

Suhonen, Tuomo

Conference Paper

Distance deterrence effect and intergenerational similarity in the field-of-study choices of Finnish university students

52nd Congress of the European Regional Science Association: "Regions in Motion - Breaking the Path", 21-25 August 2012, Bratislava, Slovakia

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Suhonen, Tuomo (2012) : Distance deterrence effect and intergenerational similarity in the field-of-study choices of Finnish university students, 52nd Congress of the European Regional Science Association: "Regions in Motion - Breaking the Path", 21-25 August 2012, Bratislava, Slovakia, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/120664>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Distance deterrence effect and intergenerational similarity in the field-of-study choices of Finnish university students

Tuomo Suhonen

Jyväskylä University School of Business and Economics, P.O. Box 35, 40014, University of Jyväskylä, Finland. Tel.: +358503762537. E-mail: tuomo.j.suhonen@jyu.fi.

June 30, 2012

ABSTRACT

The field of study in postsecondary education is a highly influential determinant of labour market outcomes among highly educated individuals, which stresses the importance of studying mechanisms behind students' field-of-study choices. This paper examines the field-of-study choice among Finnish university students by focusing on two research questions that have thus far received very little attention in empirical studies: Does an increase in the geographical distance to a field decrease the likelihood of choosing that field? To what extent are students inclined to choose the same fields as their parents? For the econometric analysis, register-based data on a sample of high school graduates from 1991–1996 is utilised, and the choice among 10 field-of-study alternatives is modelled using conditional logit and mixed logit approaches. The obtained evidence suggests that an increase in the shortest distance to enrol in a field has a significant negative effect on the likelihood of choosing that field. Furthermore, evidence of considerable intergenerational similarity in field-of-study choice is found. However, whereas the results suggest that the distance deterrence is fairly homogeneous across individuals, evidence of sizeable variation in the parental effect is found with respect to both observed and unobserved characteristics.

JEL classification: C35; I23; J24

1 Introduction

The choice of field of study in postsecondary education is known to be an important determinant of students' future occupations and earnings. For instance, the completion of a university degree in business, law or engineering is generally observed to result in higher average earnings than a degree in education or humanities.¹ Thus, it may be argued that by the recent rapid rise of the general educational level, 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 mechanisms behind these choices. In numerous empirical studies on field-of-study choice, the role of several socio-demographic variables (for reviews, see Leppel et al., 2001; Goyette & Mullen, 2006) and the effect of expected earnings (Berger, 1988; Montmarquette et al., 2002; Bourdarbat, 2008; Beffy et al., 2012) have been covered extensively. However, whereas many studies have analysed the effect of geographical distances on educational attainment and the choice of university institution (e.g., Ordovensky, 1995; Long, 2004; Sá et al., 2004; Alm & Winters, 2009; Jepsen & Montgomery, 2009; Gibbons & Vignoles, 2012), very little evidence on the impacts of the *distance deterrence effect* in the choice of field has been presented. In theory, if a specific field is unavailable in a student's closest university, the various costs involved in moving to a more distant university location could lower his or her willingness to apply for that field. The distance deterrence effect could matter especially in countries where distances between university locations are relatively long and differences in the variety of fields exist across institutions. For instance, in the country studied in this paper, Finland, university degrees in many small fields (such as arts and law) are available only in a very limited number of locations, which may have generated regional disparities in access to these fields.

¹ For instance, Arcidiacono (2004) found large earnings premiums for certain college fields in the U.S. even after accounting for differences in individual ability across fields.

Another gap in the literature exists regarding parental influence: whereas many researchers have included measures of parents' educational level or socioeconomic status in their analyses of field-of-study choice (e.g., Leppel et al., 2001; Bratti, 2006; Goyette and Mullen, 2006; Boudarbat & Montmarquette, 2007; Wolniak et al., 2008), very little evidence of the association between individuals' and their parents' field-of-study choices has been presented thus far. On theoretical grounds, there are several reasons to expect that individuals are inclined to choose the same field as one of his or her parents: for instance, individuals could learn behavioural patterns from parental role models (Dryler, 1998), or educational and occupational preferences could be genetically inherited (Nicolau & Shane, 2010). The resulting intergenerational reproduction of field of study (that is, the 'doctor's son becoming a doctor' effect) could be an important factor that limits the openness and social cohesion of today's societies (Van de Werfhorst et al., 2001); this notion motivates the study of similarity between parents' and their children's field-of-study choices using nationally representative data sets.

This paper examines the choice of field of study among Finnish university students by focusing on the effects of geographical distances and parents. For the analysis, administrative data on a sample of Finnish high school graduates from 1991–1996 are utilised. More specifically, the study seeks answers to the following two main questions: “Does an increase in the distance to enrol in a field decrease the likelihood of choosing that field?” and “If a student's mother or father has graduated from a particular field, does this increase the student's likelihood of ending up in that field?” The Finnish university system offers a particularly good opportunity for the analysis of the distance effect: as Finland is a relatively vast and scarcely populated country – with 5.4 million people (in 2011) and a land mass of 338 432 km² – and has a decentralized university system, a great deal of variation in distances between students' pre-university residential locations and university locations is available for

the analysis. The current data set is also strong regarding the analysis of intergenerational similarity: in the data, parents' educational degrees are observed at the same level of accuracy as individuals' own degrees. Therefore, one may explicitly observe whether individuals chose the same fields as their parents. Also some of the general characteristics of the Finnish university system may be seen as advantages for the analysis. In particular, studying in the Finnish universities is free of charge, and therefore, virtually no financial barriers are involved in attending different institutions; the results on the determinants of field-of-study choice are therefore unlikely to be confounded by differences in financial resources.

The paper is organised as follows. Section 2 discusses theoretical explanations and previous empirical evidence regarding the distance deterrence effect and intergenerational similarity in postsecondary educational choices. Section 3 briefly describes the institutional setting of the studied Finnish university system. In Section 4, the data set of the study is discussed and descriptive evidence of the determinants of field-of-study choice is shown. Section 5 discusses the methodology used in the empirical analysis: the baseline results are obtained by the estimation of conditional logit models that control for individual-level observables such as matriculation examination grades and parents' characteristics, whereas robustness to unobserved individual heterogeneity is studied using random coefficient logit models. The results of the empirical analysis are presented in Section 6, and Section 7 concludes the paper.

2 Previous literature

2.1 Distance deterrence effect

Researchers have for long recognised that geographical distances may create barriers to individuals' educational choices because of various costs involved in migrating from home to school. In the context of postsecondary choices, the choice of a distant alternative instead of

one located within a close proximity could involve costs arising from at least three sources (see, e.g., Sjaastad, 1962; Leppel, 1993). First, transportation to the more distant location costs more time and money (direct costs). Second, information on the more distant alternative may be less available (informational costs). Third, individuals may feel uncomfortable about leaving from a familiar environment and living far from their family and friends (psychic costs). While no empirical evidence on the importance of different types of costs has been presented, one might suppose that in modern societies, the rapid development of transportation and communication technologies during the recent decades has ensured a fairly low level of direct and informational costs. Therefore, the psychic costs of migration could play a key role in explaining findings regarding the negative effect of distance on educational choices.

The role of the distance deterrence effect in postsecondary choices has already attracted a significant amount of attention in the literature. Closely related with the current paper are studies that use individual-level data for examining the effect of distance on the choice of postsecondary institution.² In one of the first studies within this branch, Ordovensky (1995) uses survey data from the U.S. to study the effects of tuition costs and distance to the choice of postsecondary institution type. Ordovensky's results from conditional and multinomial logit analyses indicate that an increase in the distance to the nearest 2-year and 4-year colleges significantly decreases a student's probability to enrol in these institution types. In another study from the U.S., Long (2004) uses a conditional logit approach for analysing over-time changes in the determinants of a student's choice among over 2000 college alternatives. Long's results suggest that although college choices are affected by distance deterrence, the magnitude of this effect has declined over time. More recently, Jepsen and Montgomery (2009) find evidence of distance deterrence in the context of community college

² Apart from micro-level studies, there is another branch of studies that uses aggregate-level data on student flows between students' home regions and university locations to study the effect of distances on postsecondary choices (e.g., Sá et al., 2004; Alm and Winters, 2009).

choice among mature workers of a specific U.S. state (Georgia); their methodological approach involves estimating a nested logit model for the joint decision to attend a community college and which college to attend. In a recent study by Gibbons and Vignoles (2012), the effects of distance on participation in higher education and the choice of institution are studied using administrative data from England. In their methodological approach, Gibbons and Vignoles deviate from the previously discussed papers by estimating a flexibly specified binary choice model for each institution alternative instead of a multinomial choice model. The authors find a significant distance elasticity in the probability of attending a specific institution.

The existing empirical evidence therefore indicates that the distance to a university institution and the likelihood of attending that institution are inversely related. In the light of this evidence, one might suppose that distances may also affect the choice of postsecondary field, in particular, if the variety of fields differs across institutions. However, one may also suspect that for many individuals, preferences regarding field-of-study choice are fairly strong (e.g., compared to preferences across institutions). Therefore, if individuals are willing to move large distances to be able to study their favourite fields, the effect of distance on field-of-study choice may prove to be economically insignificant. Very little empirical evidence on this matter exists. In a recent study, Bertrand-Cloodt et al. (2010) analyses the distance deterrence effect in the choice of upper secondary field among Dutch students. Their findings from the estimation of binary and multinomial logit models with three field-of-study alternatives suggest that a significant inverse relationship between the shortest distances to the fields and the choice probabilities of these fields exist. The current study differs from that of Bertrand-Cloodt et al. by focusing on postsecondary field-of-study choice. Another difference is that the currently studied educational system is far more geographically

dispersed than the Dutch upper secondary system. Therefore, more credible distance variation exists in the data of the present study.

2.2 Intergenerational similarity

There is a vast consensus among social scientists that parents may influence individuals' educational and occupational choices in important ways. A significant amount of theoretical and empirical work has been conducted on the role of family of origin in career development (for a review, see Whiston & Keller, 2004). In addition, the role of genetic inheritability in the intergenerational transmission of educational and occupational status has been studied to some extent (e.g., Nicolau & Shane, 2010). However, it seems that parental effect on the choice of educational field has received only a little attention in the literature thus far. In particular, only Dryler (1998) and Van de Werfhorst et al. (2001) seem to have previously presented theoretical explanations and quantitative evidence regarding the association between individuals' and their parents' educational fields.

In her study from Sweden, Dryler (1998) uses a large sample of upper secondary students to study gender differences in parental effect on field-of-study choices. She links her analysis to sociological theories stating that individuals learn behavioural patterns from adult role models and, in particular, from their parents. Therefore, individuals may be inclined to choose the same field as one of their parents did (same-sector effect) and, in particular, the field chosen by the parent of the same sex (same-sex effect). Alternatively, individual could be inclined to imitate the more influential parent (dominance effect). To test these theoretical hypotheses, Dryler models the selection into four broad field-of-study categories using binary logit models. She finds support for the same-sector effect in field-of-study choices of both boys and girls, whereas the same-sex hypothesis is supported only in the case of boys. Also weak support for the dominance effect is found.

The paper by Van de Werfhorst et al. (2001) studies the intergenerational resemblance in field of study using data from the Netherlands. Instead of studying field-of-study choice at a particular level of education, they pool educational degrees at different levels into seven broad field-of-study categories because of the small sample size. In addition, they focus only on the effect of a father on field-of-study choice. Apart from the role model explanation, Van de Werfhorst et al. give several other explanations for the intergenerational similarity of field of study. For instance, cultural capital acquired through family may provide students with extra skills and motivation for studies in fields chosen by their parents. Another explanation states that individuals may take their parents' social status as a reference point for their own career. Therefore, to achieve at least the same status as their parents, individuals self-select into educational fields that established their parents' status. In addition, a parent may encourage children to choose his or her own field, e.g., by providing study counsel or information regarding job opportunities in the field. In the empirical analysis, Van de Werfhorst et al. estimate conditional logit models including an indicator for each field for whether the field was chosen by the father. The estimated odd ratios for these indicators suggest that having a father graduated from a field significantly increases the odds of choosing that field.

The current analysis differs from those of Dryler (1998) and Van de Werfhorst et al. (2001) by focusing on parental effect in field-of-study choice at the university level. Another difference is that the current analysis uses a larger number of categories (10) for both individuals' own field and parents' fields. Therefore, a more accurate picture of the intergenerational similarity may be provided.

3 Institutional setting

The present study focuses on field-of-study choices within the Finnish university system.³ In the 1990s when high school graduates in the study sample entered postsecondary studies, the 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 today, but some of the institutions have merged. The institutions are state owned and strictly under the governance of the Ministry of Education; only recently in the late 2000s, the universities became more independent with more freedom in both academic and financial matters. All Finnish universities are free of charge, because of which virtually no financial barriers are involved in enrolling different institutions or fields. When applying for university studies, individuals apply directly for a specific major or field of study within a university institution. Acceptance is generally based on applicants' high school grades (in particular grades from the matriculation examination) and results from field-of-study-specific entrance examinations.

The geographical distribution of the university system is depicted in Figure 1. The system has been distributed in 10 city regions across the country: the Helsinki metropolitan area, Turku, Tampere, Lappeenranta, Kuopio, Joensuu, Jyväskylä, Vaasa, Oulu and Rovaniemi.⁴ Thus, the system offers university education relatively close to everyone's place of residence, which promotes regional equality in access to university education. However, as we see from Figure 1, not all fields of study are available in every university location, and therefore significant regional variation in distances to certain fields exists. For example, a university

³ In addition to the university sector, the Finnish higher education system has since 1996 included another sector, the polytechnics (a.k.a. universities of applied sciences), consisting of 28 schools located across the country. Whereas the purpose of the university sector is to provide high quality research and education at the bachelor's, master's and doctoral level, the polytechnic sector rewards vocational degrees that are usually paralleled by the bachelor's degree.

⁴ In addition, smaller university consortiums – serving as branch campuses of the university institutions – are currently located in six other cities. These consortiums are not, however, taken into account in the current empirical analysis when calculating distances to enrol in university fields: as the consortiums were only established in 2004 and mainly dedicated to adult education, they have unlikely had significant impact on field-of-study choices of the students in the study sample.

degree in law is available in only three locations, Helsinki, Turku and Rovaniemi, because of which some of the high school graduates have to move several hundred kilometres in order to obtain this degree.

4 Data and descriptive analysis

The study data is based on a seven per cent 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 utilizes a sub-sample consisting of individuals who graduated from high school between 1991 and 1996 at a young age (less than 23 years old) and, after that, completed a university degree no later than 2006; using these criteria, the sample size is limited to 3869 observations.⁵ The analysis of field-of-study choice is based on the first university degree that a student in the sample completed after high school.⁶ Based on a three-digit code depicting the level and field of the obtained degree, the degrees are divided into ten field-of-study categories: 'education', 'arts', 'humanities', 'business', 'social sciences', 'law', 'natural sciences', 'technology', 'medicine and health sciences', and 'other' (including agriculture and forestry, military sciences and sports sciences).⁷

⁵ Apart from the above-mentioned restrictions, only a minor restriction for the sample is conducted by excluding 4 individuals whose high school location could not be determined.

⁶ The data set includes information only on individuals' completed educational degrees, whereas no information on degrees in progress is observed. Regarding the analysis of field-of-study choice, this limitation has two implications that must be kept in mind when interpreting the results. First, a part of the individuals' field-of-study choices are not their initial choices, as some individuals change their field during the studies. Second, individuals who dropped out from university are excluded from the analysis. Therefore, if the variables of interest (distance to enrol and parent's field information) and the likelihood of dropping out are somehow interconnected, the estimates may suffer from selection bias.

⁷ Apart from field of study, the degree level (bachelor's or master's), graduation year and university location are observed from the data. Based on these data, only 15% of the completed 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 an average student in the sample graduated from university 7.4 years after graduating from high school. The average distance between a student's university location and pre-university residential location was 167 kilometres.

To study the effect of geographic distance on field-of-study choice, shortest road distances from 81 locations to 10 university cities are matched with the micro data on the basis of a student's NUTS4-level residential location one year before the year of high school graduation.⁸ Based on the distances, the shortest and second shortest distance to enrol in a field is determined. For the analysis of parental effects, the data include father's and mother's highest educational degrees. The parents' degrees are given at the same level of accuracy as the individuals own degree information. Therefore, one may explicitly observe whether individuals chose the same field category as their parents did. In the empirical analysis of the parental effect, two approaches are used: most of the analysis is based on using a single *parent's field* indicator, i.e., a 0/1 variable for whether either one of the individual's parents graduated from a field, but also models including separate variables for father's and mother's field-of-study choices, i.e., *father's field* and *mother's field* indicators, are estimated.

For the conditional logit and mixed logit analyses, the data is organised in the following way: the ten field-of-study alternatives are assigned for each individual in the study sample, and an indicator for whether a field was chosen or not is generated; this indicator serves as the dependent variable in the analysis. Thus, with 10 choice alternatives and 3869 individuals, the resulting data includes in total 38690 field-of-study choice observations. To obtain descriptive evidence of the relationship between field-of-study choice and the variables of interest, variable means for the chosen and non-chosen observations are calculated, yielding the results in Table 1. These results suggest that an inverse relationship between the distance to enrol in a field and the likelihood of choosing that field exists: both the average shortest distance and the average second shortest distance to enrol in a field are smaller (by 22 km) for chosen observations than for the non-chosen observations. In most cases, this inverse relationship shows also at the level of individual fields: in 8 out of 10

⁸ The NUTS-4 high school regions and university cities were depicted above in Figure 1. 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.

fields, the average shortest distance to enrol is smaller for the chosen than for the non-chosen observations. Particularly large differences in the shortest distances are measured in the three smallest field categories, arts, law and ‘other’, in which the number of available university locations is the most limited one (three locations for each field); for instance, in the case of law, the shortest and the second shortest distances to enrol are on average 68 and 53 kilometres smaller for those that actually chose to study the field than for those that did not.

Table 1 also yields evidence of considerable intergenerational similarity in field-of-study choice: higher frequencies of ‘ones’ for the parent’s field, father’s field and mother’s field indicators are observed for the chosen than for the non-chosen field alternatives. That is: an average individual in the sample is more likely to choose a field studied by one of his or her parents than a non-parental field. A particularly high degree of similarity between individuals’ and their parents’ field choices are observed in business, law, technology and medicine/health sciences. Moreover, we see that in most cases a larger relative difference between the chosen and non-chosen observations exists in the frequency of father’s field than in the frequency of mother’s field, suggesting that individuals more often follow their father’s field-of-study choice than the choice of their mother.⁹

Apart from the variables of interest, a rich set of individual-level variables are used as controls in the analysis of field-of-study choice. Table 2 shows field-of-study differences in several individual and family characteristics. Of 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 (64%), humanities (81%), 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

⁹ One should also note that in Table 1, the mean of the mother’s field indicator is nearly always smaller than the mean of the father’s field indicator. This finding reflects the fact that because of the women’s low educational level in the past, having a highly educated mother was essentially a rare event for the studied generation. As women’s educational level has rapidly increased during the last few decades, the weight of the ‘maternal effect’ may have also been higher for the younger generations.

even. Also noteworthy differences in mother tongue exist across fields: 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 arts and ‘other field’ (3% in both cases).¹⁰ In addition, Table 2 shows that students’ grades from the matriculation examination are somewhat correlated with field-of-study choice. In particular, the average first language grade is relatively high among humanities students, whereas students in technology have a relatively high A-level math grade. Obviously, the differences in the average grades reflect both differences in the difficulty of gaining acceptance in a field and students’ preferences over different types of subjects (e.g., mathematical and non-mathematical subjects). As for family background, 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 this field. Considerable field-of-study differences also exist in the type of the region that a student originated from. In particular, students from the Helsinki metropolitan area are clearly overrepresented in arts, business and law, but underrepresented in education and natural sciences.

5 Methodology

Although the descriptive evidence of Section 4 suggests that students’ field-of-study choices are associated with the distance to enrol and similarity with parents’ field-of-study choices, one may suspect that this evidence is confounded by the fact that only the outcomes of a constrained choice, i.e., student’s field of study conditional of getting accepted, is observed in the data: a part of the raw correlations could therefore arise from differences in the

¹⁰ 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 Swedish language is relatively abundantly provided in two Swedish-speaking universities, whereas more scarcity of Swedish-speaking education exists in other fields.

opportunity of getting accepted into different fields rather than from students' revealed preferences across fields. Furthermore, a particular selection problem is involved in measuring the distance effect: as a student's pre-university residential location is in most cases chosen by his or her family, the correlation between the distance to enrol in a field and the likelihood of that field could reflect an uneven distribution of individual ability or family endowments across space rather than a causal effect of the distance (e.g., Gibbons & Vignoles, 2012).¹¹

As most of the earlier studies on postsecondary educational choices, the current econometric analysis is limited to cross-sectional analysis: the choice of field is modelled in a static discrete choice framework, and individual-level observables such as matriculation examination grades and parents' socioeconomic status are used to alleviate the above-discussed problems arising from constraints and selection. For the econometric analysis, two alternative discrete choice methods are used: the *conditional logit model* and the *mixed logit model*. In both approaches, an individual i 's field of study choice is assumed to be determined by the maximization of the latent utility u_{ij}^* over the field-of-study alternatives $j = 1, 2, \dots, 10$. That is, the observed choice j^* is assumed to satisfy $u_{ij^*}^* = \max(u_{ij}^*)$. The baseline results are obtained by assuming that the latent utilities are of the following conditional logit form:

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

where α_j is a field-specific constant term; d_{ij} is the shortest distance from i 's pre-university residential location to field j (divided by 100 km); p_{ij} is 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), the parameters of interest are β and γ which based

¹¹ For example, the finding that compared to other students, law students more often originate from a close proximity of universities that have a faculty of law could partly reflect differences in the likelihood of getting accepted into these universities.

on discussion above, are expected to take negative and positive signs, respectively. Several extensions of model (1) are also considered during the analysis. These extensions – serving as robustness checks and allowing for detecting heterogeneity in the effects – include 1) a model in which β and γ are allowed to vary across genders, 2) a model that includes both the shortest and the second shortest distance to field j and separate indicators for mother's and father's field, 3) a model including also quadratic terms of the distance variables, and 4) a model in which β and γ are allowed to vary across the field-of-study alternatives (that is, parameter β is replaced with β_j).¹²

The conditional logit model allows for a fast estimation of the distance and parental effects. However, this is achieved by the cost of making restrictive assumptions regarding the role of unobservables: the slope coefficients of the model variables are assumed to be constant across observably similar individuals, and error terms ϑ_{ij} are assumed to be uncorrelated across the field-of-study alternatives.¹³ These assumptions guarantee the following closed-form solution for the probability of selecting field j :

$$P_{ij}^{CL} = \frac{e^{u_{ij}^*}}{\sum_{k=1}^{10} e^{u_{ik}^*}}, \quad (2)$$

which is used to construct the log-likelihood function for the maximum likelihood estimation of the model parameters. In practise, the conditional logit estimation is carried out by using STATA's package *asclogit*.

¹² It should be noted that the small sample size (3869) combined with the large number of field-of-study categories (10) limits the choice of model specification to some extent. For instance, in the estimated models, the coefficients of the individual-level controls are restricted to be equal across genders, since these would be inefficiently estimated otherwise. Furthermore, the cross-field coefficients for the variables of interest (e.g., the coefficients of the shortest distance to enrol in humanities in the utilities of other fields) are restricted to zero, since including all 90 cross terms in the model inevitably results in inefficiency and multicollinearity problems. This restriction implies that the cross-field marginal effects of the variables of interest are of an opposite sign than the own-field marginal effects. By intuition, the restriction on cross terms is more plausible for the distance to enrol than for the parent's field indicator: for instance, it may be realistic to assume that an increase in the shortest distance to enrol in humanities causes individuals to substitute towards other fields, whereas having a parent graduated from humanities could have positive cross-field effects for similar fields such as arts.

¹³ The uncorrelated error terms assumption is known to imply that the choice probabilities of the conditional logit model have the independence of irrelevant alternatives (IIA) property: the odds ratio P_{ik}/P_{il} between choice alternatives k and l is assumed to remain constant after including any other choice alternative in the choice set (Train, 2003).

As many of the potential determinants of field-of-study choice are unobserved, there is a high risk that the assumptions of the conditional logit model do not hold. For this possibility, a more flexible estimation approach, known as the *mixed logit model*, is used as a robustness check during the analysis. In the applied mixed logit model, the coefficients of the variables of interest in model (1) are allowed to vary randomly across individuals, whereas the coefficients of the control variables are still assumed to be fixed. Thus, in this approach, the latent utility from choosing field j is given by:

$$u_{ij}^* = \alpha_j + \beta_i d_{ij} + \gamma_i p_{ij} + \delta_j' x_i + \vartheta_{ij}, \quad (3)$$

In model (3), the idiosyncratic parameters β_i and γ_i are assumed to be normally distributed, and, in addition, they are allowed to be correlated. Thus, the mixed logit approach involves estimating mean parameters $E(\beta_i) \equiv \mu_\beta$ and $E(\gamma_i) \equiv \mu_\gamma$ and variance-covariance matrix elements $Var(\beta_i) \equiv \sigma_\beta^2$, $Var(\gamma_i) \equiv \sigma_\gamma^2$ and $Cov(\beta_i, \gamma_i) \equiv \sigma_{\beta,\gamma}$. By the introduction of these parameters, the assumptions of taste homogeneity and uncorrelated error terms are partially relaxed, while the model is still kept tractable.¹⁴

As the idiosyncratic terms in model (3) are 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 the chosen parameter distribution as follows:

$$P_{ij}^{ML} = \iint P_{ij}^{CL} f(\beta_i, \gamma_i | \mu, \sigma) d\beta_i d\gamma_i, \quad (4)$$

where P_{ij}^{CL} is the conditional logit probability, and $f(\beta_i, \gamma_i | \mu, \sigma)$ is the joint density of β_i and γ_i with vector of means μ and covariance matrix σ ; in the analysis, the coefficients are assumed to be normally distributed, and thus $f(\beta_i, \gamma_i | \mu, \sigma)$ in model (4) denotes the bivariate normal density. The method of *maximum simulated likelihood* is used for the estimation of

¹⁴ Alternatively, one could directly allow the error terms to be correlated across choice alternatives, e.g., by using the multinomial probit model (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.

the mixed logit parameters; in practise, the estimation is carried out by using STATA's user-written command *mixlogit* (see Hole, 2007).

A rich variety of individual-level controls is included in the above-described models in order to alleviate problems arising from individual heterogeneity. The detailed descriptions of these variables are included in the Appendix. The basic controls include a female indicator, a Swedish speaker indicator and dummy variables for high school graduation year; these variables may be considered as 'safe' controls whose purpose is only to level off regional and family differences in demographic characteristics.¹⁵ The students' matriculation examination grades in first language and mathematics are included in the model in order to control for individual differences in the opportunity of getting accepted into different fields; thus, these variables are perhaps the most crucial identifying variables in the model. In addition, in order to control for other family and regional characteristics that are potentially affecting the field-of-study choice along with the variables of interest, the models are augmented with indicators for mother's and father's educational levels and vocational statuses and the type of pre-university residential region; these variables may be considered as potential 'bad controls', as they are determined along with distances and parents' field-of-study choices and may therefore cause biases in the estimates (Angrist & Pischke, 2009). For this possibility, the estimates are reported in Section 6 both with and without the inclusion of the parental and regional controls in the models.

The final methodological note concerns the interpretation of the results. After estimating the conditional and mixed logit models, the magnitudes of the distance and parental effects are difficult to interpret directly based on coefficient estimates $\hat{\beta}$ and $\hat{\gamma}$, as these coefficients depict changes in the 'log-odds' of choosing field j in response to unit changes in the variables. Thus, for easier interpretation, the effects reported in Section 6 are based on

¹⁵ 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 at that year.

$\exp(\hat{\beta}) - 1$ and $\exp(\hat{\gamma}) - 1$, which give the implied effects in terms of relative changes in the odds of choosing field j , i.e., $P_{ij}/(1 - P_{ij})$. Alternatively, the interpretation could be based on marginal effects $\partial P_{ij}/\partial d_{ij}$ and $\partial P_{ij}/\partial p_{ij}$ evaluated for example at the means of the explanatory variables. However, reporting odds ratios is preferred for two reasons. First, as most of the estimated models do not allow for field heterogeneity in coefficients β and γ , reporting marginal effects for different fields would be contrived. Second, as the number of choice alternatives is relatively large (10), $P_{ij}/(1 - P_{ij})$ is on average very close to P_{ij} , and odds ratios therefore approximate choice probability elasticities reasonably well.

6 Results

In this section, results from the estimated conditional logit and mixed logit models are reported. At first, the distance and parental effects in field-of-study choice are studied using conditional logit models in which the effects of the variables of interest are restricted to be constant both across individuals and choice alternatives. The baseline results shown at upper part of Table 3 are obtained by including the shortest distance to enrol and the parent's field indicator as the only alternative-specific variables in the model. These results are in line with the above-discussed theoretical predictions and descriptive evidence: in all specifications, the shortest distance to enrol receives a significantly negative coefficient, whereas the coefficient for the parent's field indicator is significantly positive. The estimate for the shortest distance to enrol is extremely robust to the choice of specification: the implied change in the odds of choosing a field arising from a 100 km increase in the shortest distance is only slightly larger in models controlling for matriculation examination grades (−21%) than in models without these controls (−20%); the inclusions of demographic, parental or regional variables have only negligible impacts on this estimate. The estimate for the parent's field indicator is more sensitive to controlling for individual heterogeneity: whereas the implied effect of having a

parent graduated from a field on the odds of choosing that field is 117% in the model controlling only for demographic variables, the inclusion of matriculation examination grades decreases this estimate to 93%;¹⁶ apart from this change, the estimate for the parent's field indicator is fairly insensitive to individual-level controls. By allowing the effects to vary by gender (bottom part of Table 3), no statistically significant gender difference is found in the effect of the shortest distance to enrol; the point estimates for this variable are roughly similar across men and women in all specifications. However, the estimates for the parent's field indicator are larger for men than for women, suggesting that men have a higher probability than women to end up in a field chosen by their parents. Furthermore, the estimate for the parent's field indicator is more sensitive to controlling matriculation examination grades in the case of women than in that of men: by the introduction of these controls, the implied effect for women (men) decreases from 100% (138%) to 72% (124%). Thus, the estimated gender difference in intergenerational similarity is larger and more significant when controlling for the grades.

In the remainder of this section, the robustness of the baseline results is studied by using alternative model specifications that allow for different sources of heterogeneity in the effects of the variables of interest. Given that the inclusion of the parental and regional variables in the models may be problematic (as these variables are determined along with the variables of interest), the results are hereafter reported with and without these controls; for brevity, the model including only demographic variables and matriculation examination grades as individual-level controls is referred to as the *short specification*, whereas the model including also the parental and regional variables is referred to as the *long specification*.

Table 4 shows results from an extended conditional logit model in which two variables are used to indicate both the distance to enrol (the shortest distance and the second shortest

¹⁶ This finding suggests that part of the observed intergenerational similarity in field-of-study choice arises because of individual differences in access to certain fields rather than from differences in preferences across parental and non-parental fields.

distance) and similarity with parent's choices (mother's field and father's field). In this model, the evidence of distance effect is somewhat sensitive to the choice of specification: in the short specification, both the shortest distance and the second shortest distance receive negative and significant estimates, the implied effects of 100 km increases in these distances being -18% and -10% , respectively, whereas only the shortest distance is significant (with an estimated effect of -20%) in the long specification.¹⁷ No statistically significant gender differences are found in the effects of the shortest distance or the second shortest distance; the point estimates for the second shortest distance are still clearly larger and more significant for men than for women. The estimates for the mother's field and father's field indicators suggest that both parents' field-of-study choices are significantly associated with their children's choices. However, individuals seem to be more likely to follow their fathers' choices than their mothers' choices: according to both the short and long specification, having a father graduated from a field is associated with a roughly 2.1 times higher odds of choosing that field, whereas the corresponding effect of a mother is only 1.3. The results with gender interactions suggest that the father's influence is greater for both genders.¹⁸ Furthermore, the estimated effects of the mother's field and father's field indicators are somewhat larger for men than for women, but these gender differences are not statistically significant.

Thus far, in the estimated models, the distances have been assumed to affect the latent utilities of the field-of-study alternatives in a linear manner. To test whether the linearity assumption is plausible, the distance effects are re-estimated with including also the quadratic terms of the distances in the models. Table 5 shows that the results from these models are to

¹⁷ A more careful sensitivity check reveals that the main part of the large change in the estimate of the second shortest distance (from -10% to -4%) arises from the inclusion of the type of pre-university region in the model. Thus, the results suggest that certain unspecified regional heterogeneity explains part of the correlation between field-of-study choice and the second shortest distance to enrol.

¹⁸ Based on Wald tests conducted using STATA's *test* command, the difference between father's and mother's effects is statistically significant in all reported specifications.

some extent sensitive to the choice of control variables: in the short specification, both the linear and quadratic terms of the shortest distance to enrol receive sizeable negative coefficients, implying that negative effect of the shortest distance increases by an increase in the distance, whereas the coefficient for the quadratic term reduces to approximately zero after including the parental and regional controls in the model. Compared to the linear model, the nonlinear model therefore yields a slightly smaller estimate for the mean effect of the shortest distance to enrol when using the short specification (−20% in Model I and −17% in Model II) but an approximately similar estimate with the long specification. The quadratic term for the second shortest distance to enrol is close to zero in both specifications, and the implied mean effect of this variable is also roughly similar in the linear and nonlinear models. Thus, the linear model seems to perform reasonably well in capturing the mean distance effects. As the quadratic terms are not significant in any of the specifications, and as the standard errors for the estimates increase by the inclusion of these terms, the linear model may even be preferred on the basis of efficiency.

The fourth set of results is obtained by allowing the own-field effects of the shortest distance to enrol and the parent's field indicator to vary across fields in the conditional logit model. Qualitatively, these results reported in Table 6 are in line with those obtained from the restricted model: with both the short and long specification, the field-specific estimates for the shortest distance to enrol are negative (apart from the positive but insignificant estimate for education in the short specification), whereas the estimates for the parent's field indicator are positive. However, the magnitude and statistical significance of the estimated effects varies to some extent across fields and specifications. According to the results obtained with the short specification, an increase in the shortest distance to enrol has a significantly negative and sizeable impact on the odds of choosing a field in the cases of arts (−22%), business (−25%), social sciences (−25%), law (−34%), technology (−17%), medicine/health

sciences (−14%) and ‘other field’ (−35%), whereas the estimates for the remaining three fields are smaller and insignificant. The inclusion of the parental and regional controls has dissimilar effects on the distance effect estimates for different fields: in particular, the negative estimates for humanities and natural sciences increase and become significant, whereas the estimates for arts, business and medicine/health sciences decrease and lose their significance.¹⁹ The estimate for the parent’s field indicator is significantly positive in both specifications in the cases of education, business, social sciences, law, technology and medicine/health sciences. In addition, the estimate for arts is significant only in the short specification. The suggested degree of intergenerational similarity is clearly highest in law: the estimate for this field indicates that a student having a parent graduated from law has a 5.7 or a 4.6 times higher odds of graduating from that field compared to a similar student without that type of parent. Among other fields with a significant estimate, the corresponding multiplier varies between 1.9 (in technology) and 2.8 (in business).²⁰ Table 6 also reports estimates for the mean effects of the variables obtained by taking weighted averages of the field-specific estimates (field sizes are used as weights). The estimated mean effects of the shortest distance to enrol are −14% and −19% in the short and long specifications, respectively. Thus, the model assuming constant own-field effects (see Table 3) seems to overestimate the distance effect to some extent. The mean estimate for the parental effect is then again slightly larger than in previous results: 107% in the short specification and 104% in the long specification.

Finally, the effects of the shortest distance to enrol and the parent’s field indicator are studied using the mixed logit approach which allows for individual-level heterogeneity in the

¹⁹ Because of the large standard errors, most of the differences between the field-of-study-specific estimates for the shortest distance to enrol are not statistically significant according to the calculated Wald tests. As for these differences, the most notable finding is that in the short specification, the positive estimate for education differs significantly (at the 10 per cent level) from the estimates of all other fields except humanities.

²⁰ According to the calculated Wald tests, the difference in the parental effect between law and other fields is also statistically significant in most cases.

effects arising from unobservables. Table 7 shows results from two types of mixed logit models: models assuming uncorrelated random coefficients (Columns 1 and 2) and models allowing for correlation across the random coefficients (Columns 3 and 4). The evidence of the distance effect is found to be fairly robust to relaxing the assumption of taste homogeneity: regardless of the specification, the mean coefficient estimate for the shortest distance to enrol is approximately similar to the conditional logit estimate reported earlier in Table 3. Thus, according to the mixed model assuming uncorrelated random coefficients, the implied average effect of a 100 km increase in the shortest distance to enrol is -22% (short specification) or -21% (long specification). The estimated standard deviations for the distance effect are approximately zero and statistically significant, indicating that the effect is homogeneous across observably similar individuals. As for the parent's field indicator, the mixed logit results suggest that although the mean effect of this variable is positive, significant variation in the effect exists even among observably similar individuals. The implied mean effect of this variable is somewhat smaller than in the conditional logit results (Table 3): after controlling for individual heterogeneity, having a parent graduated from a field is estimated to increase the odds of choosing that field by an average of 76% (short specification) or 79% (long specification). However, given that the standard deviation estimates for the parent's field indicator are large and significant, the results also suggest that for a sizeable proportion of individuals (20–24% depending on specification), the effect of this variable is negative.²¹

7 Conclusions

A vast literature has studied determinants of educational choices. This paper contributes to this literature by showing evidence of distance deterrence and intergenerational similarity in

²¹ Discussion of the results from models with correlated random coefficients are to be added later on in this section.

the choice of university field of study. The estimates obtained for a sample of Finnish university students suggested that an increase in the shortest distance to enrol in a field is significantly negatively associated with the likelihood of choosing that field; in previous studies, a similar distance effect has been found in the context of postsecondary institution choice (e.g., Ordovensky, 1995; Long, 2004; Sá et al., 2004; Alm & Winters, 2009; Jepsen & Montgomery, 2009; Gibbons & Vignoles, 2012). The suggested impact of the shortest distance is sizeable: depending on model specification used, a 100 km increase in the shortest distance to enrol was found to decrease the odds of choosing a field by the average of 14–22%. This evidence of the negative distance effect was found to be robust across several model specifications, and no strong support for individual heterogeneity in the effect was obtained. However, the distance effect was found to be somewhat heterogeneous across fields.

The analysis also provided evidence of substantial intergenerational similarity in field-of-study choice. The estimated parental effect varied to some extent across model specifications, but even according to the most conservative estimate, a student having a parent graduated from a specific field has on average a 1.8 times higher odds of graduating from that field than a similar student without that type of parent. However, the results also suggested that the parental effect is heterogeneous in many ways. In particular, the field in question was found to matter: for example, having a parent with a similar field was found to have a particularly sizeable impact on choosing business or law but very little or no impact on choosing humanities or natural sciences. The parental effect was also found to differ across genders – the estimated effects were larger for men than for women – and even across observably similar individuals: the results from the mixed logit analysis suggested that a significant proportion of individuals in fact place a negative weight on fields chosen by one of their

parents. In addition, the results indicated that individuals are more likely to imitate their father's field-of-study choice than the choice of their mother.

Certain implications for educational policy may be derived based on the obtained results. In particular, the evidence of the distance deterrence effect in field-of-study choice should be accounted for by policy makers in countries similar to Finland: although transforming small regional universities from multidisciplinary institutions into specialized institutions may be beneficial from productivity perspective, regional disparities in access to certain fields may increase as a result of such policy. For example, one might consider the consequences of abolishing the supply of certain fields from the universities located in the Northern Finland. The results of the current analysis suggest that this action would have a significant negative impact on the likelihood of sorting into these fields for high school graduates in the northern regions of Finland.

Acknowledgements

Financial support from the Academy of Finland is gratefully acknowledged (grant number 127049). I also wish to thank Mika Haapanen, Kaisa Kotakorpi, Timo Kuosmanen and participants at the 2009 meeting of the Finnish Economic Society for providing helpful comments.

References

- Alm, J., & Winters, J.V. (2009). Distance and intrastate college student migration. *Economics of Education Review*, 28(6), 728–738.
- Angrist, J.D., & Pischke, J.-S. (2009). *Mostly harmless econometrics*. Princeton University Press, New Jersey.

- Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, 121 (1–2), 343–375.
- 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(1), 334–347.
- Berger, M. (1988). Predicted future earnings and choice of college major, *Industrial and Labour Relations Review*, 41(3), 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: Mancebón-Torrubia, M.J., Ximénez-de-Embún D.P. & Gómez-Sancho, J.M. & Esteban, G.G. (Eds.), *Investigaciones de Economía de la Educación*, 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(1), 79–93.
- Boudarbat, B., & Montmarquette, C. (2007). Choice of fields of study of Canadian university graduates: The role of gender and their parents' education, *IZA Discussion Paper 2552*.
- Bratti, M. (2006). Social class and undergraduate degree subject in the UK, *IZA Discussion Paper 1979*.
- Dryler, H. (1998). Parental models, gender and educational choice, *British Journal of Sociology*, 49(3), 375–398.
- Gibbons, S., & Vignoles, A. (2012). Geography, choice and participation in higher education in England. *Regional Science and Urban Economics*, 42(1), 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(3), 497–538.

- Hole, A.R. (2007). Fitting mixed logit models by using maximum simulated likelihood. *Stata Journal*, 7(3), 388–401.
- 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(1), 64–73.
- Leppel, K. (1993). Logit estimation of a gravity model of the college enrolment decision. *Research in Higher Education*, 34(3), 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(4), 373–394.
- Long, B.T. (2004). How have college decisions changed over time? An application of the conditional logistic model. *Journal of Econometrics*, 121(1–2), 271–296.
- Montmarquette, C., Cannings, K., & Mashedjian, S. (2002). How do young people choose college majors? *Economics of Education Review*, 21(6), 543–556.
- Nicolau, N., & Shane, S. (2010). Entrepreneurship and occupational choice: Genetic and environmental influences. *Journal of Economic Behavior & Organization*, 76(1), 3–14.
- Ordovensky, J. (1995). Effects of institutional attributes on enrollment choice: Implications for postsecondary vocational education. *Economics of Education Review*, 14(4), 335–350.
- Sá, C., R., Florax, & Rietveld, P. (2004). Determinants of the regional demand for higher education in the Netherlands: A gravity model approach. *Regional Studies*, 38(4), 375–392.
- Sjaastad, L. (1962). The costs and returns of human migration, *Journal of Political Economy*, 70(5), 80–93.
- Train, K. (2003). *Discrete choice methods with simulation*, New York: Cambridge University Press.

- Van de Verhorst, H.G., De Graaf, N.D., & Kraaykamp, G. (2001). Intergenerational resemblance in field of study in the Netherlands. *European Sociological Review*, 17(3), 275–293.
- Whiston, S.C., & Keller, B.K. (2004). The influences of the family of origin on career development. *Counseling Psychologist*, 32(4), 493–568.
- 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(1), 123–139.

Appendix

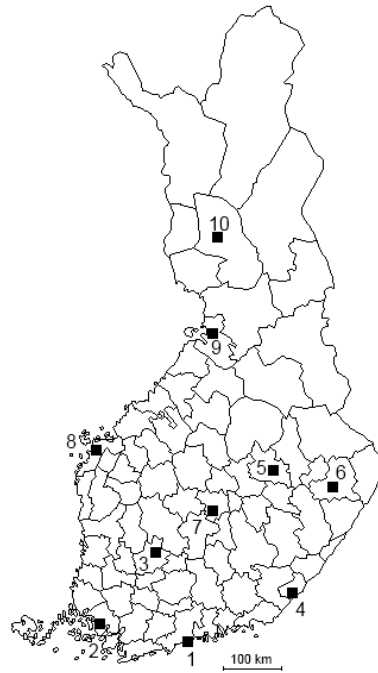
List of variables used in the analysis.

- *Field of study*. The field of the first completed university degree (bachelor's or master's) after graduation from high school. This variable includes 10 categories: education; arts; humanities; business; social sciences; law; natural sciences; technology; medicine/health sciences; other (agriculture/forestry, military sciences and sports sciences).
- *Shortest distance to enrol*. The road distance from the individual's pre-university residential location to the nearest city in which the field of study is available. The distance is measured from the central municipality of the individual's NUTS-4 region one year before high school graduation year.
- *Second shortest distance to enrol*. The road distance from the individual's pre-university residential location to the second nearest city in which the field of study is available.
- *Parent's field*. This variable gets 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 a 0 otherwise.
- *Mother's/father's field*. This variable gets a value of 1 if the individual's mother/father has a university degree (bachelor's, master's or PhD) in the field, and a 0 otherwise.
- *Female*. This variable gets a value of 1 if a woman, and a 0 otherwise.
- *Swedish speaker*. This variable gets a value of 1 if the individual's mother tongue is Swedish, and a 0 otherwise.
- *Mother's/Father's educational 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; secondary-level vocational qualification; lowest tertiary-level qualification; bachelor's degree (including polytechnic degrees); master's degree; doctoral degree.

- *Mother's/Father's vocational status.* The latest observed socioeconomic class of mother/father (observations in 1970, 1980, or 1990). These variables include 5 categories: farmer; entrepreneur; high-ranking official; low-ranking official; worker or no classification available.
- *Matriculation examination grade in first language.* 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.
- *Matriculation examination grade in mathematics.* 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.
- *High school graduation year.* This variable includes 6 categories (years from 1991 to 1996).
- *Type of pre-university region.* The type of the NUTS-4 region in which the individual resided one year before high school graduation year. This variable includes 6 categories; sparsely populated region; rural region; industrial centre; regional centre; large university region; metropolitan region.

Figure 1

Geography of the Finnish university system: cities, institutions and fields of study



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

Notes: NUTS-4 regional classification (from 2003) is used in the map. The institution names and field alternatives reflect the situation for high school graduates from 1991–1996. ‘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

Table 1

Descriptive evidence of the determinants of field-of-study choice: mean distance to enrol and similarity with parents' fields for chosen and non-chosen field-of-study alternatives

Group	Nr. of obs.	Shortest distance to enrol	2nd shortest distance to enrol	Parent's field	Father's field	Mother's field
Chosen field alternatives	3869	75	206	0.106	0.077	0.043
Non-chosen field alternatives	34821	98	228	0.045	0.027	0.023
Education chosen	546	70	180	0.10	0.05	0.06
Education not chosen	3323	58	175	0.07	0.03	0.05
Arts chosen	118	121	274	0.03	0.03	0.01
Arts not chosen	3751	144	283	0.01	0.00	0.00
Humanities chosen	567	68	199	0.10	0.04	0.08
Humanities not chosen	3302	65	194	0.07	0.02	0.05
Business chosen	422	56	188	0.13	0.09	0.05
Business not chosen	3447	69	198	0.04	0.03	0.02
Social sciences chosen	376	57	188	0.09	0.06	0.04
Social sciences not chosen	3493	67	184	0.05	0.03	0.03
Law chosen	160	100	265	0.12	0.11	0.01
Law not chosen	3709	168	319	0.02	0.02	0.01
Natural sciences chosen	455	71	186	0.09	0.07	0.04
Natural sciences not chosen	3414	72	186	0.06	0.04	0.03
Technology chosen	794	84	231	0.12	0.12	0.01
Technology not chosen	3075	95	249	0.05	0.05	0.00
Medicine/health sciences chosen	291	75	198	0.14	0.10	0.08
Medicine/health sciences not chosen	3578	77	191	0.06	0.04	0.03
Other field chosen	140	115	261	0.06	0.06	0.01
Other field not chosen	3729	147	284	0.03	0.03	0.01

Table 2

Differences in individual observables across fields of study: means of selected variables

	Education	Arts	Humanities	Business	Social sciences	Law	Natural sciences	Technology	Medicine/ health sciences	Other field
Female	0.83	0.64	0.81	0.52	0.75	0.55	0.53	0.22	0.71	0.44
Swedish speaker	0.08	0.03	0.06	0.12	0.10	0.07	0.07	0.05	0.08	0.03
First language grade ^a	4.89	5.10	5.44	5.00	5.25	5.27	5.04	4.97	5.20	4.97
A-level math grade ^a	3.68	4.33	4.16	4.50	4.34	4.39	4.84	5.21	4.88	4.27
B-level math grade ^a	3.99	3.94	4.55	4.80	4.36	4.79	4.65	4.85	4.64	4.44
Mother's educational level ^b	3.20	3.80	3.52	3.54	3.41	3.89	3.58	3.74	3.71	3.84
Mother's vocational status										
Farmer/entrepreneur	0.11	0.10	0.11	0.14	0.11	0.11	0.13	0.13	0.16	0.16
High-/low-ranking official	0.71	0.81	0.73	0.74	0.74	0.82	0.74	0.74	0.72	0.73
Worker/other	0.18	0.08	0.16	0.13	0.15	0.08	0.13	0.13	0.12	0.11
Father's educational level ^b	3.27	3.88	3.63	3.99	3.75	4.14	3.82	4.14	3.92	3.86
Father's vocational status										
Farmer/entrepreneur	0.20	0.17	0.17	0.20	0.18	0.17	0.16	0.16	0.21	0.24
High-/low-ranking official	0.51	0.64	0.59	0.63	0.57	0.66	0.62	0.64	0.61	0.55
Worker/other	0.29	0.19	0.25	0.18	0.25	0.17	0.22	0.20	0.18	0.21
Type of pre-university region										
Metropolitan area	0.14	0.35	0.21	0.32	0.25	0.42	0.17	0.25	0.26	0.26
University region	0.24	0.28	0.24	0.23	0.27	0.31	0.28	0.25	0.26	0.16
Other	0.61	0.37	0.55	0.44	0.48	0.27	0.55	0.50	0.47	0.58

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 educational levels take values from 1 (primary school) to 7 (doctoral degree).

Table 3

Relative change in the odds of choosing a field associated with an increase in distance to enrol and similarity with a parent's field: estimates from a conditional logit model

Basic model								
Shortest distance to enrol (/100 km)	-0.203	***	-0.201	***	-0.213	***	-0.212	***
	(0.033)		(0.035)		(0.032)		(0.027)	
Parent's field	1.131	***	1.165	***	0.931	***	0.921	***
	(0.209)		(0.185)		(0.182)		(0.161)	
Model with gender interactions								
Shortest distance to enrol * Male			-0.204	***	-0.214	***	-0.219	***
			(0.041)		(0.042)		(0.042)	
Shortest distance to enrol * Female			-0.199	***	-0.213	***	-0.207	***
			(0.045)		(0.041)		(0.035)	
Parent's field * Male			1.381	***	1.239	*** ^d	1.282	*** ^d
			(0.171)		(0.181)		(0.179)	
Parent's field * Female			1.004	***	0.716	***	0.679	***
			(0.247)		(0.220)		(0.197)	
Controls included:								
Field-specific constants	Yes		Yes		Yes		Yes	
Demographic variables	No		Yes		Yes		Yes	
Matriculation examination grades	No		No		Yes		Yes	
Parental and regional variables	No		No		No		Yes	

Notes: The estimates are to be interpreted as relative changes in the odds of choosing a field caused by one unit increases in the variables. The set of demographic variables includes a woman indicator, a Swedish speaker indicator and high school graduation year. The set of parental and regional variables includes mother's/father's vocational status, mother's/father's educational level and the type of pre-university region. Standard errors (in parentheses) are clustered at the level of 81 high school regions. Significant estimates are indicated by * (p<0.1), ** (p<0.05) or *** (p<0.01). Significant gender differences (p<0.05) are indicated by ^d.

Table 4

Relative change in the odds of choosing a field associated with an increase in distance to enrol and similarity with a parent's field: estimates from a conditional logit model

Basic model				
Shortest distance to enrol (/100 km)	-0.181	***	-0.201	***
	(0.027)		(0.030)	
Second shortest distance to enrol (/100 km)	-0.097	**	-0.040	
	(0.039)		(0.038)	
Mother's field	0.337	***	0.312	***
	(0.147)		(0.129)	
Father's field	1.079	***	1.110	***
	(0.172)		(0.181)	
Model with gender interactions				
Shortest distance to enrol * Male	-0.171	***	-0.192	***
	(0.043)		(0.046)	
Shortest distance to enrol * Female	-0.191	***	-0.209	***
	(0.043)		(0.045)	
Second shortest distance to enrol * Male	-0.134	***	-0.089	
	(0.046)		(0.053)	
Second shortest distance to enrol * Female	-0.066		0.001	
	(0.061)		(0.054)	
Mother's field * Male	0.421	**	0.402	**
	(0.203)		(0.204)	
Mother's field * Female	0.283	*	0.253	*
	(0.171)		(0.145)	
Father's field * Male	1.242	***	1.346	***
	(0.195)		(0.210)	
Father's field * Female	0.926	***	0.906	***
	(0.243)		(0.242)	
Specification	Short		Long	

Notes: The estimates are to be interpreted as relative changes in the odds of choosing a field caused by one unit increases in the variables. The short specification includes a field-specific constant, a female indicator, a Swedish speaker indicator, high school graduation year and matriculation examination grades in first language and math as control variables. The long specification also includes mother's/father's vocational status, mother's/father's educational level and the type of pre-university region. Standard errors (in parentheses) are clustered at the level of 81 high school regions. Significant estimates are indicated by * (p<0.1), ** (p<0.05) or *** (p<0.01). None of the estimated gender differences in the effects are statistically significant.

Table 5

Relative change in the odds of choosing a field associated with an increase in distance to enrol:
estimates from a conditional logit model with nonlinear distance effects

Model I			
Shortest distance to enrol (/100 km)			
Linear term	-0.149 (0.108)	-0.214 (0.086)	**
Quadratic term	-0.025 (0.043)	0.001 (0.036)	
Effect at mean distance (95.4 km)	-0.196 *** (0.039)	-0.212 *** (0.029)	***
Model II			
Shortest distance to enrol (/100 km)			
Linear term	-0.131 (0.113)	-0.204 (0.093)	*
Quadratic term	-0.019 (0.041)	0.002 (0.037)	
Effect at mean distance (95.4 km)	-0.167 *** (0.042)	-0.200 *** (0.036)	***
Second shortest distance to enrol (/100 km)			
Linear term	-0.105 (0.102)	-0.061 (0.106)	
Quadratic term	0.001 (0.013)	0.003 (0.013)	
Effect at mean distance (225.7 km)	-0.098 * (0.053)	-0.047 (0.055)	
Specification	Short	Long	

Notes: The estimates are to be interpreted as relative changes in the odds of choosing a field caused by one unit increases in the variables. The short specification includes a field-specific constant, a female indicator, a Swedish speaker indicator, high school graduation year and matriculation examination grades in first language and math as control variables. The long specification also includes mother's/father's vocational status, mother's/father's educational level and the type of pre-university region. In addition to the individual-level controls, Model I includes a parent's field indicator, whereas Model II includes mother's field and father's field indicators. Standard errors (in parentheses) are clustered at the level of 81 high school regions. Significant estimates are indicated by * (p<0.1), ** (p<0.05) or *** (p<0.01).

Table 6

Relative change in the odds of choosing a field associated with an increase in distance to enrol and similarity with a parent's field: estimates from a conditional logit model allowing for heterogeneous own-field effects

	Shortest distance to enrol (/100 km)				Parent's field		
Education	0.075 (0.100)	-0.161 (0.108)			1.054 *** (0.289)	1.484 *** (0.518)	
Arts	-0.218 ** (0.079)	-0.107 (0.085)			2.424 * (2.464)	1.330 (1.620)	
Humanities	-0.043 (0.076)	-0.190 * (0.102)			0.333 (0.243)	0.102 (0.279)	
Business	-0.245 ** (0.089)	-0.149 (0.110)			1.840 *** (0.351)	1.803 *** (0.392)	
Social sciences	-0.254 *** (0.057)	-0.287 *** (0.081)			0.572 * (0.367)	0.696 ** (0.380)	
Law	-0.337 *** (0.078)	-0.259 *** (0.072)			4.729 *** (1.764)	3.589 *** (1.428)	
Natural sciences	-0.121 (0.078)	-0.257 *** (0.059)			0.289 (0.260)	0.436 (0.320)	
Technology	-0.165 *** (0.049)	-0.164 ** (0.059)			0.863 *** (0.201)	0.912 *** (0.217)	
Medicine/health sciences	-0.139 ** (0.053)	-0.058 (0.100)			1.610 *** (0.707)	1.423 *** (0.633)	
Other field	-0.261 *** (0.068)	-0.351 *** (0.094)			0.496 (0.701)	0.423 (0.540)	
Mean effect	-0.136 *** (0.034)	-0.190 *** (0.039)			1.073 *** (0.112)	1.038 *** (0.117)	
Specification	Short	Long			Short	Long	

Notes: The estimates are to be interpreted as relative changes in the odds of choosing a field (specified in the left column) caused by one unit increases in the variables. The mean effect is calculated as the weighted average of the field-specific effects (field sizes are used as weights). The short specification includes a field-specific constant, a female indicator, a Swedish speaker indicator, high school graduation year and matriculation examination grades in first language and math as control variables. The long specification also includes mother's/father's vocational status, mother's/father's educational level and the type of pre-university region. Standard errors (in parentheses) are clustered at the level of 81 high school regions. Significant estimates are indicated by * (p<0.1), ** (p<0.05) or *** (p<0.01).

Table 7

Relative change in the odds of choosing a field associated with an increase in distance to enrol and similarity with a parent's field: estimates from a mixed logit model allowing for taste heterogeneity

Shortest distance to enrol (/100 km)							
Mean coefficient	-0.242	***	-0.240	***	-0.242	***	t.b.a.
	(0.033)		(0.037)		(0.033)		
Standard deviation	0.002		0.004		0.056		t.b.a.
	(0.152)		(0.189)		(0.117)		
Exp(mean coef.)-1	-0.215	***	-0.214	***	-0.215	***	t.b.a.
	(0.026)		(0.029)		(0.026)		
Pr(coef. < 0)	1.000		1.000		1.000		t.b.a.
Parent's field							
Mean coefficient	0.568	***	0.582	***	0.560	***	t.b.a.
	(0.102)		(0.108)		(0.100)		
Standard deviation	0.791	**	0.698	*	0.819	***	t.b.a.
	(0.322)		(0.387)		(0.298)		
Exp(mean coef.)-1	0.764	***	0.789	***	0.751	***	t.b.a.
	(0.180)		(0.193)		(0.175)		
Pr(coef. > 0)	0.764		0.798		0.753		t.b.a.
Covariance between random coefficients allowed for							
	No		No		Yes		Yes
Covariance estimate							
					0.046		t.b.a.
					(0.097)		
Specification							
	Short		Long		Short		Long

Notes: The random coefficients are assumed to be normally distributed. The short specification includes a field-specific constant, a female indicator, a Swedish speaker indicator, high school graduation year and matriculation examination grades in first language and math as control variables. The long specification also includes mother's/father's vocational status, mother's/father's educational level and the type of pre-university region. Significant estimates are indicated by * (p<0.1), ** (p<0.05) or *** (p<0.01).