *Four-year earnings* are constructed by summing the annual values of taxable income from the four-year period between the third and sixth year after the

graduation year. This variable is used as the dependent variable for the metro area sample.

- *Graduated from a metro area university* takes a value of 1 if the individual graduated from one of the universities located in the Helsinki metropolitan area and 0 otherwise.
- *Resident in metro area* takes a value of 1 if the individual resided in the Helsinki metropolitan area during the four consecutive years between the third and sixth year after the graduation year and 0 if the individual resided outside Helsinki for the entire period.
- *Field of study*. The field of the latest university degree. This variable includes 10 categories: education; arts; humanities; business; social sciences; law; natural sciences; technology; agriculture and forestry; medicine and health sciences.
- *Graduation year.* The year the latest university degree was completed. This variable includes seven categories (years from 1994 to 2000).
- *Age*. The individual's age six years after the graduation year. The square of this variable is also included in each model.
- *Woman* takes a value of 1 if a woman and 0 otherwise.
- *Swedish speaker* takes a value of 1 if the individual's mother tongue is Swedish and 0 otherwise.
- *Bachelor's degree* takes a value of 1 if the obtained degree is a bachelor's degree and a 0 otherwise.
- *Mother's/Father's education*. The latest observed educational qualification of mother/father (observations in 1970, 1980, 1990, and 2000). This variable includes 12 categories: primary school; high school diploma; secondary-level vocational qualification; lowest tertiary-level qualification; higher education degree (bachelor's or master's) in education; higher education degree in humanities (including arts); higher education degree in social sciences (including business and law); higher education degree in natural sciences or technology; higher education degree in medicine; other higher education degree; doctoral degree; no classification available.
- *Mother's/Father's socioeconomic status*. The latest observed socioeconomic class of mother/father (observations in 1970, 1980, or 1990). These variables include six categories: farmer; entrepreneur; high-ranking official; low-ranking official; worker; no classification available.
- *Matriculation grade in first language*. The highest grade obtained in the first language test of the Finnish matriculation examination. The variable includes five categories (starting from the lowest grade): approbatur (A); lubenter approbatur (B); cum laude approbatur (C); magna cum laude approbatur (M); eximia cum laude approbatur (E) or laudatur (L).
- *Matriculation grade in mathematics*. The highest grade obtained in the mathematics test of the Finnish matriculation examination. This variable includes 13 categories: B-level improbatur (I); B-level approbatur (A); B-level lubenter approbatur (B); B-level cum laude approbatur (C); B-level magna cum laude approbatur (M); B-level eximia cum laude approbatur (E) or

laudatur (L); A-level improbatur (I); A-level approbatur (A); A-level lubenter approbatur (B); A-level cum laude approbatur (C); A-level magna cum laude approbatur (M); A-level eximia cum laude approbatur (E) or laudatur (L); no math grade.

- *High school region*. The region in which the individual completed a high school degree. This variable includes 20 categories: Uusimaa; Eastern Uusimaa; Finland Proper or Aland Islands; Satakunta; Tavastia Proper; Pirkanmaa; Päijänne Tavastia; Kymenlaakso; South Karelia; Southern Savonia; Northern Savonia; North Karelia; Central Finland; Southern Ostrobothnia; Ostrobothnia; Central Ostrobothnia; Northern Ostrobothnia; Kainuu; Lapland; unknown.
- *High school graduation year* includes 12 categories: 1985 or earlier, one category for each year between 1986 and 1995, and 1996 or later.
- *Region of residence*. The region in which the individual resided six years after the graduation year. The categories for this variable are similar to those for high school region.
- *Number of years resided in the metro area* is obtained by summing the years in which the individual resided in the Helsinki metropolitan area from the six-year period after the graduation year.
- *Supply of metro area student places.* The Helsinki metro area's proportion of first-year students in the individual's field of study at the high school graduation year. The information on first-year students was obtained from the KOTA database (Ministry of Education, 2012).
- *Parent's residence near Helsinki* takes a value of 1 if one of the individual's parents resided in the Uusimaa or Eastern Uusimaa region during at least one of the years between the third and sixth year after the graduation year.

APPENDIX B Exclusion restrictions

The instrumental variables strategy used in this paper relies on the hypothesis that temporal changes in regional supplies of student places governed by the Ministry of Education have created exogenous variation in students' locational choices. In particular, changes in the Helsinki metro area's proportion of student places in certain fields of study may have affected the probability of sorting into the metro area universities for students applying to these fields. Figure A1 shows that noticeable changes in the metro area's proportion of first-year students occurred in several fields between 1985 and 1995.[89] Specifically, the metro area's proportions in two large fields, business and technology, decreased by 10.7 and 10.3 percentage points, respectively, making metro area student places in these fields relatively more scarce over time.

---

89 For 99% of the sample observations, the high school graduation year lies within this interval.

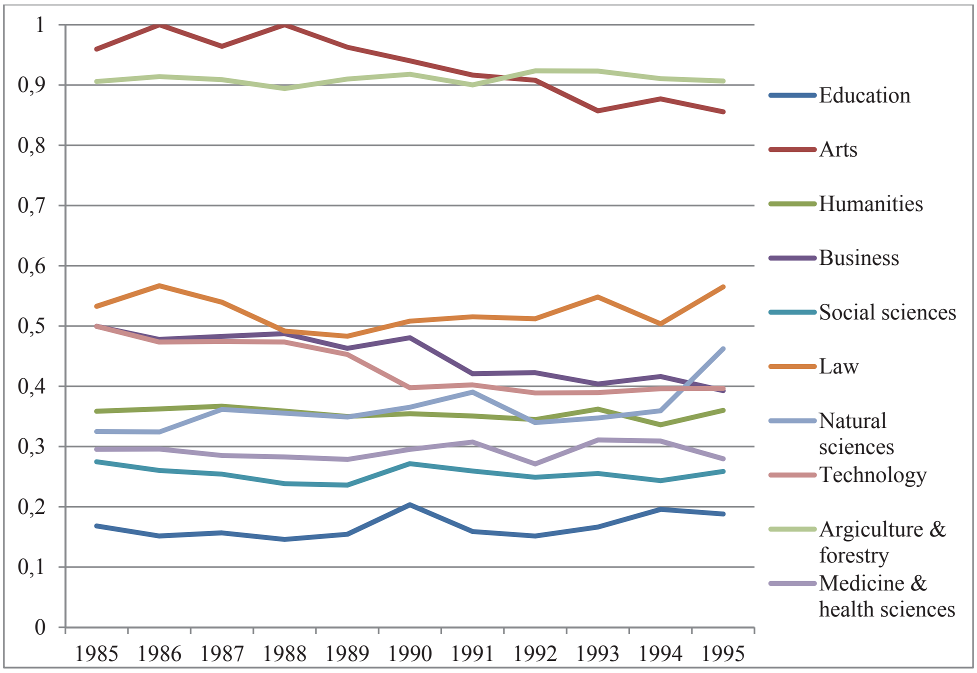


FIGURE A1 Helsinki metropolitan area's proportion of first-year students in ten fields of study. Source: Ministry of Education (2012).

To use these policy changes as an IV, the metro area's proportions depicted in Figure A1 are matched with the micro data based on the individuals' fields of study and high school graduation years, and the resulting variable *supply of metro area student places* is included in the reduced-form model of university choice (equation (3) in Section 4). Furthermore, when implementing the IV strategy, the supply of metro area student places is interacted with one's high school location. The reason for this choice is illustrated in Table A1: when included as a separate variable in the first-stage model (columns 1 and 2), the supply of metro area student places is insignificant – and thus does not form a sufficient exclusion restriction – but becomes significant after allowing its effect to vary across students' high school regions (columns 3–6). The pattern of heterogeneity suggested by the first-stage estimates is predictable: the estimate for the supply of metro area student places differs significantly between high school graduates coming from remote regions (Western, Eastern and Northern Finland) and those coming from proximate regions (Southern Finland), being positive for the first group and negative for the latter group. Thus, the preferred strategy is to use both the supply variable and its interaction with 'high school in south' as instruments. The first-stage F-statistic for this strategy is approximately 10, i.e., near the rule-of-thumb level often used for detecting whether an instrument is relevant (Angrist and Pischke, 2009).[90]

90 As the supply of metro area student places varies both by field of study and high school graduation year, one might suspect that the weakness of the IV strategy arises

Under what conditions is the employed IV strategy valid? Because the fixed effects of both field of study and high school graduation year are controlled for in the first-stage model, the identification of the 'supply effect' in this model arises through differences in the temporal changes of the supply across fields. Thus, the strategy is valid unless there were other simultaneous changes – such as those in the ability distributions of different fields – that correlate with both unobserved earnings determinants and the supply changes. Although one may not entirely dismiss the possibility of such distortions, at least two arguments for the validity may be presented. First, in the presence of changes in field-of-study-specific ability distributions, one would expect the IV estimates to be sensitive to the inclusion of additional ability controls in the model. However, after controlling for field of study, high school graduation year and high school region, the inclusion of other individual-level variables has almost no effect on the obtained IV estimates. Second, as the Sargan test of overidentifying restrictions is not significant after including all of the control variables in the model (Table A1, column 6), no strong evidence of correlation between the unobserved earnings determinants and the instruments is found.

Although the supply of metro area student places may in principle also be used to form an exclusion restriction in Heckman's and Wooldridge's sample selection approaches, selectivity-corrected results using another exclusion restriction – a dummy variable for whether one of the individual's parents resided near Helsinki during the four-year period – are also reported in Section 5. The reason for this robustness check is illustrated in Table A2: the estimate for the supply variable is positive but only weakly significant in the estimated reduced-form model for the choice of post-university region (equation (5) in Section 4). Thus, identification – particularly in Wooldridge's approach that has two endogenous variables – may in practice require another exclusion restriction. The obvious logic behind the second strategy is that as individuals may prefer residing near their parents, an individual with a parent located near Helsinki may have a higher probability of self-selecting into the region after graduation than a similar individual without that type of parent. As shown in Table A2, the indicator for a parent's residence near Helsinki receives a highly significant positive estimate and thus serves as a sufficiently strong exclusion restriction. However, as parents' locational choices are of a non-random nature and thus potentially correlated with individuals' unobserved characteristics, the validity of the exclusion restriction is not completely warranted, even after controlling for individuals' and their parents' pre-determined characteristics. For the sake of caution, the sample selection approaches are therefore also estimated without using the second exclusion restriction.

---

partly because of the lack of variation resulting from controlling for both high-school-graduation-year and field-of-study fixed effects. Therefore, excluding high school graduation year from the model could be beneficial, in particular, as a direct link between this variable and earnings is not particularly obvious. However, Table A1 shows that the IV strategy is only marginally weaker in the model controlling for high school graduation year. The inclusion of the high-school-graduation-year dummy variable merely has the effect of shifting the coefficients for the supply variable in a negative direction, which could arise from multicollinearity.

TABLE A1 Supply of metro area student places as an instrumental variable for graduation from a metro area university: first-stage OLS estimates and tests of relevance and validity

Supply of metro area student places	.125	-.005			.322	.193
	(.300)	(.322)			(.302)	(.324)
Interaction with high school location						
South			-.083	-.213	-.421 ***	-.417 ***
			(.303)	(.324)	(.094)	(.094)
West			.409	.273		
			(.308)	(.329)		
East			.360	.226		
			(.322)	(.341)		
North			.167	.038		
			(.318)	(.339)		
Abroad			.070	-.063		
			(.569)	(.579)		
Conditional on high school graduation year	No	Yes	No	Yes	No	Yes
Tests for the instruments						
Joint significance (F-stat.)	0.17	0.00	4.70 ***	4.55 ***	10.10 ***	9.80 ***
Overidentifying restrictions (Sargan χ^2-stat.)			8.62 *	6.82	4.11 **	2.28
N	2658	2658	2658	2658	2658	2658

Notes: Standard errors are in parentheses. A significant estimate/test statistic is indicated by * ($p < .1$), ** ($p < .05$) or *** ($p < .01$). The basic control variables in each model are age, age squared, a woman indicator, a Swedish speaker indicator, mother's and father's education and socioeconomic status, high school region, matriculation grades in math and first language, field of study, a bachelor's degree indicator and university graduation year.

TABLE A2 Exclusion restrictions in Heckman's and Wooldridge's sample selection approaches: probit estimates from the reduced-form models of metro area residence

Variable used as an exclusion restriction	Sample selection approach							
	Heckman I		Heckman II		Wooldridge I		Wooldridge II	
Supply of metro area student places	2.465	*			2.311	*	2.179	*
	(1.365)				(1.304)		(1.311)	
Supply interacted with high school in south	0.063				-0.364		-0.352	
	(0.414)				(0.391)		(0.391)	
Parent's residence near Helsinki			0.584	***			0.665	***
			(0.166)				(0.157)	
N	2477		2477		2477		2477	

Notes: A significant estimate is indicated by * (p<0.1), ** (p<0.05) or *** (p<0.01). The estimation sample includes graduates who resided either inside or outside the Helsinki metropolitan area for the entire four-year period between the third and sixth year after graduation. An indicator for whether one resided in the metro area is used as the dependent variable in the probit model. The basic independent variables in each model are age, age squared, a woman indicator, a Swedish speaker indicator, mother's and father's education and socioeconomic status, high school region, matriculation examination grades in math and first language, field of study, a bachelor's degree indicator and university graduation year. Heckman's model also includes a dummy variable for graduation from a metro area university.

CHAPTER 5: QUALITY OF HIGHER EDUCATION AND EARNINGS: EVIDENCE FROM FINLAND USING FIELD-OF-STUDY-LEVEL QUALITY MEASURES[91]

Using administrative data from Finland, this paper empirically examines the relationship between university graduates' early career earnings and three measures of university quality: the number of teachers per student, the number of publications per researcher and the number of applicants per admitted student. A distinction to previous studies is made by paying special attention to field-of-study heterogeneity: the quality measures are allowed to vary by a student's field, while the heterogeneity of earnings and individuals across fields is accounted for in the analysis. For the most part, the results indicate that the relationship between institution quality and earnings is rather weak; however, certain significant quality effects are also found. In particular, the teachers/student ratio is found to be positively associated with the earnings of women and graduates from the humanities. Overall, the results indicate considerable heterogeneity in quality effects across genders and fields.

Keywords: earnings; higher education; university choice; field of study; school quality

91 This paper has been accepted for publication in the *International Review of Applied Economics*. Financial support from the Academy of Finland is gratefully acknowledged (grant number 127049). I also wish to thank Tor Eriksson, Martin Nybom, Terhi Maczulskij, Tuomas Pekkarinen, Hannu Tervo and the seminar participants at the EALE 2011 congress in Cyprus and at the annual meeting of the Finnish Economic Society in Vaasa in 2012 for providing helpful comments.

1 Introduction

In today's world, higher education is a strongly valued product, and significant resources are dedicated to maintaining and improving its quality. Therefore, a relevant question is whether these investments make a difference for students and society as a whole. One way to evaluate this question is to study the labour market outcomes of students who graduated from higher education institutions of varying quality. For example, a researcher may estimate a relationship between students' subsequent earnings and a variable measuring the resources, reputation or selectivity of a higher education institution and – having controlled for student heterogeneity across institutions in a credible manner – make a claim that this relationship represents a causal effect of institution quality on earnings.[92] In the empirical literature, this type of analysis has been conducted numerous times and with datasets from different countries, including the U.S. (e.g., Behrman, Rosenzweig, and Taubman 1996; Brewer, Eide, and Ehrenberg 1999; Monks 2000; Dale and Krueger 2002; Black and Smith 2004, 2006; Long 2008; Hoekstra 2009), the U.K. (Chevalier and Conlon 2003), Sweden (Lindahl and Regnér 2005; Eliasson 2006; Holmlund, 2009), Italy (Brunello and Cappellari 2008; Triventi and Trivellato 2012), Israel (Lang and Siniver 2011) and Finland (Suhonen 2013). Most of these studies find that graduating from a prestigious institution is positively associated with success in the labour market; however, there has also been some evidence from both the U.S. (Dale and Krueger 2002) and Europe (Eliasson 2006; Holmlund 2009; Triventi and Trivellato 2012; Suhonen 2013) indicating a rather weak quality-earnings relationship.

In this paper, the relationship between university quality and graduates' early career earnings is examined in Finland. In this country, university education is offered by a system of state-funded, tuition-free institutions whose main campuses are located in ten city regions.[93] Although this system has, in principle, supported low quality differences between institutions by allocating educational resources in a rather equal manner, there is still evidence suggesting that quality differences exist within the system. For example, the findings documented in Suhonen (2013) show that there have been significant differences between universities in the average matriculation grades of students and in the way educational resources are divided across faculties and

---

92 It is worth noting that a positive relationship between the quality of higher education and labour market outcomes does not necessarily imply that there are social benefits from investments in quality. In particular, the social benefits could be marginal if the positive relationship merely arises because quality serves as a signal of high worker productivity in the labour market or because the choice of a prestigious institution enhances a student's networks (see Lang and Siniver 2011). In contrast, if the quality-earnings relationship reflects differences in students' human capital accumulation across institutions arising, for example, from differences in the quality of instruction, positive effects from quality investments, for example, on economic growth are to be expected.

93 For a comprehensive description of the Finnish university system, see, e.g., Ministry of Education (2005) and Suhonen (2013).

departments, which suggests that the level of peer effects and quality of instruction may have been different for students studying the same field but in different institutions. Therefore, an interesting question is whether these quality differences have generated earnings inequality among university graduates from different university units. The previous findings from Finland regarding returns from university choice – also reported in Suhonen (2013) – indicate that despite the locational heterogeneity of the universities, a student's decision of 'where to study' is not generally significant regarding his or her early career earnings. However, the effect of institution quality cannot be directly inferred from this evidence because the quality effects could be masked by other effects such as those arising from differences in job opportunities across regions. Therefore, the present study takes a more narrow perspective to returns to university choice by examining the relationship between early career earnings and explicit measures of institution quality, whereas the location of a university is merely considered as a variable to be controlled for.

The main contribution of the present study to the previous literature arises from the utilisation of quality measures that are allowed to vary within universities with respect to students' fields of study. Because a difference between universities, for example, in educational resources, peer quality or academic prestige may crucially depend on the type of field, major subject or study program in question, one may suspect that quality measures taking into account some of this heterogeneity can significantly alleviate biases arising from measurement errors (see Black and Smith 2006). The study data are obtained by merging three quality measures from the KOTA database of the Ministry of Education with a Statistics Finland administrative dataset, including a sample of individuals who graduated from one of the Finnish universities between 1995 and 2002. The quality measures include the number of teachers per student, the number of peer-reviewed publications per researcher and the number of applicants per admitted student. [94] Of these measures, teachers/student and applicants/admitted are commonly used as proxies of institution quality (e.g., Black and Smith 2006; Brunello and Cappellari 2008; Long 2008; Holmlund 2009); the first measure indicates the availability of teaching staff and, therefore, can be interpreted as a proxy for a university unit's resources offered to educational quality, whereas the second measure is associated with selectivity and peer quality. However, none of the previous studies appear to have used publications/researcher or other research performance measures in their analyses, although these measures are likely to be associated with the staff's academic competence, the institution's reputation and other important aspects of quality.

In the econometric analysis, the relationship between the quality measures and students' monthly earnings during the fourth year after the year of graduation is estimated using linear regression. During the analysis, a serious effort is made to alleviate biases arising from the non-random sorting of

---

94 For conciseness, the quality measures are from now on mainly referred to as 'teachers/student', 'publications/researcher' and 'applicants/admitted'.

students across university units based on earnings-enhancing individual characteristics. In the current context, a particular concern is that because both the quality measures and earnings vary considerably across fields of study, one's inability to control for individual heterogeneity across fields may result in biased quality effect estimates. In prior studies, problems arising from selection have been solved, for example, by focusing on identical twins (Behrman, Rosenzweig, and Taubman 1996) and, more recently, by using universities' admission cut-offs to generate exogenous variation in students' university choices (Hoekstra 2009; Lang and Siniver 2011). Many of these techniques are currently unavailable. For instance, admission cut-offs cannot be used because admission to Finnish universities is usually based on both high school grades and entrance examination scores, of which the latter are not available in the current dataset. However, similar to Suhonen (2013), changes in the supply of student places over time may be considered as a source of exogenous variation: because Finnish students are known to be relatively immobile – a large proportion of them attends a university in the region nearest to their place of origin – one may suspect that a change governed by the Ministry of Education in the supply of student places of a particular field in a student's nearest university region may have caused a change in the student's likelihood to end up in that field. Thus, in the current estimation strategy, these supply changes are used as an instrumental variable when conducting a correction for unobserved individual heterogeneity across fields of study; this correction is implemented using the method of Dubin and McFadden (1984). Otherwise, the results rely on controlling for a rich set of students' pre-university characteristics, including, for example, matriculation examination grades in math and first language and parents' socioeconomic characteristics.

The paper is organised as follows. In Section 2, the data are discussed in more detail, and descriptive evidence is presented to set up the empirical analysis. Section 3 describes the empirical approach, which – in addition to the above-discussed selection problem – addresses problems arising from measurement errors and 'fixed' differences across regions and fields. The estimation results are reported in Section 4, and Section 5 presents a summary and concluding remarks.

2 Data and Descriptive Analysis

The dataset for this study is obtained by matching three quality measures from the KOTA database of the Ministry of Education[95], teachers/student, publications/researcher and applicants/admitted, with administrative micro data provided by Statistics Finland. The micro dataset is based on a 7% random

95 The KOTA database includes yearly information on Finnish universities (disaggregated at the level of fields of study) from 1981 onward. The database is publicly available at https://kotayksi.csc.fi.

sample drawn from the Finnish population in 2001. The dataset includes yearly panel information, for instance, on individuals' earnings, job characteristics, educational attainment and family background for the period 1970–2006. This study uses only a small portion of the full dataset. The analysis is restricted to individuals who completed a university degree (master's or bachelor's) between 1995 and 2002.[96] Graduates from all fields of study except the 'arts' (music, theatre, fine arts or industrial arts) are included in the sample. Furthermore, the study sample only includes individuals who, prior to university, participated in the Finnish matriculation examination in 1985, 1987–1989 or 1991 or later because matriculation grades in math and first language are observed only from these years in the data.[97] With these restrictions, the sample size is limited to 3,586 graduates.

The labour market performance of the graduates is studied for the year that falls four years after the year of university graduation. Thus, for 1995 graduates, the observation year is 1999; for 1996 graduates, the observation year is 2000, etc. Average monthly earnings within the observation year are used as the dependent variable. This variable is obtained by dividing annual earned income (including both wage and entrepreneurial income) by the number of months worked. In the data, the number of months worked are reported only at the integer level and include both full-time and part-time months; therefore, the resulting earnings variable includes some measurement error. The use of monthly earnings is still preferred over annual earnings because the former are, obviously, highly sensitive to early career distortions in participation such as those arising from parental leaves.[98] With the exclusion of individuals having either zero or missing monthly earnings, the sample size is reduced by 301 observations; thus, the final sample comprises of 3,285 observations. Table 1 shows that the average student in the sample graduated from university at the age of 27 – most likely from the technology field – and earned 3,119 euros per

---

[96] With regard to an analysis of university quality effects, the data set has one potentially important limitation: university dropouts are not observed. This limitation must be kept in mind when interpreting the results: if the quality of the university institution affects both earnings and the probability of graduation, estimates could suffer from selection bias.

[97] Finnish high school students participate in the matriculation examination during their final year of high school. The examination includes numerous standardised tests, some of which are mandatory and some are not. Passing the first language test is mandatory for the completion of the examination; due to this requirement, the current sample is in practice formed by excluding individuals whose first language grade is missing from the sample of university graduates. However, those with missing grades for mathematics are not excluded because the math test is not compulsory: dropping these observations would result in a selection problem. In the sample used, the students without a math grade only comprise 13%.

[98] The possible bias arising from self-selection in participation could be further minimised by using earnings from several consecutive years. However, because the period for measuring the graduates' labour market performance had to be chosen to be rather short – to enable a sufficient sample size – it was regarded as preferable to use earnings that are determined as far off from the year of graduation as possible; graduates' 'starting wages' could be very sensitive to distorting factors, such as the gap between graduation and employment, yielding an unreliable picture of the quality effects.

month during the fourth year after the year of graduation. Furthermore, Table 1 illustrates that, despite the decentralised nature of the Finnish university system, university education is still geographically rather highly concentrated because 63% of the graduates completed their degree in one of the three largest city regions (Helsinki, Turku or Tampere).

TABLE 1 Means of selected variables (N = 3285)

Monthly earnings four years after graduation (€)	3119
Age at graduation	26.96
Woman	0.573
Swedish speaker	0.062
First language grade	4.112
Math grade	3.636
Math test taken at the A-level	0.645
Math test not completed	0.127
Bachelor's degree	0.065
Field of study	
Education	0.148
Humanities	0.141
Business	0.132
Social sciences	0.106
Law	0.047
Natural sciences	0.094
Technology	0.221
Medicine	0.083
Agriculture & forestry/sports	0.028
Region of graduation	
Helsinki	0.331
Turku	0.153
Tampere	0.150
Lappeenranta	0.041
Kuopio	0.036
Joensuu	0.056
Jyväskylä	0.089
Vaasa	0.030
Oulu	0.096
Rovaniemi	0.019
University quality measures	
Teachers/student	0.060
Publications/researcher	0.538
Applicants/admitted	4.383

Note: For calculating the average matriculation grades in math and first language, numeric values were assigned to the grades as follows: I = 0; A = 1; B = 2; C = 3; M = 4; E or L = 5.

The university quality measures are constructed and matched with the micro data according to the region of graduation, field of study and graduation year linked to the latest completed degree.[99] For example, all individuals who received a master's degree in education from Helsinki in 1995 have the same quality measure values. It should be emphasised that the university institution is not directly observed from the micro data but rather the geographical region in which the institution is located. Nevertheless, university institutions may be reasonably well identified. As we see from the list of Finnish universities in Appendix B, only two regions have more than one institution offering education in a specific field: in the Helsinki region, the business field is available in both a Finnish-speaking and Swedish-speaking university, and in the Turku region, most fields are available in both a Finnish-speaking and Swedish-speaking university. In the case of these two regions, an individual's first language is used as an additional criterion to identify the university by assuming that a Finnish speaker attended the Finnish-speaking university and that a Swedish speaker attended the Swedish-speaking university. Using these criteria, the data divide the Finnish university system into 59 units. From the bottom of Table 1, we see that in these university units, there were, on average, one teacher per every 17 students and 4.4 applicants per every admitted student. Furthermore, Table 1 shows that the average university unit published one peer-reviewed article per every two researchers within a year.

Figure 1 depicts the variation in the quality measures across fields and time. Above all, this figure illustrates that because of certain fundamental differences, for example, in teaching resource requirements and research practices, the average levels of the quality measures vary to some extent across fields. In particular, both the average amount of teaching resources and the average number of published peer-reviewed articles are considerably larger in medicine than in any other field, whereas the remaining field differences in these measures are more moderate. Another notable finding is that, relative to the other fields, the number of applicants per admitted student has been small in natural sciences and technology, reflecting the fact that it has been relatively easy to get admitted to study these disciplines. These findings stress the importance of controlling for 'fixed' differences between fields when using field-of-study-level quality measures. Figure 1 also highlights certain temporal changes in the field-of-study-specific measures during the period of measurement (the 1990s and early 2000s). In particular, in most of the fields, a noticeable declining trend occurred in teachers/student because of the rapid increase in the size of the student population. The publications/researcher ratios were rather steady throughout the period – with the exception of the declining trend in medicine – whereas certain diverging trends occurred in applicants/admitted: toward the end of the period, the relative number of applicants increased rapidly in education, while it decreased in business and law.

---

99 Detailed descriptions of the quality measures, as well as those of all other variables used in the analysis, are included in Appendix A.

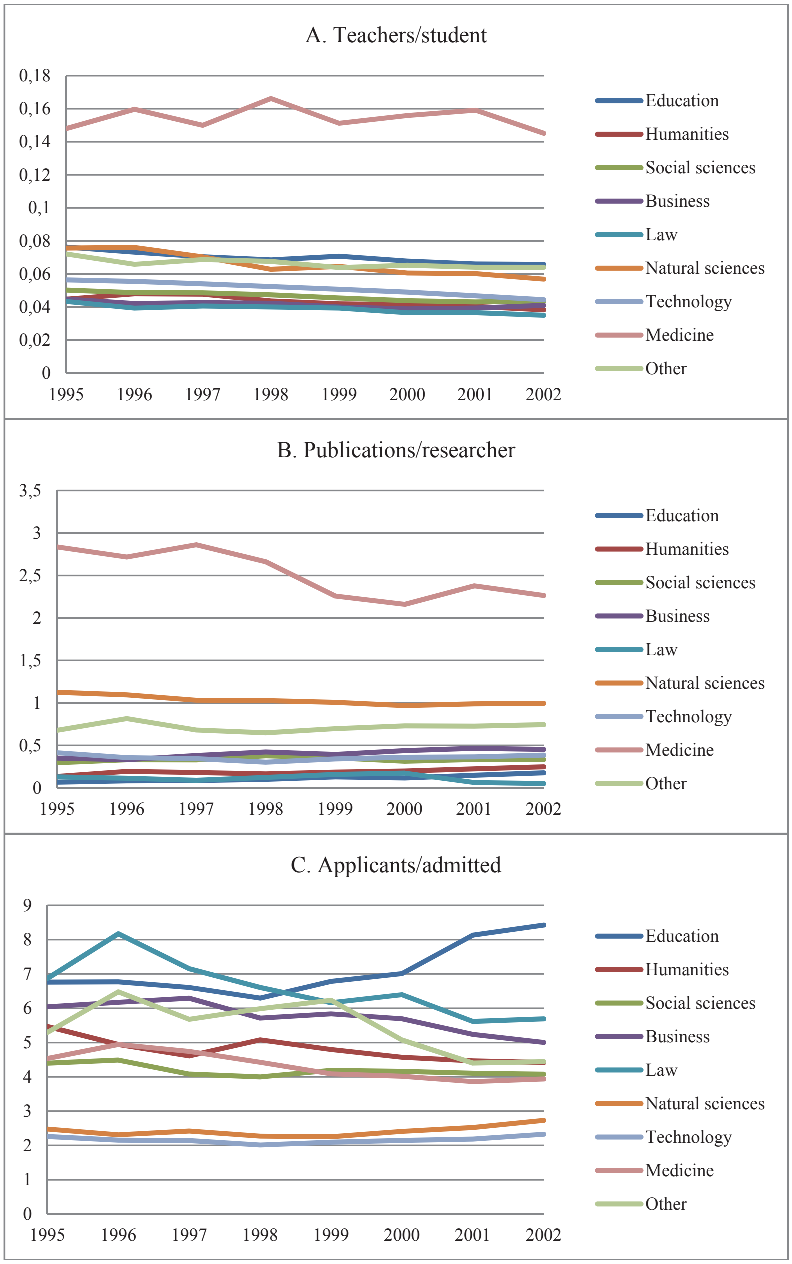


FIGURE 1 Average quality measure values by field of study and graduation year

To provide descriptive evidence of the interconnections between the quality measures and their association with monthly earnings and student ability, a number of inter-variable correlations is presented in Table 2. The correlations of the quality measures as such are unlikely to be very useful for causal interpretation because they are heavily driven by heterogeneity across fields in quality measure levels; therefore, the reported correlations are based on field-average-scaled versions of the quality measures, that is, the quality measures divided by their field-of-study- and graduation-year-specific average values.[100] A remarkable finding regarding the interconnection of the quality measures is that teachers/student and applicants/admitted are positively and significantly correlated with publications/researcher but negatively and significantly correlated with each other. This correlation pattern crucially differs from those presented in studies utilising institution-level quality measures (e.g., Black and Smith 2006; Holmlund 2009) because those studies systematically find large and positive correlations between all quality measures. When interpreted in the 'latent quality factor framework' (see Black and Smith 2006), the correlations presented in Table 2 suggest that the quality measures may serve as proxies for two separate quality factors, instead of a single quality factor, as follows: 1) because both teachers/student and publications/researcher represent a university unit's resources offered to educational quality, the positive correlation between these variables may arise from a 'resource factor', and 2) because a university unit with a good reputation is likely to have both competent researchers and a large number of applicants, the positive correlation between publications/researcher and applicants/admitted may arise from a 'reputation factor'. In the empirical analysis presented below, these interpretations are utilised for constructing the two quality factors that are used along with the original quality measures.

TABLE 2 Quality measures, monthly earnings and the math grade: inter-variable correlations

	Correlation coefficient	
Teachers/student - Publications/researcher	.108	***
Teachers/student - Applicants/admitted	-.212	***
Publications/researcher - Applicants/admitted	.051	***
Monthly earnings - Teachers/student	.035	**
Monthly earnings - Publications/researcher	-.021	
Monthly earnings - Applicants/admitted	.037	**
Math grade - Teachers/student	.010	
Math grade - Publications/researcher	.021	***
Math grade - Applicants/admitted	.063	***

Notes: The correlations are calculated using quality measures that have been divided by their field-of-study- and graduation-year-specific average values. The math grade variable is specified as in Table 1. A significant correlation coefficient is indicated by * ($p<0.1$), ** ($p<0.05$) or *** ($p<0.01$).

Table 2 further shows that two of the quality measures, teachers/student and applicants/admitted, are positively and significantly correlated with students' monthly earnings four years after the year of graduation, whereas the

100 For instance, the teachers/student ratio is .068 for an individual who graduated from the University of Helsinki in 1995, whereas the average value of this ratio is .074 for graduates in that field and year. Thus, the field-average-scaled teachers/student ratio for this individual is .912.

correlation between publications/researcher and monthly earnings is insignificant. Thus, these correlations provide some evidence suggesting that students from university units of high relative quality (relative to the field average) have earned more than students from units of low relative quality; however, the correlations are very small in magnitude (less than .04). Finally, for assessing whether there has been ability sorting of students across high- and low-quality university units, the correlations between the quality measures and the students' matriculation examination grades in math are shown at the bottom of Table 2. As expected, applicants/admitted is positively and significantly correlated with the math grade – although only by a factor of .06 – suggesting that a larger number of applicants is indeed associated with higher peer quality. A small but still significant correlation (.02) also exists between publications/researcher and the math grade, whereas the correlation for teachers/student is close to zero and insignificant. Thus, overall, these correlations suggest that within the Finnish university system, institution quality and student ability are rather weakly related.

3 Methodology

The empirical analysis of the quality-earnings relationship is based on the following linear earnings equation:

$$y_i = \alpha' \mathrm{q}_{\mathrm{rft}} + \beta' x_i + \gamma_r + \delta_f + \theta_t + \varepsilon_{irf}, \qquad (1)$$

where y_i denotes the natural logarithm of individual i's average monthly earnings during the fourth year after the year of university graduation; $\mathrm{q}_{\mathrm{rft}}$ is a vector of university quality measures for those graduated from region r and field of study f in year t; x_i is a vector of individual-level control variables; and ε_{irf} is the error term.[101] Because the quality measure values vary by i's region of graduation, field and graduation year, the possibility of fixed differences in mean earnings by these variables is accounted for in model (1) by the inclusion of terms γ_r, δ_f and θ_t. The graduation-year fixed effect θ_t is included to capture earnings differences arising from inflation, business cycles and other time-variant factors, whereas the purpose of the region-of-graduation- and field-of-study-specific fixed effects, γ_r and δ_f, is to consider the fact that earnings may vary across graduates from different regions and fields not only because of a variation in institution quality. In particular, controlling for δ_f may be crucial because one may suspect that the earnings differences across fields of study are largely attributable to certain fundamental differences in the characteristics of the fields, such as those in the skills sets and formal occupational qualifications

101 The error term is allowed to include university-unit-specific effects using a cluster robust covariance matrix estimator in which the number of clusters equals the number of university units in the data (59 in the full sample).

offered by them. Accounting for the region-of-graduation fixed effect γ_r may also be important because of locational differences between universities: students from large and economically active city regions – such as the Helsinki metropolitan area in Finland's case – could be more able than students from smaller regions to acquire relevant work experience already prior to graduation and could also be more likely to reside in 'lucrative' regions after graduation because of limited interregional mobility. Thus, there may be significant 'university location effects' that, in principle, must be accounted for when analysing the effect of institution quality on earnings (see *Suhonen* 2013). Furthermore, the sensitivity of the results to the inclusion of interactions between γ_r, δ_f and θ_t is studied during the analysis.

Throughout the analysis, two slightly different versions of model (1) – referred to as the *three-variable model* and the *quality factor model* – are employed. The three-variable model amounts to simply including all of the three quality measures, teachers/student, publications/researcher and applicants/admitted, in vector q_{rft}.[102] The purpose of the quality factor approach is to address a problem involved in the three-variable model arising from measurement error: at best, university quality measures are always only crude proxies for true quality factors; therefore, evidence based on these measures may biased. Black and Smith (2006) argue that measurement error biases in quality effect estimates can be reduced by replacing the original quality measures with a factor combining two or more of these measures. Therefore, based on the correlations between the field-average-scaled versions of the quality measures discussed in Section 2, the following two quality factors are constructed and used in the analysis: 1) a *resource factor* that combines the field-average-scaled versions of teachers/student and publications/researcher and 2) a *reputation factor* that combines the field-average-scaled versions of publications/researcher and applicants/admitted.[103]

In addition to non-quality-related effects and measurement errors, the evidence of university quality effects may be confounded because of student heterogeneity across university units. For instance, one may suspect that attending a high-quality university unit is associated with high individual ability and that, consequently, quality effect estimates are upward biased (see, e.g., Dale and Krueger 2002). When using field-of-study-level quality measures, a particular concern is that the choice of field is correlated with students' expected earnings through different unobserved individual characteristics such as occupational preferences and non-cognitive skills, which are not adequately captured by observed individual characteristics. Therefore, the estimates for the field-of-study fixed effects δ_f – and consequently also the estimates for the

---

102 In practice, logarithmic transformations of the quality measures are used because the log-log specification was found to yield slightly more precise estimates than the log-linear specification.

103 Because only two variables are used to form each quality factor, the resulting factors are simple linear combinations of the variables with unit standard deviation and similar factor loadings for both variables.

quality measures – are potentially biased.[104] In the current estimation strategy, the selection problem is addressed both by using observable control variables and by implementing a selectivity correction procedure to account for the endogeneity of the field-of-study choice. The set of variables controlling for 'selection on observables' includes a woman indicator, a Swedish-speaker indicator, age four years after graduation and its square, high school region, matriculation examination grades in math and first language and mother's and father's educational qualifications and occupational statuses. One may suspect that these variables already capture individual heterogeneity across regions and fields rather effectively. In particular, the grades are known to directly affect university admission in Finland and are likely to be strongly associated with students' mathematical and linguistic skills, preferences across different types of fields, etc. In addition, the former results presented in *Suhonen* (2013) suggest that high school region is an important variable to be considered when analysing returns to university choice in Finland because Finnish high school graduates are, in general, highly inclined to remain near their place of origin both during the university studies and afterwards.

To conduct an additional selectivity correction for the endogenous field-of-study choice, the control function approach proposed by Dubin and McFadden (1984) is implemented.[105] The first step of the procedure involves modelling a student's choice among the nine field-of-study alternatives using multinomial logistic regression. Thus, individual i is assumed to choose field j if this decision maximises his or her latent utility u_{ij}^*, which is assumed to be given by

$$u_{ij}^* = \mu z_{ij} + \pi' x_i + \vartheta_{ij}, \qquad (2)$$

where vector x_i denotes a set of individual-level control variables similar to that in equation (1), and ϑ_{ij} is a logistic error term.[106] The estimation of model (2)

104 Empirical studies have indeed found evidence that individuals choose their field of study in higher education partly based on their anticipated earnings in the fields (e.g., Arcidiacono 2004; Beffy, Fourgère, and Maurel 2012). These findings stress the importance of accounting for the endogeneity of the field of study with respect to earnings.

105 When correcting for selection bias in a multinomial-discrete-choice framework, studies have often relied on Lee's (1983) approach, in which the selection bias component (the control function) is formed solely on the basis of the predicted probability of the chosen alternative. However, according to the Monte-Carlo comparisons of Bourguignon, Fournier, and Gurgand (2007), the approach of Dubin and McFadden (1984) – which utilises the entire set of choice probabilities – is often preferable to Lee's approach.

106 The presence of explanatory variables that are determined after or along with the field of study – region of graduation, graduation year, the bachelor's degree indicator and the quality measures – does impose some problems for the estimation of the selection model. In particular, because the region of graduation and the quality measures are highly collinear with the choice of field – some of the region-of-graduation dummies even perfectly – the inclusion of these variables prevented STATA's maximum likelihood algorithm from converging to the solution. Therefore, the Dubin-McFadden procedure was conducted with excluding these problematic

yields the predicted probability P_{ij} for each individual i and field-of-study alternative j; in the second step of the procedure, these probabilities are used for constructing the selectivity correction terms for the earnings equation. An exclusion restriction for the earnings equation is generated by augmenting the selection equation with variable z_{ij} depicting i's nearest university region's proportion of the supply of student places in field j in i's high school graduation year. The logic behind the inclusion of this variable is that because individuals may be inclined to attend a university in their nearest university region, temporal changes in the supply of fields in that region may have generated variation in the individuals' field-of-study choices.[107] For simplicity and to enable efficient estimation, the coefficient of variable z_{ij} is assumed to be fixed across the field-of-study alternatives. While the relevance of z_{ij} as an instrument may be examined simply by testing whether $\mu \neq 0$ holds, the validity is, as usual, non-testable. Given that the fixed effect of i's high school region is controlled for in model (2), the validity relies on the assumption that high school graduates' unobserved earnings-enhancing characteristics have not changed over time in accordance with changes in z_{ij}.

When the selection model (2) is estimated using the full sample (N = 3285), the estimate for μ is positive (.020) with a p-value of .092. The supply effect implied by this estimate is sizeable: a one-percentage-point increase in the nearest university region's proportion of student places in field j is associated with a 2% increase in the odds of choosing that field. However, results from a supplemental analysis, conducted by estimating the selection model separately for men and women, indicate that the supply effect differs between the genders: the estimate of μ is positive and significant for women (.037 with a p-value of .02), whereas the estimate for men is very close to zero (–.002) and insignificant. Thus, it appears that the changes in z_{ij} over time have generated significant variation only in women's field-of-study choices, whereas this variable does not form a credible exclusion restriction for men. The selectivity-corrected estimates for men should therefore be interpreted with caution; these results rely solely on the functional form assumed for the selection bias component.

---

variables from the first-stage model. Dummy variables for graduation year and bachelor's degree are still included in both the first-stage and second-stage model.

107 Naturally, in order for the exclusion restriction to work, sufficient time variation in the regions' proportions of student places is required. In many cases, these proportions were rather steady during the 1980s and 1990s; however, several radical changes also occurred. For example, in the business field, Helsinki's proportion of the supply of student places was reduced from 50% to 35% between 1985 and 1999. At the same time, in the beginning of the 1990s, Oulu and Lappeenranta began to award university degrees in business, which rapidly shifted these regions' proportion of the supply of business education from 0% to 10%. Furthermore, figures calculated from the data yield certain indication that this regional expansion of business education may have indeed affected students' field choices: the proportion of students originating from remote regions, such as those located near Joensuu and Rovaniemi, is considerably higher among business graduates who graduated from high school in 1991 or later, that is, after the reform was initiated; at the same time, a noticeable decline occurred in the proportion of students originating from a close proximity of Helsinki.

In the second step of the Dubin-McFadden procedure, the earnings equation is estimated with controlling for a linear function of mean-centred selection unobservables $\vartheta_{ij} - E(\vartheta_{ij})$ as follows:

$$y_i = \alpha' \mathrm{q}_{\mathrm{rft}} + \beta' x_i + \gamma_r + \delta_f + \theta_t + \sum_{j=1}^{9} \rho_{fj} \lambda_{ij} + \eta_i, \qquad (3)$$

where $\lambda_{ij} = E[\vartheta_{ij} - E(\vartheta_{ij}) | u_{if}^* > \max_{k \neq f}(u_{ik}^*), \Gamma]$, and Γ denotes the full set of predictions from the first-stage selection model; η_i is an independent error term. Furthermore, with the logistic form assumed for ϑ_{ij}, the selection term in the case of $j = f$ may be written as $\lambda_{ij} = -ln(P_{ij})$, whereas for all $j \neq f$, these terms are given by $\lambda_{ij} = P_{ij} ln(P_{ij})/(1 - P_{ij})$ (see Bourguignon, Fournier, and Gurgand 2007). With nine choice alternatives, the total number of Dubin-McFadden selection terms included in equation (3) is 81.[108] Given the two-step nature of the procedure, the standard errors for the selectivity-corrected earnings equation estimates are obtained by bootstrapping with 500 replications.

4 Estimation Results

The baseline results regarding the relationship between university quality and graduates' monthly earnings are reported in Table 3. Let us first concentrate on the results from different specifications of the 'three-variable model' presented in the upper section of Table 3. Column 1 shows that, controlling for graduation year, degree level and graduates' pre-university characteristics, only one of the three quality measures, teachers/student, is significant. The estimate for this variable (.146) indicates that a 10% increase in teaching resources is associated with a 1.4% average increase in monthly earnings.[109] The estimates for publications/researcher and applicants/admitted are again approximately zero. In the second specification (Column 2), dummy variables for the field of study are added to the earnings equation to control for fixed differences across fields in the level of earnings and institutional characteristics. A remarkable finding is that despite considerable heterogeneity across fields, the estimates for teachers/student and publications/researcher are rather insensitive to the inclusion of these dummy variables; the first estimate increases slightly (to

108 Originally, Dubin and McFadden (1984) used an additional assumption that the correlations between error terms and sum to zero, which reduces the number of selectivity terms in each control function by one. However, Bourguignon, Fournier, and Gurgand (2007) have later on argued that this assumption is unnecessary and potentially harmful; thus, it is not used in the current analysis.

109 The coefficients of the three-variable model approximate percent changes in earnings for one-percent increases in the quality measures, whereas the coefficients of the quality factor models approximate percent changes in earnings for one-standard-deviation increases in the quality factors. In the text, estimates with a p-value of less than 0.10 are referred to as significant. More precise significance levels are included in the tables.

.155), whereas the latter estimate remains close to zero and insignificant. However, a sizeable change occurs in the estimate for applicants/admitted, which increases to .076 and becomes weakly significant with a p-value of .12. Judging from Column 3, the estimates of the three-variable model are also rather insensitive to the inclusion of region-of-graduation dummy variables that are expected to capture earnings heterogeneity arising from locational differences between the university units. In comparison with the previous specification, the estimate for teachers/student is slightly larger (.197), whereas the estimate for applicants/admitted is smaller (.051) and less significant. Finally, in Column 4, the three-variable model is augmented with 81 Dubin-McFadden selection terms to control for unobserved individual heterogeneity across fields. Above all, this selectivity correction has the consequence of reducing the estimate for teachers/student by 40% (to .118). Thus, it appears that the baseline OLS model overestimates the average association between teaching resources and earnings due to unobservables that are correlated with both the choice of field and earnings. In contrast, applicants/admitted receives a larger estimate (.071) in the selectivity-corrected model, indicating that OLS may even underestimate the return from attending a selective institution. The estimate for publications/researcher is close to zero and insignificant also in the fourth specification; thus, the evidence systematically suggests that, on average, there are no returns from graduating from an institution with a high research performance.

TABLE 3 Effect of university quality on log monthly earnings: ordinary least squares (OLS) and Dubin-McFadden-corrected estimates

(1) Three-variable model							
log Teachers/student	.146	**	.155	**	.197	**	.118
	(.055)		(.077)		(.079)		(.090)
log Publications/researcher	.009		-.014		-.006		-.011
	(.018)		(.022)		(.030)		(.027)
log Applicants/admitted	.008		.076		.051		.071
	(.039)		(.049)		(.060)		(.049)
Quality factor models							
(2) Resource factor	.028		.023		.028		-.009
	(.067)		(.037)		(.049)		(.044)
(3) Reputation factor	.027		.039		.025		.006
	(.067)		(.048)		(.071)		(.063)
Controls included:							
Basic controls	√		√		√		√
Field of study dummies			√		√		√
Region of graduation dummies					√		√
Dubin-McFadden selection terms							√
N	3285		3285		3285		3285

Notes: [a] The set of basic controls includes age, age squared, a woman indicator, a Swedish speaker indicator, mother's and father's educational qualifications and occupational statuses, high school region, matriculation grades in math and first language, a bachelor's degree indicator and graduation year. Standard errors are in parentheses. A significant estimate is indicated by * ($p<0.1$), ** ($p<0.05$) or *** ($p<0.01$).

The estimates from the 'quality factor models' are presented in the bottom part of Table 3. A comparison of Columns 1–3 reveals that these estimates are rather insensitive to the choice of control variables. Each of the first three specifications yields a small and positive coefficient estimate for both the resource factor (.023–.028) and the reputation factor (.025–.039). However, due to the large standard errors, none of these estimates are statistically significant. The Dubin-McFadden selectivity correction further reduces the estimates of the quality factors to approximately zero, indicating an upward selection bias in the OLS estimates. Thus, the quality factor models provide very little support for the existence of quality effects.

Given that the results have thus far relied on controlling for the region of graduation, field of study and graduation year separately in the earnings equation, it is reasonable to ask whether one should also control for some of the interactions between these variables. Interaction effects between the region of graduation and field of study could result, for example, from differences between institutions in the varieties of narrower disciplines within the broad fields of study. It is also possible that the region-of-graduation- and field-of-study-specific effects vary across graduates from different years because of changes in business cycle conditions, demand for skills, etc. These mechanisms are not directly related to institution quality and should therefore be controlled for. In Table 4, Columns 3–6 demonstrate that the estimates are fairly insensitive to allowing the field-of-study or region-of-graduation-specific effects to vary across years; the only noteworthy change occurs in the estimate for teachers/student, which is slightly smaller after these interactions are included. However, based on Columns 1 and 2, it appears that without controlling for the interaction term between the region of graduation and field of study, many of the quality effect estimates are biased in the positive direction. In particular, after the inclusion of these interactions, the estimate for applicants/admitted is close to zero, and the estimate for the reputation factor is even substantially negative, albeit not significant. The positive relationship between teachers/student and earnings is again robust to the region-field interaction terms; the non-selectivity-corrected OLS estimate for this relationship is even somewhat larger (.237) when conditioning on these interactions.

After analysing the average associations between the quality measures and earnings, we turn the discussion to the heterogeneity of these associations across genders and fields of study. At first, the gender differences are examined by estimating the baseline OLS and Dubin-McFadden-corrected models separately for women and men. Table 5 shows that the results differ considerably between the genders: in the case of women, all point estimates for the quality measures are positive, whereas all point estimates for men are negative. However, due to the large standard errors, very few of the gender-

TABLE 4 Sensitivity of results to the inclusion of additional interaction terms

	Included interaction term					
	Region of graduation X Field of study		Field of study X Graduation year		Region of graduation X Graduation year	
	OLS estimate	DMF estimate	OLS estimate	DMF estimate	OLS estimate	DMF estimate
(1) Three-variable model						
log Teachers/student	.237 ***	.118	.172 **	.089	.175 **	.093
	(.080)	(.148)	(.079)	(.090)	(.085)	(.093)
log Publications/researcher	-.024	-.020	-.006	-.019	-.002	-.007
	(.041)	(.039)	(.029)	(.029)	(.028)	(.028)
log Applicants/admitted	-.013	-.011	.064	.070	.070	.091 *
	(.063)	(.063)	(.059)	(.052)	(.063)	(.053)
Quality factor models						
(2) Resource factor	.000	-.047	.031	-.009	.027	-.011
	(.081)	(.076)	(.050)	(.043)	(.049)	(.046)
(3) Reputation factor	-.113	-.133	.034	.011	.045	.026
	(.090)	(.094)	(.074)	(.064)	(.073)	(.066)
N	3285	3285	3285	3285	3285	3285

Notes: Apart from the added interaction terms, the control variables in each model are the same as in Table 3. Standard errors are in parentheses. A significant estimate is indicated by * ($p<0.1$), ** ($p<0.05$) or *** ($p<0.01$).

specific estimates are significant.[110] The most noteworthy finding is that a university unit's teaching resources appear to be strongly associated with the earnings of women: both the OLS and Dubin-McFadden-corrected estimates for teachers/student obtained for women are larger than the previous full-sample estimates (.265 and .195). The OLS estimate for the resource factor is also large and weakly significant for women, implying a 9% increase in earnings for a standard-deviation increase in this factor; however, the selectivity correction considerably reduces the estimate for this variable. For men, publications/researcher is the most significant quality measure: both the OLS and the selectivity-corrected estimates for this variable are substantially negative and statistically significant (–.113 and –.097), indicating that a high research performance of a male student's university unit may even be harmful with regard to his early career earnings. In the case of men, large and negative estimates are also obtained for applicants/admitted and the two quality factors; however, these estimates lack statistical significance.

TABLE 5 Effect of university quality on log monthly earnings by gender

	Women				Men			
	OLS estimate		DMF estimate		OLS estimate		DMF estimate	
(1) Three-variable model								
log Teachers/student	.265	***	.195	*	-.002		-.037	
	(.081)		(.113)		(.133)		(.141)	
log Publications/researcher	.041		.031		-.113	**	-.097	*
	(.036)		(.040)		(.051)		(.053)	
log Applicants/admitted	.070		.065		-.008		-.077	
	(.072)		(.072)		(.066)		(.073)	
Quality factor models								
(2) Resource factor	.090	*	.058		-.109		-.100	
	(.053)		(.059)		(.069)		(.072)	
(3) Reputation factor	.094		.060		-.108		-.122	
	(.073)		(.087)		(.100)		(.104)	
N	1883		1883		1402		1402	

Notes: The control variables in each model are the same as in Table 3. Standard errors are in parentheses. A significant estimate is indicated by * (p<0.1), ** (p<0.05) or *** (p<0.01).

Finally, heterogeneity across fields of study is examined by including interaction terms with the quality measures and graduates' fields in the earnings equation estimated for the full sample. The point estimates from these

110 In the case of the gender-specific samples, there were difficulties in obtaining bootstrapped standard errors for the Dubin-McFadden-corrected estimates because the estimation of the first-stage conditional logit model did not converge to the solution for every bootstrap sample. In the case of men, this problem was particularly severe because almost all bootstrap rounds failed. Therefore, the ordinary cluster-robust standard errors are reported for men's Dubin-McFadden estimates in Table 5. The standard errors for women are again based on 261 successful bootstrap replications (out of 500 attempts).

models reported in Table 6 should be interpreted with caution: very few individual- and institution-level observations are used for identifying many of the field-of-study-specific estimates; therefore, they are somewhat imprecise.[111] Certain significant results are still obtained. In particular, a highly positive and significant relationship between teachers/student and earnings is observed in humanities regardless of whether the Dubin-McFadden correction is used (the point estimates are .764 and .746). The OLS estimate for teachers/student is also substantially large and significant in the case of education (.528), whereas the selectivity-corrected estimate for this field is somewhat smaller (.414) and less significant. Aside from the positive and weakly significant OLS estimate for natural sciences (.231), the remaining estimates for teachers/student – some of which are even negative – are not distinguishable from zero. The estimates for publications/researcher are mainly small and insignificant. However, a weakly significant and negative estimate for this variable is obtained in the case of business (–.121 and –.115 with and without the selectivity correction); the positive estimate for law again reduces considerably (from .109 to .066) and becomes insignificant after applying the selectivity correction. The relationship between applicants/admitted and earnings is most significant in the case of natural sciences; the large positive estimate for this field implies that a 10% increase in the applicants/admitted ratio is associated with 4% higher earnings. The estimate for technology is again large and negative (–.344 and –.316 with and without the selectivity correction) but only weakly significant due to the large standard error. Once again, the quality factor models yield very little significant results. Most notably, these results suggest that there is a strong positive association between earnings and the reputation factor in the case of natural sciences: a standard-deviation increase in this factor is associated with up to twice higher monthly earnings for the graduates of this field. Furthermore, evidence of a negative relationship between the resource factor and earnings is found in the case of business. The remaining estimates for the quality factors lack statistical significance and should therefore be interpreted cautiously.

5 Summary and Concluding Remarks

In this paper, the relationship between the quality of higher education and students' subsequent earnings was studied using a sample of Finnish students who graduated from university between 1995 and 2002. Although most of the results for the employed quality measures were insignificant, suggesting a rather weak quality-earnings relationship, certain significant linkages were also found. In particular, a positive and significant relationship was found between

111 The estimates for the category 'other field' (including 'agriculture and forestry' and 'sports') are excluded from Table 6 because these estimates are obviously unreliable given the small number of observations in these fields.

the teachers/student ratio and earnings, suggesting that there may be benefits for students from increasing the amount of educational resources. However, the results also indicate heterogeneity in this relationship: graduating from an institution with a high teachers/student ratio appears, on average, to be beneficial for women but not for men and in some fields (education, humanities and natural sciences) but not in others. Furthermore, the selectivity-corrected results obtained by the method of Dubin and McFadden (1984) indicate that, when using field-of-study-level quality measures, evidence of the association between educational resources and earnings may be biased upwards if the unobserved individual heterogeneity across fields is not controlled for.

TABLE 6 Effect of university quality on log monthly earnings by field of study

	(1) Three variable model			Quality factor models	
	log Teachers/ student	log Publications/ researcher	log Applicants/ admitted	(2) Resource factor	(3) Reputation factor
Education					
OLS estimate	.528 ***	.024	.043	.188	.036
DMF estimate	.414	.050	.071	.183	.068
Humanities					
OLS estimate	.764 **	-.016	.154	.119	-.193
DMF estimate	.746 **	-.036	.170	.088	-.209
Business					
OLS estimate	.162	-.115	.120	-.126 *	-.130
DMF estimate	.154	-.121 *	.095	-.132 **	-.162
Social sciences					
OLS estimate	-.243	.034	-.005	-.059	.186
DMF estimate	-.253	.034	-.011	-.071	.144
Law					
OLS estimate	.147	.109 **	.048	.242	.178
DMF estimate	.185	.066	.196	.126	.135
Natural sciences					
OLS estimate	.231 *	.120	.392 ***	.017	.703 ***
DMF estimate	.205	.166	.426 **	.001	.767 ***
Technology					
OLS estimate	-.287	-.026	-.316	-.031	.235
DMF estimate	-.395	-.053	-.344	-.075	.127
Medicine					
OLS estimate	.143	-.241	.178	.124	.293
DMF estimate	.046	-.159	.134	.016	.143

Notes: The control variables in each model are the same as in Table 3. A significant estimate is indicated by * (p<0.1), ** (p<0.05) or *** (p<0.01).

Very little evidence of returns from graduating from a selective institution – proxied by the applicants/admitted ratio – was obtained during the analysis. A significantly positive selectivity-earnings relationship was merely found in the case of natural sciences graduates. The remaining quality measures, the publications/researcher ratio and the two quality factors, were likewise insignificant in most cases; the results even indicated a negative association between these variables and the early career earnings of men.

Unlike previous studies, the present study utilised university quality measures that are allowed to vary within universities by students' fields of study. The use of these field-of-study-level quality measures offers a way to reduce biases from measurement errors and, furthermore, forces one to explicitly consider the possibility that there may be certain institution-level confounding heterogeneity that drives a statistical relationship between a measure of institution quality and students' earnings. Namely, when higher education institutions are located in different geographical regions and offer differing sets of field-of-study options, it might be the case that returns from the choice of location or field – rather than resources, selectivity or other factors of interest – explain the estimated quality effects. In the analysis presented above, the particular sensitivity of the results for the applicants/admitted ratio to different sets of region- and field-specific fixed effects may be interpreted as a sign of this problem.

Finally, given that certain evidence of quality effects was obtained, it is reasonable to contemplate potential mechanisms behind these effects. In particular, because the teachers/student ratio was, in many cases, positively associated with early career earnings, one might ask whether this positive relationship arises, for example, because better educational resources enhance students' human capital accumulation or because there are labour market signalling effects associated with graduating from an institution with good resources. The obtained results do not provide a direct answer to this question. However, if the explanation based on labour market signalling was correct, one would expect a positive correlation between educational resources and selectivity because signalling effects are usually assumed to arise from attending a highly selective institution. However, the teachers/student ratio and the primary measure of selectivity (applicants/admitted) were found to be inversely related in the current dataset. Furthermore, if the estimated positive returns from a high teachers/student ratio in the education, humanities and natural sciences fields were explained by signalling effects, one would expect the most prestigious institution to have the highest teachers/student ratio in these fields. However, the statistics show that the University of Helsinki, which is generally regarded as Finland's flagship university, has in fact had a relatively small amount of teaching staff in these fields. Based on this simple reasoning, the 'signalling story' does not provide a likely explanation for the positive relationship between educational resources and earnings. It is therefore possible that this relationship reflects 'genuine' human capital effects resulting from higher educational quality.

REFERENCES

Arcidiacono, P. 2004. "Ability Sorting and the Returns to College Major." *Journal of Econometrics* 121 (1–2): 343–375.

Beffy, M., D. Fougère, and A. Maurel. 2012. "Choosing the Field of Study in Postsecondary Education: Do Expected Earnings Matter?" *Review of Economics and Statistics* 94 (1): 334–347.

Behrman, J., M. Rosenzweig, and P. Taubman. 1996. "College Choice and Wages: Estimates Using Data on Female Twins." *Review of Economics and Statistics* 78 (4): 672−685.

Black, D., and J. Smith. 2004. "How Robust is the Evidence on the Effects of College Quality? Evidence from Matching." *Journal of Econometrics* 121 (1−2): 99−124.

Black, D., and J. Smith. 2006. "Estimating the Returns to College Quality with Multiple Proxies for Quality." *Journal of Labor Economics* 24 (3): 701–728.

Bourguignon, F., M. Fournier, and M. Gurgand. 2007. "Selection Bias Corrections Based on the Multinomial Logit Model: Monte Carlo Comparisons." *Journal of Economic Surveys* 21 (1): 174–205.

Brewer, D., E. Eide, and R. Ehrenberg. 1999. "Does It Pay to Attend an Elite Private College? Cross-Cohort Evidence on the Effects of College Type on Earnings." *Journal of Human Resources* 34 (1): 104–123.

Brunello, G., and L. Cappellari. 2008. "The Labour Market Effects of Alma Mater: Evidence from Italy." *Economics of Education Review* 27 (5): 564–574.

Dale, S. B., and A. Krueger. 2002. "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables." *Quarterly Journal of Economics* 117 (4): 1491–1527.

Chevalier, A., and G. Conlon. 2003. "Does It Pay to Attend a Prestigious University?" CEP Discussion Paper No. 33, London School of Economics and Political Science.

Dubin, J. A., and D. L. McFadden. 1984. "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption." *Econometrica* 52 (2): 345–362.

Eliasson, K. 2006. "The Role of Ability in Estimating the Returns to College Choice: New Swedish Evidence." Umeå Economic Studies No. 691, Umeå University.

Hoekstra, M. 2009. "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach." *Review of Economics and Statistics* 91 (4): 717–724.

Holmlund, L. 2009. "The Effect of College Quality on Earnings: Evidence from Sweden." Umeå Economic Studies No. 781, Umeå University.

James, E., N. Alsalam, J. C. Conaty, and D. To. 1989. "College Quality and Future Earnings: Where Should You Send Your Child to College?" *American Economic Review* 79: 247–252.

Lang, K., and E. Siniver. 2011. "Why is an Elite Undergraduate Education Valuable? Evidence from Israel." *Labour Economics* 18 (6): 767–777.
Lee, L. F. 1983. "Generalized Econometric Models with Selectivity." *Econometrica* 51 (2): 507–512.
Lindahl, L., and H. Regnér. 2005. "College Choice and Subsequent Earnings: Results Using Swedish Siblings Data." *Scandinavian Journal of Economics* 107 (3): 437–457.
Long, M. C. 2008. "College Quality and Early Adult Outcomes." *Economics of Education Review* 27 (5): 588–602.
Ministry of Education. 2005. *OECD Thematic Review of Tertiary Education: Country Background Report for Finland*. Helsinki: Publications of the Ministry of Education.
Monks, J. 2000. "The Returns to Individual and College Characteristics: Evidence from the National Longitudinal Survey of Youth." *Economics of Education Review* 19 (3): 279–289.
Suhonen, T. 2013. "Are There Returns from University Location in a State-Funded University System?" *Regional Science and Urban Economics* 43 (3): 465–478.
Triventi, M., and P. Trivellato. 2012. "Does Graduating from a Private University Make a Difference? Evidence from Italy." *European Journal of Education* 47 (2): 260–276.

APPENDIX A List of variables used in the analysis

- *Teachers/student*. This variable is based on the ratio (number of teaching staff) / (number of students) calculated from the KOTA database for each university unit at different years. The value of this variable for an individual graduated at year t is the average of these ratios at years t, t-1, t-2 and t-3.
- *Publications/researcher*. This variable is based on the ratio (number of published articles in international peer-reviewed journals within a year) / (number of research staff at the previous year) calculated from the KOTA database for each university unit at different years. The value of this variable for an individual graduated at year t is the average of these ratios at years t, t-1. (The smaller two-year time window had to be used for the calculation of this variable, as the publication information was available only from 1993 onward.)
- *Applicants/admitted*. This variable is based on the ratio (number of student applications) / (number of accepted applications) calculated from the KOTA database for each university unit during different years. The value of this variable for an individual graduated at year t is the average of these ratios at years t, t-1, t-2 and t-3.
- *Resource factor.* A factor combining field-average-scaled teachers/student and publications/researcher.
- *Reputation factor.* A factor combining field-average-scaled publications/researcher and applicants/admitted.

- *Monthly earnings*. This variable is constructed by dividing the annual earned income (both wage and entrepreneurial income) in the year that falls four years after the year of graduation by the number of months worked during that year.
- *Region of graduation*. The region or city in which the latest university degree was completed. This variable includes 10 categories: Helsinki metropolitan area, Turku, Tampere, Lappeenranta, Kuopio, Joensuu, Jyväskylä, Vaasa, Oulu and Rovaniemi.
- *Field of study.* The field of the latest university degree. This variable includes 9 categories: education; humanities; business; social sciences; law; natural sciences; technology; medicine; other (including sports and agriculture and forestry).
- *Graduation year.* The year the latest university degree was completed. This variable includes eight categories (years from 1995 to 2002).
- *Age*. The individual's age four years after graduation. The square of this variable is also included in each model.
- *Woman*. This variable gets a value 1 if a woman, and a 0 otherwise.
- *Swedish speaker*. This variable gets a value 1 if the individual's mother tongue is Swedish, and a 0 otherwise.
- *Bachelor's degree*. This variable gets a value of 1 if the obtained degree is a bachelor's degree, and a 0 otherwise.
- *Mother's/Father's educational qualification*. The latest observed educational qualification of mother/father (observations in 1970, 1980, 1990, and 2000). These variables include 12 categories: primary school or no classification available; high school diploma; secondary-level vocational qualification; lowest tertiary-level qualification; higher education degree (bachelor's or master's) in education; higher education degree in humanities; higher education degree in social sciences (including business and law); higher education degree in natural sciences or technology; higher education degree in medicine; other higher education degree; doctoral degree.
- *Mother's/Father's occupational status*. The latest observed occupational status of mother/father (observations in 1970, 1980, or 1990). These variables include six categories: farmer; entrepreneur; high-ranking official; low-ranking official; worker; no classification available.
- *Matriculation grade in first language*. The highest grade obtained in the first language test of the Finnish matriculation examination. The variable includes five categories (starting from the lowest grade): approbatur (A); lubenter approbatur (B); cum laude approbatur (C); magna cum laude approbatur (M); eximia cum laude approbatur (E) or laudatur (L).
- *Matriculation grade in mathematics*. The highest grade obtained in the mathematics test of the Finnish matriculation examination. This variable includes 13 categories: B-level improbatur (I); B-level approbatur (A); B-level lubenter approbatur (B); B-level cum laude approbatur (C); B-level magna cum laude approbatur (M); B-level eximia cum laude approbatur (E) or laudatur (L); A-level improbatur (I); A-level approbatur (A); A-level

lubenter approbatur (B); A-level cum laude approbatur (C); A-level magna cum laude approbatur (M); A-level eximia cum laude approbatur (E) or laudatur (L); no math grade.

- *High school region*. The region in which the individual completed a high school degree. This variable includes 19 categories: Uusimaa; Eastern Uusimaa; Finland Proper or Aland Islands; Satakunta; Tavastia Proper; Pirkanmaa; Päijänne Tavastia; Kymenlaakso; South Karelia; Southern Savonia; Northern Savonia; North Karelia; Central Finland; Southern Ostrobothnia; Ostrobothnia; Central Ostrobothnia; Northern Ostrobothnia; Kainuu; Lapland.
- *Nearest university region's proportion of the supply of student places in field j*. The number of first-year students in field j in the individual's nearest university region divided by the total number of first-year students in field j in the year of high school graduation. The nearest university region is determined according to road distances calculated from the largest municipality of the individual's high school region. The numbers of first-year students were obtained from the KOTA database.

APPENDIX B University institutions included in the analysis

University institution	Region	Fields of study
University of Helsinki	Helsinki	education, humanities, social sciences, law, natural sciences, medicine, agriculture and forestry
Helsinki School of Economics	Helsinki	business
Swedish School of Economics and Business Administration (S)	Helsinki	business
Helsinki University of Technology	Helsinki	technology
University of Turku	Turku	education, humanities, social sciences, law, natural sciences, medicine
Turku School of Economics	Turku	business
Åbo Akademi University (S)	Turku	education, humanities, business, social sciences, natural sciences, technology, medicine
University of Tampere	Tampere	education, humanities, business, social sciences, natural sciences, medicine
Tampere University of Technology	Tampere	technology
Lappeenranta University of Technology	Lappeenranta	business, technology
University of Kuopio	Kuopio	social sciences, natural sciences, medicine
University of Joensuu	Joensuu	education, humanities, social sciences, natural sciences, agriculture and forestry
University of Jyväskylä	Jyväskylä	education, humanities, business, social sciences, natural sciences, sports
University of Vaasa	Vaasa	humanities, business, social sciences
University of Oulu	Oulu	education, humanities, business, natural sciences, technology, medicine
University of Lapland	Rovaniemi	education, social sciences, law

Notes: The institution names and varieties of fields reflect the situation for graduates from 1995–2002. 'S' stands for a Swedish-speaking university institution.

SUMMARY IN FINNISH (YHTEENVETO)

Tämä väitöskirja koostuu johdantoluvusta ja neljästä empiirisestä tutkimuksesta, jotka käsittelevät suomalaisnuorten lukion jälkeisiä opiskeluvalintoja Suomen yliopistojärjestelmässä keskittyen erityisesti sijainnin rooliin potentiaalisena, yksilöiden valintoja ja myöhempiä työmarkkinatulemia ohjaavana tekijänä. Johdantoluku esittelee lyhyesti Suomen yliopistojärjestelmän erityispiirteitä sekä aikaisempaa tutkimuskirjallisuutta korkeakoulutusvalinnoista ja niiden yhteydestä työmarkkinatulemiin. Lisäksi esitellään yksinkertainen, teoreettinen malli, jossa yksilön yliopisto- ja koulutusalavalintoja selittävät muiden tekijöiden ohella muuttokustannukset, jotka liittyvät yksilön siirtymiseen kotipaikkakunnalta opiskelupaikkakunnalle. Johdantoluvun lopuksi tehdään yhteenveto väitöskirjan empiiristen tutkimusten tutkimuskysymyksistä, menetelmistä ja keskeisistä tuloksista.

Empiiristen tutkimusten pääasiallisena aineistona on käytetty Tilastokeskuksen yksilöaineistoa, joka pohjautuu seitsemän prosentin satunnaisotokseen Suomen vuoden 2001 väestöstä. Aineistoon on liitetty aikasarjainformaatiota mm. Tilastokeskuksen väestölaskennan ja työssäkäyntitilaston pitkittäistiedostoista sekä Ylioppilaslautakunnan rekisteristä vuosilta 1970–2006. Lisäksi tutkimuksissa hyödynnetään Opetus- ja kulttuuriministeriön KOTA-tietokantaa, joka sisältää yliopisto- ja koulutusalakohtaista pitkittäisinformaatiota Suomen yliopistojärjestelmästä. Empiirinen analyysi perustuu mikroekonometristen menetelmien hyödyntämiseen, ml. erilaiset moniluokkaisen valinnan mallit sekä lineaariset regressiomallit instrumenttimuuttujalaajennuksineen.

Suomen yliopistojärjestelmä on maantieteellisesti pitkälle hajautettu, minkä vuoksi monet sen yliopistoyksiköistä ovat suhteellisen pienikokoisia ja voivat tarjota vain muutamia koulutusalavaihtoehtoja opiskelijoilleen. Tämän vuoksi voidaan sanoa, että järjestelmässä on merkittävää alueellista variaatiota koulutusalatarjonnassa. Luvussa 2 tutkitaan, vaikuttavatko tästä alueellisesta variaatiosta johtuvat erot ”sisäänkirjautumisetäisyyksissä” yksilöiden koulutusalavalintoihin. Toisin sanoen: jos etäisyys tietyn koulutusalan lähimpään opinahjoon kasvaa, pieneneekö yksilön todennäköisyys valikoitua tälle alalle? Aikaisemmassa kansainvälisessä kirjallisuudessa on havaittu, että etäisyydet ovat merkittävässä yhteydessä korkeakoulun valintaan, mutta aiempaa evidenssiä etäisyyden vaikutuksesta koulutusalan valintaan on olemassa hyvin vähän. Tutkimuksen aineistona käytetään Tilastokeskuksen yksilöaineistosta poimittua aliotosta, johon kuuluvat vuosina 1991–1996 lukiosta valmistuneet opiskelijat. Kun näiden yksilöiden lukuisia ominaisuuksia, kuten ylioppilasarvosanoja äidinkielessä ja matematiikassa, vanhempien koulutustaustaa sekä lähtöalueen urbanisaatioastetta, on kontrolloitu conditional logit -valintamallissa, havaitaan, että 100 kilometriä suurempi etäisyys tiettyyn koulutusalaan on keskimäärin yhteydessä noin 15 prosenttia pienempään todennäköisyyteen valikoitua kyseiselle alalle. Etäisyysvaikutuksen havaitaan kuitenkin olevan varsin heterogeeninen riippuen valintavaihtoehdosta: etäisyydellä ei havaita olevan vaikutusta kasvatustieteiden, taidealan tai lääke-

/terveystieteiden opintoihin valikoitumiseen, mutta suuri ja tilastollisesti merkitsevä vaikutus suurimpaan osaan muista koulutusalavalinnoista. Conditional logit –mallien lisäksi tutkimuksessa estimoidaan nested logit –malleja, jotka mahdollistavat havaitsemattomista tekijöistä aiheutuvan korrelaation valintavaihtoehtojen hyötyjen välillä, mutta näillä estimoinneilla ei havaita olevan merkittävää vaikutusta tutkimuksen tulemiin.

Etäisyyksien ohella opiskelijoiden vanhemmilla voidaan olettaa olevan merkittävä vaikutus koulutusalavalintoihin sekä ”luonnon” (synnynnäiset kognitiiviset kyvykkyydet ja muut piirteet) että ”kasvatuksen” kautta. Koulutuksen siirtymistä sukupolvelta toiselle ollaan kuitenkin aikaisemmassa kirjallisuudessa tutkittu lähes pelkästään koulutustason (kvantiteetti) näkökulmasta kiinnittämättä huomiota koulutusalojen (kvaliteetti) periytyvyyteen. Luku 3 pyrkii täyttämään tätä kirjallisuudessa olevaa aukkoa esittämällä evidenssiä koulutusalan siirtymisestä vanhemmilta lapsille Suomen yliopistojärjestelmän kontekstissa. Toisin sanoen ollaan kiinnostuneita, kuinka paljon suurempi todennäköisyys opiskelijalla on valmistua tietyltä alalta, jos hänen vanhemmistaan vähintään toisella on yliopistotutkinto kyseiseltä alalta. Luvun 2 tavoin tutkimuksen otoksena ovat vuosina 1991–1996 lukiosta valmistuneet opiskelijat. Moniluokkaisten valintamallien estimoinneista saadut tulokset osoittavat, että koulutusalat siirtyvät merkittävässä määrin vanhemmilta lapsille: keskimäärin yksilöllä on noin kaksinkertainen mahdollisuus valikoitua tietylle alalle, jos hänen vanhemmallaan on tutkinto kyseiseltä alalta. Vanhempien vaikutuksessa havaitaan kuitenkin merkittävää vaihtelua niin valintavaihtoehdon, yksilön sukupuolen, vanhemman sukupuolen kuin havaitsemattomienkin tekijöiden mukaan. Oikeustieteellisen koulutusalan havaitaan erityisesti periytyvän muita koulutusaloja huomattavasti useammin. Lisäksi havaitaan, että miehet ovat naisia alttiimpia seuraamaan vanhempiensa koulutusalavalintoja, ja että sekä miehet että naiset seuraavat useammin isän kuin äidin valintoja.

Väitöskirjan kahdessa viimeisestä luvussa siirrytään tarkastelemaan korkeakoulutusvalintojen vaikutuksia yksilöiden valmistumisen jälkeisiin työmarkkinatulemiin. Luvussa 4 kiinnostuksen kohteena on yliopiston sijainnin merkitys, joka on aikaisemmassa kirjallisuudessa jäänyt varsin vähälle huomiolle. Tarkasteltavana tutkimushypoteesina on, että suomalaisopiskelija hyötyy (rahallisessa mielessä) päätöksestä sijoittua yliopisto-opintojen ajaksi Helsingin metropolialueelle pienempien yliopistokaupunkien sijaan, koska metropolialue pystyy tarjoamaan opiskelijoilleen mm. parempia työmahdollisuuksia ja (selektiivisyyden perusteella) korkealaatuisempia yliopistoja. Tutkimuksen otokseen sisällytetään vuosina 1994–2000 suomalaisista yliopistoista valmistuneet yksilöt, ja näiden yksilöiden tuloja seurataan valmistumisvuoden jälkeisinä kuutena vuotena. Empiirisessä analyysissä keskitytään yliopiston ja valmistumisen jälkeisen asuinalueen valinnan endogeenisuudesta johtuvien valikoitumisharhojen korjaamiseen käyttämällä sekä havaittujen tekijöiden kontrolloimiseen (”selection on observables”) että instrumenttimuuttujiin perustuvia menetelmiä; jälkimmäisessä lähestymistavassa eksogeenisen variaation lähteenä käytetään muutoksia metropolialueen opiskelijapaikkamäärissä yli ajan ja koulutusalojen.

Tulokset osoittavat, että metropolialueen yliopistoista valmistuneet opiskelijat ansaitsevat keskimäärin noin 14 prosenttia enemmän kuin muut opiskelijat valmistumisen jälkeisellä kuuden vuoden periodilla; tämä "metropolipreemio" selittyy kuitenkin kokonaan yksilöiden yliopisto-opintoja edeltävillä ominaisuuksilla, kuten ylioppilasarvosanoilla, perhetaustalla ja kotipaikkakunnan sijainnilla, kun taas sijainnin valinnalla itsellään ei havaita olevan merkitsevää kausaalivaikutusta tuloihin. Yliopistopaikkakunnan valinnan vaikutuksessa havaitaan kuitenkin jonkin verran koulutusalasta riippuvaa heterogeenisuutta, ja humanististen alojen opiskelijoiden havaitaan jopa hyötyvän sijoittumisesta metropolialueen ulkopuolelle. Tutkimuksessa löydetään lisäksi evidenssiä yksilöiden epäsatunnaisesta valkoitumisesta eri työmarkkina-alueille valmistumisen jälkeen havaitsemattomien, odotettuja ansioita selittävien tekijöiden mukaan, mikä vaikeuttaa yliopistopaikkakunnan ja työmarkkina-alueen vaikutusten erottelemista toisistaan.

Väitöskirjan viimeisessä luvussa (Luku 5) jatketaan yliopiston valinnan ansiovaikutusten tarkastelua, keskittyen eksplisiittisemmin yliopistoinstituution laadun vaikutukseen. Yliopistoyksikön laatua mitataan tutkimuksessa KOTA-tietokannan tietojen perusteella rakennetuilla kolmella laatumittarilla: opettaja-opiskelija –suhde, vertaisarvioitujen kansainvälisten julkaisujen määrä suhteessa tutkimushenkilökunnan määrään sekä hakijoiden määrä suhteessa opiskelijapaikkojen määrään. Erona aikaisempaan kirjallisuuteen laatumittarien sallitaan vaihtelevan yliopistoinstituutioiden sisällä koulutusalan mukaan, minkä voidaan odottaa pienentävän mittavirheistä aiheutuvaa harhaa estimaateissa. Tutkimuksen otos ja selitettävä muuttuja poikkeavat hieman Luvun 4 vastaavista: tällä kertaa tutkimuksen kohteena ovat vuosina 1995–2002 yliopistosta valmistuneiden henkilöiden keskimääräiset kuukausiansiot neljä vuotta valmistumisen jälkeen. Havaittujen tekijöiden, kuten ylioppilasarvosanojen ja vanhempien ominaisuuksien, lisäksi ekonometrisessa analyysissä pyritään kontrolloimaan koulutusalan valintaan vaikuttavia havaitsemattomia tekijöitä Dubinin ja MacFaddenin vuonna 1984 esittämän moniluokkaisen selektiivisyyskorjausmenetelmän avulla. Tulosten perusteella yliopistoyksikön laadun ja ansioiden välinen yhteys on keskimäärin heikko, joskin myös varsin heterogeeninen sukupuolen ja koulutusalan mukaan. Erityisesti opettaja-opiskelija –suhteen havaitaan olevan positiivisessa yhteydessä naisten ja humanististen alojen opiskelijoiden ansioihin. Useilla muilla koulutusaloilla ja miesten keskuudessa laatumittarien ja ansioiden väliset yhteydet ovat sen sijaan usemmiten ei-merkitseviä tai jopa negatiivisia.

Eräs väitöskirjan analyysien tuottama keskeinen johtopäätös on, että sijainnilla on merkittävä vaikutus suomalaisnuorten opiskeluvalintoihin – niin opiskelupaikkaa kuin -alaakin koskevaan valintaan. Tämä havainto voi olla merkittävä käynnissä olevan yliopistouudistuksen kontekstissa, sillä uudistuksen keskeisenä päämääränä on yliopistojen erilaisuuden lisääminen: mikäli tulevaisuuden yliopistojärjestelmä koostuu nykyistä erikoistuneemmista ja/tai laadultaan epätasaisemmista yliopistoista, voi yksilön kotipaikkakunnan sijainnilla olla entistä suurempi vaikutus yksilön vastaanottaman yliopistokoulutuksen laatuun. Tällainen kehitys voisi olla haitallista koulutusjärjestelmän tehokkuuden kannalta, koska yliopistojen opiskelupaikat allokoituisivat yhä enemmän ”epäoptimaalisille” opiskelijoille. Lisäksi alueiden väliset taloudelliset erot voisivat lisääntyä entisestään joidenkin alueiden kärsiessä tiettyjen alojen osaajien puutteesta. Toisaalta, mikäli uudistus parantaa yliopistojärjestelmän kansainvälistä kilpailukykyä, ja mikäli opiskelijoiden alueellinen liikkuvuus on tulevaisuudessa nykyistä suurempaa, voivat uudistuksen haitalliset vaikutukset jäädä vähäpätöisiksi suhteessa sen tuottamaan hyötyyn.

Toinen väitöskirjan keskeinen havainto on, että suomalaisopiskelijan maantieteellisellä sijoittumisella opiskelujen aikana ei yleisesti ottaen ole merkittävää vaikutusta hänen alku-uransa työmarkkinatulemiin. Tämän havainto implikoi, että yliopistojärjestelmä voidaan maantieteellisesti hajauttaa varsin pitkälle ilman, että tämä synnyttää merkittävää epätasa-arvoa eri yliopistoista valmistuneiden välille – ainakin, mikäli koulutukselliset resurssit allokoidaan suhteellisen tasaisesti yliopistojen kesken, kuten Suomessa on tehty. Näiden tulosten valossa – jotka siis heijastelevat 1990-luvun lopun ja 2000-luvun alun tilannetta – onkin mielenkiintoista nähdä, tuleeko meneillään oleva yliopistouudistus muuttamaan Suomen yliopistojärjestelmää heterogeenisempaan, suurempia alueellisia eroja tuottavaan suuntaan.