Regional Supply of Education and Student Sorting

Hanna Virtanen*, Lotta Väänänen**

* *Aalto University School of Economics, HECER, FDPE*

** *University of Mannheim, Department of Economics*

This version: 27.1.2011
(Very preliminary, please do not quote)

ABSTRACT: We examine how the regional supply of education affects the schooling choices of young people. We use rich individual-level data on 4 cohorts of Finns leaving comprehensive school in 2000-2003 to study the sorting of individuals across 8 schooling alternatives (gymnasium, 7 vocational fields) and the option of not to study. We use discrete choice models with random coefficients, allowing for heterogeneous preferences. We find that the distance to the nearest educational institute offering an alternative has a statistically significant effect on the choice of a schooling alternative. The estimated effect is larger for boys than girls. The demand estimates can be used to investigate various policy counterfactuals (work in progress). Under the counterfactual scenario of no supply constraints, we find that a large fraction of individuals would move away from the choice of gymnasium to vocational fields. Quantitatively, we can also see some interesting results when comparing girls and boys, those from low- and high-income families and individuals with low and high gpas.

Keywords: mixed logit; random coefficients; schooling choices; supply of education; upper secondary education; vocational education,

Corresponding author: Hanna Virtanen, hanna.virtanen@hse.fi, Aalto University, School of Economics, P.O.Box 21240, FI-00076 AALTO.

We would like to thank Pekka Ilmakunnas, Tuomas Pekkarinen and Otto Toivanen for comments. We also thank Satu Nurmi at Statistics Finland for her work in constructing the data. We would like to thank the Yrjö Jahnsson Foundation for financial support. Väänänen also gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15. The usual caveat applies.

1 Introduction

Education, not only the level but also the field, is an important determinant of the labour market outcomes and career opportunities of individuals. There are substantial differences in the earnings and unemployment risks between fields of study given the education level (see e.g. Arcidiacono, 2004). Particularly important is the decision of the field of the upper secondary education, because this choice is a crucial step in defining the later schooling opportunities and the direction of the career. For the society, this has implications for the labour supply of different kinds of workers. Understanding how individuals sort to different fields is therefore important, in particular with respect to factors that can be affected by policy measures. A key aspect of educational policy is the regional supply of education. Our goal is to estimate how the supply of education affects the demand for the schooling alternatives and the sorting of individuals across the alternatives. The demand estimates can be used to investigate various policy counterfactuals: how changing different aspects of the supply of education affects the sorting of individuals across schooling alternatives

Literature on schooling choices has examined how individuals sort across different fields of education, majors and schooling tracks (academic/vocational). The focus has been on the effects of factors such as individuals' characteristics (Dustmann, 2004; Boudarbat & Montmarquette, 2007), expected earnings and risk related to it (Arcidiacono 2004; Beffy et. al., 2009; Berger, 1988; Saks & Shore, 2005), and option value (Eide & Waehrer, 1998). Supply side factors, such as tuition fees and distance-related costs, have also been shown to play a role in determining the participation in higher education (e.g. Gibbons & Vignoles, 2009; Frenette, 2004; 2006; Spiess & Wrohlich, 2008)[1] as well as the choice of what and where to study (Long 2004; Kelchtermans and Verboven 2010a, 2010b). The focus of our paper is to examine the effects of the availability of schooling on the sorting of individuals across schooling alternatives and fields.

If the supply of education has an impact on the schooling decisions, then possible regional disparities in the availability of education can introduce inequalities in human capital accumulation between individuals living in different regions. Furthermore, it may boost differences in skill composition between regions. A poor availability of education may also be linked to social exclusion, as participation in upper-secondary schooling is considered to be crucial to the later success of the individuals. Availability of schooling may also have larger effects on some sub-groups of individuals (low income families etc.) and thus, a deficiency in the local supply of education may further increase the inequality of these groups.

In this study we use a rich cross section data of four complete cohorts of Finns leaving compulsory schooling to examine how the supply of education affects the demand for the schooling alternatives and the sorting of individuals across the alternatives, namely gymnasium, 7 fields of vocational education and

[1] Supply characteristics are also often used as instruments for the schooling decision (to study the returns to education or other educational outcomes), as they are perceived to be uncorrelated with the studied outcomes of the education (see Card 2001 for a survey).

the outside option of not to study. We use a conditional logit model to model the discrete choice among the alternatives, and use a random coefficients specification to allow for heterogeneous effects of the supply variables. Our data allows us to thoroughly analyze the individuals' application decisions, taking into account all the schooling alternatives available and their features, while controlling for detailed family background information, school performance, characteristics of the comprehensive schools, and regional characteristics. The detailed data together with the use of a mixed logit model permits a flexible specification of demand that allows for random taste variation and unrestricted substitution patterns. These features help to provide good demand predictions and meet the ultimate goal of performing counterfactual simulations of demand under different supply environments.

The large regional differences in the supply of upper secondary education in Finland[2] provide a good opportunity to analyse the sensitivity of schooling decisions on the availability of schooling alternatives. We measure the supply of schooling using the distance to the nearest municipality offering each alternative and the number of open positions in each alternative per size of the cohort in a region (work in progress). Distance to schooling (the regional availability) can affect the costs of schooling in several ways: direct financial costs of re-allocation or commuting, emotional costs associated with leaving home, and information costs when seeking information on the schooling options. Besides the distances to the schooling alternatives, the availability of open schooling positions in each alternative in a region affects the entrance probabilities and thus influences the kind of individuals who sort to the different fields.

Our results show that the availability and proximity of a schooling alternative has a positive and statistically significant role in influencing the decision to choose that alternative. Furthermore, the cross-effects of the proximity of other fields are negative. For boys, the estimated effect is larger than for girls, and we also find significant heterogeneity of the effect (i.e. significant standard deviation of the random coefficient). Our demand estimates allow us to investigate various policy counterfactuals. Under the counterfactual scenario of no supply constraints, we find that a large fraction of individuals would move away from the choice of gymnasium to vocational fields. In the lower income groups, and for those with lower gpas, we find a larger share of the students moving away from gymnasium than in the high income or high gpa groups. There are also differences in the fields they would change to. Preliminary results indicate that individuals are less sensitive to the supply concerning their participation decision than concerning their choice of field. This result is in line with prior studies (see eg. Kelchtermans and Verboven 2010a).

The rest of the paper is organized as follows. We begin with a brief introduction of the upper secondary education system in Finland. In Section 3 we describe the data in detail and present some descriptive analysis. In section 4 we present the empirical analysis and the results, and in section 5 the conclusions.

[2] In its recent definition of policy, the Finnish Committee for Education and Culture stated that the target of the education policy in Finland for the 2012 is to diminish the regional differences in the supply of education (the number of open positions per cohort).

2 The upper secondary education system in Finland

Compulsory education in Finland consists of nine years of comprehensive school and it typically ends at the age of 16. There is also a possibility to continue for one additional year (10th grade) in comprehensive schooling after finishing the 9th grade. After completing comprehensive schooling, one can continue to upper secondary education or decide not to study further. One of the objectives of education policy is to provide upper secondary education to all of each age group free of charge and it is strongly recommended that young people continue studying after compulsory schooling. Upper secondary education is divided into general (i.e. gymnasium) and vocational education. Vocational education and training includes 7 fields of education, which again contain 30 subfields (see appendix A). The scope of the syllabus in upper secondary education is three years[3].

Upper secondary education is provided by local authorities, municipal consortia or other organisations authorised by the Ministry of Education. According to the statistics of educational institutions by Statistics Finland, there were about 440 gymnasiums and around 200 vocational education organisations in the year 2004[4]. The licence to provide upper secondary education defines the maximum number of students an education organisation is allowed to have per schooling alternative. Schooling alternative refers to the seven vocational fields and gymnasium.

Admission to the upper secondary education takes place through the joint application system maintained by the Finnish National Board of Education. Most of the positions in the secondary education are filled through this system[5]. Individuals can simultaneously apply to five different schooling positions. A schooling position is defined as an education institution- schooling alternative combination. Thus, students can apply to the same schooling alternative in five different education institutions or to five different schooling alternatives in the same education institution, for instance. Individuals rank their requests and receive one offer at a time. If a person does not gain admittance to the school of his or her first choice, the other requests are considered in their order of ranking. The Ministry of Education determines the general student selection criteria, but an education provider can decide on the specifics of the student admission criteria (e.g. use of entrance or aptitude tests). For those positions that do not get filled at the first application, a replacement application process is organized.

3 Data and descriptive analysis

3.1 Data sources and sample

We use a very detailed individual level dataset on four full cohorts of individuals leaving the compulsory school during the years 2000-2003. The dataset is constructed from various registers at

[3] It is possible to attain the matriculation and a vocational qualification simultaneously. This usually means that the studies take 4 years.

[4] The number of education organizations is a little higher during our observation period as the amalgamation of education organizations has been ongoing trend in upper secondary education during the last decade.

[5] There are some types and fields of education, which do not use the joint application system (e.g. the smaller scale vocational qualifications, vocational qualifications in specialized fields such as music and dance)

Statistics Finland. The sample is based on the Application Register of the Finnish National Board of Education and it includes about 250,000 individuals. This contains all the individuals who are in the 9th grade in the spring of one of the examined years and obtain a leaving certificate during that year[6]. We exclude from our sample those individuals who continue in the comprehensive schooling in the following year. All the descriptive results in this document are done so far only with a 30 percent random sub-sample of the total data.

The Application Register contains information on the application requests as well as information on the individual characteristics. In addition, we include in the dataset detailed information on the parents from the Finnish Longitudinal Employer-Employee data (FLEED) of Statistics Finland. The FLEED is a register-based dataset that contains information on all Finnish individuals aged 15-70 years, including their income, socioeconomic status and education. From the FLEED we also calculate some characteristics of the alternatives, such as average income and standard deviation of income, unemployment rate, and the proportion of individuals with higher education.

We complement the dataset with information on the supply of education and regional characteristics from several data sources. The supply of education is determined from the online database WERA of the National Board of Education. In addition, we obtain a matrix of inter-municipality driving distances from the Finnish Road Administration. Furthermore, the regional characteristics are gathered from online databases of Population Register Centre, National Land Survey of Finland, and Association of Finnish Local and Regional Authorities.

3.2 Variables

3.2.1 Schooling choices

We determine the outcome variable, namely the schooling choice of the individual, based on the schooling alternative which is assigned to as first request in the joint application. From the Application Register we observe the subfield of education of the first application request. We aggregate this information to educational field level. After this we have 8 schooling alternatives from which individuals make their choice: gymnasium and 7 vocational fields[7], which include "Natural Resources", "Technology and Transport", "Administration and Commerce", "Hotel, Catering and Home Economics", "Social and Health Care Services", "Culture", and "Humanist and Teaching".

The first request is taken as the individual's optimal schooling alternative given his or her information set (expectations of the success in admission). The enrolment is, instead, determined mutually by the application and admission processes. Thus, we are able to examine how supply of upper secondary education affects the demand of the schooling alternatives.

We define the outside option as the choice of not applying to any upper secondary schooling alternative (through the joint application). There are, however, several issues that may cause

[6] We do not include individuals who are either resident or studying in Åland. We do not observe the inter-municipality distances for these observations.

[7] We use classification of vocational fields in year 1995.

"mismeasurement" here. Ideally, we would like to capture those individuals who do not apply to any schooling alternative and who decide to remain outside the upper secondary schooling system. Therefore, it is not clear how to treat those who decide to enter the 10th grade (or repeat the 9th grade), or those who end up enrolling in some upper secondary although they do not apply (through the joint application). We remove these individuals from the sample, but also run robustness checks including them.

3.2.2 Supply and other alternative specific characteristics

The variables we use to describe the regional supply of education are: i) a dummy variable equal to one if the field of study is offered in the municipality of residence; ii) the distance to the nearest municipality offering the field of study (in km), iii) the number of open positions in the field per cohort in the sub-region (work-in-progress).

We use the classification of municipalities in year 2010 which contains 326 municipalities each year. We determine the supply of education in these municipalities from the WERA data of the National Board of Education. The data has information on the annual number of open schooling positions announced by the education organisations[8]. The open schooling positions are given at the level of education organisation, municipality and education subfield. We aggregate this information to level of education field and municipality. We define the set of open schooling position to include those positions which have the requirement of basic training only in comprehensive schooling, are for young persons and are aimed at completing a degree.

The individual's municipality of residence comes from Application Register and is determined at the time of the application (during the last semester of the 9th grade). We measure the distance from the municipality of residence to the nearest municipality offering each alternative and use this to calculate the minimum distance to each of the alternatives for all the individuals in the sample. If the schooling alternative is offered in the municipality, we define the distance to be 2km[9].

The distance between the municipality of education and the municipality of residence is an approximate measure of distance to schooling as it gives the same distance to all individuals living in the same municipality and to all education provided within the same municipality[10]. However, the distance within a municipality (which we cannot measure) probably has less of an effect than the distance between municipalities, because transportation services are usually better within municipalities as the municipalities are active in planning, financing and organising the public transportation in the area. Pupils in comprehensive school are also likely to be better informed about the schooling possibilities in their municipality of residence. Furthermore, the distances between municipalities are defined as the driving

[8] They do not include information on the open positions in apprenticeship training or in education not administrated by the National Board of Education.

[9] The observed minimum of distances between municipalities is 3km.

[10] Some previous studies have utilised the zip codes of educational institutions (universities) and home to define the distance to schooling. Frenette (2004) use straight line distances to create an indicator of whether the education institute is "nearby". Gibbons and Vignoles (2009) define the distances using rail-network distances. Spiess and Wrohlich (2008) use the geographical information system ArcGIS to calculate the distances between the centres of gravity of the zip codes. There may be some measurement error in these as well, because the distances may be of different effect depending e.g. on the means of communication and transport services.

distances and thus provide good proxies for the effective distances between municipalities. To diminish the possible measurement error, we use a control for the size of the municipality in square kilometres from the National Land Survey of Finland.

There are other alternative specific characteristics for which individuals have (different) preferences and which effect the application decisions. For example, individuals may prefer higher expected income, less risk in the employment and income streams and better prospects for higher education. We have calculated information on the characteristics of the schooling alternatives from the FLEED data. By using the information on all the 40-45 years old Finns in year 2003, we calculate the average wages, the standard deviation of wages, the unemployment rates and the proportion of individuals with higher education for each schooling alternative. We do the calculations separately for each sub-region and for men and women. The regional differences in the characteristics of the schooling alternatives may affect the individuals' expectations. For one, their perceptions of the expected labour market prospects of each alternative may be based on what they observe in their region. Second, these prospects may be the relevant expectations for individuals with low regional mobility. In addition, the outcomes of the schooling alternatives differ by gender.

3.2.3 Individual and regional controls

The regions do not differ only in their supply, but also in other characteristics which may lead to the selection of different families into different areas. These characteristics can be correlated with the supply of education. This would cause correlation between the regional supply of education and individuals schooling choices, and makes it important to control for other characteristics of the region as well as the individuals' parents.

To account for regional factors, we control for the unemployment rate of young people, the municipal tax revenues, the population density and the proportion of immigrants. Furthermore, we include the proportion of jobs in manufacturing and services, the proportion of jobs in the private sector, and the inward and outward mobility of the municipality. In addition, we include a dummy variable for each province.

We have detailed information on the individuals and their parents to control for characteristics that may be related to their preferences and affect their choices. We have personal information on the date of birth, sex, nationality and native language. We also have information on the level and field of education, the income and the socioeconomic status of the parents. Finally, we have information on the comprehensive school the individuals attended and their grades in comprehensive school. We also observe the size of the comprehensive schools as well as the decisions of the school mates.

3.3 Descriptive analysis

Our data shows that substantial differences exist in the availability of the various schooling alternatives between the regions. Table 1a shows the municipality-level descriptive statistics on the availability of each alternative. Columns 1-3 show the mean, maximum and between standard deviation

of distances to the nearest municipality offering the alternatives in year 2000. The fifth column shows the fraction of municipalities offering each alternative in the same year. We see that gymnasium is most prevalent, present in 77% of the municipalities and the average distance to the nearest municipality offering gymnasium is about 7km. However, some municipalities face longer distances to gymnasium, up to a maximum of 43km. Vocational education is more scarcely available and there are also large differences between fields. 50% of the municipalities offer at least one vocational education. Of vocational fields, the Technology, Communication and Transport sector, and the Tourism, Catering and Domestic Services sector are most widespread, offered in about 35% of the municipalities, whereas the Humanities and Education sector is present only in about 6% of municipalities. With the exception of Tourism, Catering and Domestic Services sector and Social Services, Health and Sports sector, the distances to the fields of vocational education can be several hundreds kilometres, up to 570km maximum in Humanities and Education.

[Table 1a here]

Columns 4, 6 and 7 in Table 1a show the degree of within-municipality variation in the availability of the alternatives during our observation period 2000-2003. There is only little variation in the presence of schooling alternatives, with more than 10km within standard deviation only in the sectors of Natural Sciences, Humanities and Education, and Social Services, Health and Sports. In majority of the vocational fields we can see a tendency for a reduction in supply, with more municipalities ceasing to provide the educational fields than establishing new. In the fields of Culture, Natural Sciences and Technology, Communication and Transport ceasing and establishing new schooling positions are almost as common.

There are also some differences in the overall supply of education when we look at larger regions in Finland. Table 1b shows the mean, maximum, and the standard deviations of the number of open schooling positions per cohort size for each schooling alternative for the 74 sub-regions in Finland in year 2000.

[Table 1b here]

Table 2 presents the distribution of the schooling choices in our sample (for boys and girls separately). First of all, we can see that there are clear differences in the schooling choices by gender. Gymnasium is the most popular choice for both boys and girls, particularly for girls. Of the vocational fields, Technology and Transport (35%) is the most common for boys, whereas for girls it is Social and Health Care Services (11%). Humanist and Teaching is the least common choice at the upper secondary education level. Also both Culture and Natural Resources are rarely chosen by either gender. To take a first look at the possible correlation between the supply of schooling alternatives and individuals' choices, we also present these choices conditional on whether the alternative is offered in the municipality or not and report the significance of the mean comparison t-tests. Many of the differences, in particular for boys, indicate that individuals are more likely to choose an alternative when it is offered in their municipality. For both boys and girls, the probability of choosing gymnasium is significantly higher when it is offered

in their municipality. For boys, the other significant differences are for "Natural Resources", "Admin and Commerce", "Hotel, Catering and Home economics", "Humanist and Teaching" and "Culture". For girls, the field of "Humanist and Teaching" shows a significant difference in the expected direction. There are also some significant differences in the opposite direction. For boys, in the field of "Technology and Transport" the difference in distance goes in the opposite (unexpected) direction and is significant. The same is true for girls in the fields of "Hotel, Catering and Home economics" and "Social and Health care". For both boys and girls, the differences that go in the counterintuitive direction concern alternatives that are the most popular vocational fields among them. These descriptive statistics indicate that while it is possible that the regional supply of schooling alternatives plays a role in determining individuals' schooling choices, there are likely to be other factors related to the region that correlate with choices and thus confound the effect of the supply when not controlled for.

[Table 2 here]

In Table 3 we present the schooling choices conditional on parents (mother and father separately) having (not having) a degree from the particular field. In practically all the cases, individuals, in particular boys, are significantly more likely to choose a given alternative when the parents' education is in that field. For both sexes, the intergenerational correlation is particularly strong for the choice of gymnasium and for natural resources, but is stronger for boys. Boys are four times as likely to choose natural resources if their mother or father has it than if their parent does not; respectively twice as likely to choose gymnasium. Overall, girls have weaker, statistically insignificant correlations of choices with their father's vocational fields of education, with the exception of natural resources.

[Table 3 here]

In Table 4 we present descriptive statistics conditional on the schooling alternative chosen to provide a comparison in terms of the characteristics of the individuals who make the choice and their family background. We also present the significance (and direction) of the mean comparison tests between those who make the choice and those who choose something else. First of all we can see that the fields of "Technology and Transport", "Administration and Commerce" and "Natural Resources" are clearly male-dominated, whereas the "Social and Health Care Services", "Humanist and Teaching", "Hotel, Catering and Home Economics" and "Culture", as well as gymnasium, are female-dominated schooling alternatives. Overall, what stands out from this table is that individuals who choose gymnasium are significantly different from those who choose a vocational field. They have significantly higher grades in comprehensive school, and they come from families with higher incomes, and higher educational level and socioeconomic status of both mother and father. Of the vocational fields, "Technology and Transport" stands out the choice for individuals with the lowest average grade in theoretical subjects, indicating that it is an important schooling option of the academically weaker students. Individuals choosing "Natural Resources" tend to come from families with self-employed parents, who also have lower incomes on average. In addition, the parents of the individuals enrolled in

"Administration and Commerce" have higher income when compared to individuals in other fields of vocational education.

[Table 4 here]

In Table 5, we present some characteristics of the region (municipality) where the individuals live conditional on their schooling choice, as well as some alternative-specific characteristics. Some regional differences are apparent. For example, individuals who choose "Natural Resources" come from municipalities with a lower population density, lower tax revenues, higher share of agricultural jobs, and also a higher level of youth unemployment. On the other hand, those who choose gymnasium or "Admin and Commerce" come from more urban areas, i.e. areas with a higher population density, higher tax revenues and a higher proportion of service sector jobs.

[Table 5 here]

Also in Table 5, we see that there are clear differences in the alternative specific characteristics as well as between men and women. Those who have a gymnasium diploma have a higher mean income and a lower unemployment rate. Men have on average a higher income than women in all the schooling choices, and their unemployment rate tends to be lower.

4 Empirical Analysis

Our goal is to estimate a model of demand for upper-secondary schooling alternatives from the application behaviour of young people who have just finished their compulsory education. We model the schooling decisions of individuals as a utility-maximizing discrete choice among the nine alternatives they face: gymnasium and 7 fields of vocational education, and the outside option of not to study. We specify the choice as a function of the characteristics of the alternatives as well as the characteristics of the individuals. We employ the mixed logit model in order to allow for random taste variation (differences in tastes that are not related to observable factors) and flexible substitution patterns, removing the independence of irrelevant alternatives (IIA) property inherent in simple logit. As the ultimate goal is to perform counterfactual simulations of demand under different supply environments, these features are particularly important. We discuss them in more detail below (see eg. Train, 2009, for a detailed coverage of discrete choice models).

To illustrate the benefits of using the mixed logit (limitations of simple logit), consider a general specification of the utility a person derives from each possible choice, by decomposing utility into a part that is known by the researcher up to some parameters, and an unknown part that is treated as random. Different assumptions about the distribution of the unobserved part of utility lead to different discrete choice models with different properties. In the logit model, the unknown part is specified as an iid extreme value distribution. This independence means that the unobserved portion of utility for one alternative is unrelated to the unobserved portion of utility for another alternative. This gives rise to the IIA property and restrictive substitution patterns. However, there are often factors unobservable to the econometrician, and the unobserved factors related to one alternative may be similar to those of another

alternative, inducing correlation between the error terms and thus violating the assumption of independence. In the mixed logit model the unobserved part of utility is specified as a vector of random terms (random coefficients or error components), in addition to the iid extreme value part. This allows for correlation between the unobserved part of utility across alternatives.

One key advantage of the mixed logit specification is the way it allows for taste variation. The value that an individual places on an attribute of an alternative varies over individuals. To the extent that these differences vary systematically with observable characteristics of the individuals, they enter in the model as interactions and can be included in simple logit. However, tastes can vary in ways that are not related to observable characteristics (people are different in unobservable ways, individual preferences). The mixed logit model can accommodate this sort of random taste variation. Simple logit may still work as an approximation (we use it for comparison). However, the simple logit model does not provide information on the distribution of tastes around the average, which is important for counterfactual simulations.

Another limitation of the logit model concerns substitution patterns. The logit model implies a certain pattern of substitution across alternatives, which is fairly restrictive. There are two ways to illustrate this limitation: as a restriction on the ratios of probabilities and as a restriction on the cross-elasticities of probabilities. The ratio of choice probabilities exhibits the property of the independence of irrelevant alternatives. That is, the relative odds of choosing i over k are the same no matter what other alternatives are available or what the attributes of the other alternatives are. (The ratio is independent of any other alternatives). As an example, consider having two alternatives, gymnasium and a technical vocational field, with the choice probability of each being one half (and thus their ratio equal to 1). Then consider adding another vocational field. With IIA, this would mean that the new choice probabilities are one third for each alternative (keeping the ratio of gymnasium and technical vocational at 1). In reality, the two vocational alternatives are more similar to each other (closer substitutes). Thus one would expect the addition of a new vocational field to reduce the probability of technical vocational by a greater proportion than the probability of gymnasium, therefore changing the ratio of their choice probabilities (and violating IIA). In the mixed logit model, The ratio of choice probabilities (of alternative i and j) depends on all the data, including characteristics of alternatives other than i and j.

Another way to see the limitation of the simple logit is from the cross-elasticities. Changing an attribute of one alternative would have the same percentage effect on the choice probabilities of all the other alternatives (i.e. The cross elasticity of a change in the attribute of alternative j on the probability of choosing alternative i is independent of i.). For example, decreasing the distance to the nearest technical vocational institution would draw applicants from all the other alternatives in the same proportion. This seems unrealistic, as it would most likely draw from the more similar alternatives (vocational rather than gymnasium; boys rather than girls). In the mixed logit model, the cross-elasticities (i.e. substitution patterns) depend on the specification of the variables and the mixing distribution.

The flexibility of the mixed logit is important for the policy counterfactual analysis. If a certain schooling alternative is removed from a municipality (distances to the nearest schooling alternative increase in some regions), the policymaker is interested in knowing how the individuals who would have chosen that alternative are will choose (to which schooling alternative they substitute).

4.1 Specification

We specify the choice as a function of the characteristics of the alternatives as well as the characteristics of the individuals. We also include an alternative-specific fixed effect (for unobservable characteristics). The conditional logit model specifies that the probability that individual i chooses alternative j conditional on alternative-specific regressors x_{ij} and case-specific regressors z_i is

$$p_{ij} = \frac{\exp(x'_{ij}\beta + z'_i\gamma_j)}{\sum_{l=1}^{m}\exp(x'_{il}\beta + z'_i\gamma_l)}, j = 1,..m$$

The alternative-specific variables include a variable describing the supply of that alternative and two dummy variables indicating whether the mother and the father have a degree from that field. We use the following variables to describe the supply of upper-secondary education: One, we measure the distance to the closest municipality offering each alternative. Alternatively (unreported), we include a dummy variable indicating the presence of each type of education in the individual's municipality of residence. Finally, we include a variable describing the availability of schooling positions for new students for each schooling alternative in the sub-region, i.e. the number of open schooling positions relative to the size of the cohort. This measure affects the entrance probabilities into the alternatives, and we interact this variable with the individual's gpa as the effect should vary depending on the gpa.

The estimated effects of the supply factors are based on the assumption that the supply of education is exogenous to the individuals' choices. One way in which there may be a possible selection problem is the selection of parents to municipalities based on their occupation. We are, however, able to control for the parents' income, socioeconomic status and their education (the field and the level of their education) and thus significantly diminish the problem.

We include a number of variables that describe the individual. First, to examine (and control for) how previous school performance affects choices we use the grade point average in comprehensive school as a measure of academic ability. Here we encounter the problem that the gpa is observed only for those individuals who apply through the joint application, and not for those who choose the outside option. We overcome this issue by calculating a predicted gpa for those individuals, based on an OLS estimation of gpa conditional on all the observable characteristics of the individual, her parents, the comprehensive school, and the region. The results of this estimation are presented in the Appendix. It is possible that individuals who choose the outside option have lower gpas than others with the same observable characteristics, but luckily we are also able to control for this to some extent. For some of the individuals who do not apply to schooling in the year of their graduation from comprehensive school, we are able to observe their gpa when they apply in later years. By including these individuals in the OLS

estimation, and by including a dummy variable to indicate that they choose the outside option, we can control for the difference in gpa of individuals who apply and those who do not. The results show that the coefficient on this dummy is negative and statistically significant.

Second, we have variables describing family background. We include the level of education and the socioeconomic status of both the mother and the father and the family income. Third, we have variables describing the region where the individual lives. We include the youth unemployment rate in the municipality, the municipal tax revenues, the population density and the proportion of immigrants, the proportion of jobs in manufacturing and services, the proportion of jobs in the private sector, and the inward and outward mobility of the municipality. Finally, we include the nationality and native language of the individual. We also include the size of the comprehensive school the individual comes from.

In the mixed logit specification we allow individual specific coefficients for the supply variables as well as for the alternative specific constant[11].

We perform our estimations separately for boys and girls. We also divide the boys and girls into 3 smaller subsamples (each) based on the parents' income (an exogenous variable), in order to reduce the memory requirements and computational time when calculating predictions. This has the additional benefit of allowing the determinants of schooling choices to differ for those from low- and high-income families. Our samples then are boys (girls) whose parents' income is above median, below median, and a third subsample for which we do not observe family income. For this last subsample, we have to alter the specification so that it does not include parents' income as an explanatory variable.

4.2 Results[12]

Table 6 shows the estimated coefficients for the supply of education (measured by the distance) from the alternative-specific conditional logit and mixed logit specifications. We can see that distance to the nearest municipality offering each field has a negative significant effect on the decision to choose that alternative for both boys and girls. The estimated effects are larger for boys. The results from the mixed logit specification provide evidence that there is some heterogeneity in the effects of supply for boys, but not for girls.

[Table 6 here]

In Table 7 we report the marginal effects of increasing the distance to a given alternative by 100 kilometres from the alternative-specific conditional logit. As expected the own-effects are negative and the cross-effects positive, and mostly statistically significant. These marginal effects produced by the alternative-specific conditional logit reflect the IIA property, which restricts the cross-elasticities to be the same across alternatives (i.e. note that the proportional changes in the demand for each alternative, induced by a change in the distance of a given alternative, are the same). To provide more interesting (and realistic) insights into the substitution patterns between alternatives, we need the marginal effects

[11] We have also explored random coefficient for the parents' field of education. No significant hetrogeneity was visible and the main results remain.

[12] Results in tables 6-8 are not updated to reflect the latest specification i.e. the specification used in the counterfactual analysis.

from the more flexible mixed logit specification. However, due to the computational time required to calculate these marginal effects, we are unable to provide them in the current version of the draft.

[Table 7 here]

In previous literature it has been found that parental background characteristics and previous school performance are important factors in determining the schooling decisions. We are able to examine in detail the influence of various parental characteristics as well as the individual's school performance in comprehensive school. We find similar results, but our detailed analysis shows interesting differences in how these factors affect the choice of various alternatives, in particular that different factors are relevant for different alternatives. Table 8 shows the estimates of the marginal effects at the mean from the alternative specific conditional logit estimations for boys (for girls unreported).

[Table 8 here]

From Table 8, we can see that the choice among alternatives is positively affected by having a parent with the same field of education. One exception is that when the mother's education is from the field of "Culture", the son is less likely to choose that field. Similarly, the cross-effects of parents' fields of education (i.e. the effects on the probability of choosing a field that is different from the parents' field) are as expected, mostly negative and significant. The effect of father's education is stronger than that of mother's. For girls, the mother's field of education has no statistically significant effects and the father's field has fewer significant effects (of the expected sign) than for boys.

Having a mother with a degree higher than upper secondary education increases the boys' probability to apply to gymnasium and decreases their probability of choosing "Technology and Transport" and "Administration and Commerce". Father's level of education has the same effect with similar magnitudes. Furthermore, the father's higher level of education also decreases boys' probability to choose "Hotel, Catering and Home Economics". The estimated effects are again smaller for girls. However, the negative effect of both mother's and father's higher education on choosing "Social and Health Care Services" is statistically significant for girls.

In terms of the socioeconomic status of the parents (conditional on their level and field of education), the main finding that emerges is that the sons of self-employed parents are more likely to choose natural resources. For girls, this result is not statistically significant. For girls, having a mother whose socioeconomic status is "Upper level employee" increases the probability of choosing gymnasium.

Concerning previous school performance, we find that higher average grades in comprehensive school are positively correlated with choosing gymnasium (statistically significant only for boys). Negative correlations between school performance and choices emerge for all vocational fields (for both boys and girls). In particular, the negative effect is large for the field of "Technology and transport" for boys and for the field of "Social and Health Care Services" for girls.

4.3 Counterfactuals

For the counterfactual analysis, we use the mixed logit specification to obtain predicted choice probabilities under the actual supply environment and then under some counterfactual scenarios. First, we examine how the predicted choices would look in an environment where there are no supply constraints, i.e. the distances to all the alternatives equal to 2km, and enough open schooling positions for the whole cohort. Second, as the past (and current trend) in educational supply in Finland has been one of consolidation and closing down of smaller establishments, we take the actual supply structure of 2010 as another counterfactual scenario. (Later we can also compare the predictions from this scenario to actual application data from 2010).

We examine these counterfactuals separately for each of our 6 subsamples. These predictions tell us how individuals would substitute to different alternatives (in the aggregate in each subsample). Furthermore, we can look at counterfactual changes for different groups of individuals within these subsamples, eg. by looking at the counterfactuals separately for individuals with low gpas versus high gpas. Later, we want to extend the counterfactual analysis to provide more insights at the individual level, i.e. to explore in more detail what kinds of individuals change their schooling choices in response to different changes in supply.

Under the counterfactual scenario of no supply constraints, we find that a large fraction of individuals (in each of the 6 subsamples) would move away from the choice of gymnasium to vocational fields. Quantitatively, we can also see some interesting results when comparing girls and boys, and those from low- and high-income families. In the lower income groups, and for those with lower gpas, we find a larger share of the students moving away from gymnasium than in the high income or high gpa groups. There are also differences in the fields they would change to. Preliminary results indicate that individuals are less sensitive to the supply concerning their participation decision than concerning their choice of field. This result is in line with prior studies (see eg. Kelchtermans and Verboven 2010a).

4.4 Robustness checks

There are several robustness checks we plan to do.

5 Conclusion

One of the objectives of educational policy is to guarantee educational basic rights to all young people regardless of where they live. However, Finland is a sparsely inhabited and geographically large country. We observe substantial differences in the supply of upper secondary education in different regions and investigate whether these differences make individuals to sort to different career paths, which may create inequality between regions. The availability of schooling may be of particular importance to the 16-year old individuals graduating from compulsory school as they are less likely to migrate to another region than the older age groups. Finally, our data also allows us to control for most of the factors found important in the previous literature.

We have put together a very detailed and rich individual level dataset on four full cohorts of Finns leaving comprehensive school. We observe their application information personal characteristics, school performance, information on the comprehensive school they attended, and parents' information. We also observe their location of residence at the time of finishing comprehensive school and several characteristics describing the region. We also construct measures of supply of education in each municipality and link the municipalities to inter-municipality distances to get measures of proximity of each education alternative.

In this paper, we use this data and employ conditional and mixed logit models to examine the schooling choices of young people and focus on the effects of educational supply. Our results show that the supply of education, in particular the availability and proximity of education alternatives, plays a statistically significant role in influencing schooling decisions and thus influences the sorting of individuals across the schooling alternatives. The effect is larger for boys than for girls, and for boys we also find significant heterogeneity of the effect in our random coefficients. Also, in line with prior evidence, we find that parental factors and previous school performance are important determinants of schooling choices. Overall, our analysis shows that boys are very different in their schooling choices and in how they respond to the various determinants.

We have started to analyse demand in different counterfactual situations of supply: how changing the supply of different alternatives affects the sorting of individuals. Under the counterfactual scenario of no supply constraints, we find that a large fraction of individuals (in each of the 6 subsamples) would move away from the choice of gymnasium to vocational fields. Qualitatively this is not a surprising result, given that gymnasiums are the most widely spread education alternative around the country and thus often the most attractive choice due to their proximity. If other schooling alternatives were as easily available, more individuals would choose them. This result, however, has some important implications. There has been some criticism of the Finnish system as being one that encourages too many young people to choose gymnasium and thereafter university education (even individuals who could be better suited for vocational education). The magnitude of the results from this counterfactual scenario indicates that a large fraction of students are pushed towards gymnasium due to its availability; a direct outcome of educational policy.

Quantitatively, we can also see some interesting results when comparing girls and boys, and those from low- and high-income families. In the lower income groups, and for those with lower gpas, we find a larger share of the students moving away from gymnasium than in the high income or high gpa groups. There are also differences in the fields they would change to. Preliminary results indicate that individuals are less sensitive to the supply concerning their participation decision than concerning their choice of field. This result is in line with prior studies (see eg. Kelchtermans and Verboven 2010a).

In work-in-progress, we explore the heterogeneity of the effects and substitution patterns in more detail as well as investigate the sorting of individuals under various alternative counterfactual scenarios.

The careful modelling of the schooling choices of individuals forms the foundation for various other studies that examine the progress of individuals through their schooling process, and their schooling outcomes.

References

Arcidiacono, P. (2004) "Ability sorting and the return to college major", Journal of Econometrics 121, 343-375.

Beffy, M., Fougére, D. and Maurel, A. (2009) "Choosing the field of study in post-secondary education: do expected earnings matter?" IZA Discussion Paper No. 4127.

Berger, Mark C. "Predicted Future Earnings and Choice of College Major", Industrial and Labor Relations Review, April 1988, 41(3). p. 418-29.

Boudarbat, B. and Montmarquette, C. (2007) "Choice of Fields of Study of Canadian University Graduates: The Role of Gender and their Parents' Education", IZA Discussion Paper No. 2552.

Bound, J. and Turner, S. (2007) "Cohort crowding: How resources affect collegiate attainment", Journal of Public Economics 91, 877-899.

Card, D. (2001) "Estimating the return to schooling: progress of some persistent econometric problems", Econometrica 69:5, 1127-1160.

Dustmann, C. (2004) "Parental background, secondary school track choice, and wages", Oxford Economic Papers 56, 209-230.

Eide, E. and Waehrer, G., 1998, "The role of the option value of college attendance in college major choice", Economics of Education Review, Volume 17, Issue 1, Pages 73-82.

Frenette, M. (2004) "Access to college and university: does distance to school matter?", Canadian Public Policy 30:4, 427-443.

Gibbons, S. and Vignoles, A. (2009) "Access, choice, and participation in higher education", CEE Discussion Paper No. 101.

Kaukonen, R. (2009) "Valtaosa uusista ylioppilaista viettää välivuoden opinnoista", Hyvinvointikatsaus 1/2009. Tilastokeskus.

Kelchtermans, S. and Verboven, F., 2010a, Participation and Study Decisions in a Public System of Higher Education, Journal of Applied Econometrics, 25, pp. 355-391.

Kelchtermans, S. and Verboven, F., 2010b, Program duplication in higher education is not necessarily bad, Journal of Public Economics, 94, pp. 397–409.

Long, B. T. 2004, How have college decisions changed over time? An application of the conditional logistic choice model. Journal of Econometrics, 121(1-2): 271-296.

Saks, R.E. and Shore, S.H., (2005), "Risk and Career Choice", Advances in Economic Analysis and Policy, 5(1), Article 7.

Spiess, C.K. and Wrohlich, K. (2008) "Does distance determine who attends a university in Germany?", IZA Discussion Paper No. 3615.

Train, Kenneth, 2009, Discrete Choice Methods with Simulation, Cambridge University Press (second edition).

Press release of the Ministy of Education (2008) "Sivistyspoliittinen ministeriryhmä linjasi koulutuksen alueelliset tavoitteet vuodelle 2012." 16.5.2008.

APPENDIX A Fields and sub-fields of vocational education (the classification by the Finnish National Board of Education in 1995)

FIELDS OF VOCATIONAL EDUCATION	SUB-FIELDS
Natural Resources Sector	
	Agriculture
	Horticulture
	Fishery
	Other primary industries
	Forestry
Technology and Transport Sector	
	Graphics technology
	Heating and ventilation
	Machinery and metal technology
	Vehicles and transportation
	Textiles and clothing
	Food industry
	Electrical engineering
	Land survey technology
	Construction technology
	Wood industry
	Surface treatment
	Paper and chemical industry
	Seafaring
	Other technology and transportation
Administration and Commerce Sector	
	Business and administration
Hotel, Catering and Home Economics Sector	
	Hotel, restaurant and catering
	Home economics and cleaning services
Social and Health Care Services Sector	
	Social and health care services
	Beauty care
Culture Sector	
	Crafts and design
	Communications and visual arts
	Music
	Theatre and dance
Humanist and Teaching Sector	
	Leisure activities
	Physical education

Table 1a. Regional availability of the schooling alternatives at the municipality level

Schooling alternative	Distance to the nearest municipality offering each alternative (km)				Municipalities with each alternativity (%)*	Transition probability (%)**	
	Mean*	Max*	SD* (between)	SD** (within)		From 1 to 0	From 0 to 1
Gymnasium	6.6	43	9.2	0.5	76.7	0.1	0.4
Humanities and Education	74.9	570	64.8	14.4	6.4	4.1	0.9
Culture	39.6	281	35.7	7.0	18.4	2.1	2.2
Social Sciences, Business and Administration	33.5	154	24.8	7.1	20.6	1.0	0.1
Natural Sciences	52.2	325	39.0	15.1	11.0	2.9	3.1
Technology, Communications and Transport	24.4	320	27.5	1.6	34.0	0.3	0.6
Natural Resources and the Environment	33.3	144	24.4	6.0	20.9	4.0	0.3
Social Services, Health and Sports	40.5	434	42.1	12.1	18.4	6.4	1.6
Tourism, Catering and Domestic Services	23.0	125	21.2	3.8	35.6	2.3	0.6
Any vocational field	16.9	125	19.2	1.8	50.0	1.0	1.2
Any schooling alternative	6.3	43	9.0	0.5	78.2	0.1	0.5

Note. * Cross-section in year 2000 (n=326)
** During the observation period 2000-2003 (n=326, T=4)

Table 1b. Regional availability of the schooling alternatives at the sub-region level

Schooling alternative	Open positions per cohort (%)			
	Mean*	Max*	SD* (between)	SD** (within)
Gymnasium	62.3	144.0	13.6	6.4
Humanities and Education	0.7	8.4	1.6	0.8
Culture	3.3	21.6	4.2	2.0
Social Sciences, Business and Administration	6.0	18.2	4.8	1.4
Natural Sciences	1.5	9.8	2.3	1.3
Technology, Communications and Transport	25.5	44.8	12.6	3.7
Natural Resources and the Environment	5.2	28.5	6.2	1.7
Social Services, Health and Sports	6.1	20.9	5.8	1.6
Tourism, Catering and Domestic Services	10.8	44.0	7.2	2.6
Any vocational field	59.1	109.9	25.5	7.9
Any schooling alternative	121.4	171.9	26.2	11.6

Note. * Cross-section in year 2000 (n=74)
** During the observation period 2000-2003 (n=74, T=4)

Table 2. Schooling choices conditional on the availability of the alternative in the municipality (%)

Schooling choice	Gymnas		Natural Resources		Tech and Transport		Admin and Commerce		Hotel, Catering and Home Econ		Social and Health Care Services		Culture		Humanist and Teaching	
Boys																
All boys	49.5		2.1		35.2		7.3		3.8		0.5		1.4		0.2	
Available in the municipality	50.0	+++	2.3	++	33.6	---	8.5	+++	4.1	+++	0.6		1.7	+++	0.3	++
Not available	40.6		1.9		39.6		5.1		3.0		0.5		1.2		0.1	
Girls																
All girls	67.9		1.6		2.8		4.4		8.2		11.3		2.7		1.0	
Available in the municipality	68.0	+++	1.6		2.9		4.5		8.0	--	10.9	---	2.7		1.4	+++
Not available	64.5		1.6		2.8		4.3		9.0		11.9		2.7		0.9	

Table 3. Schooling choices conditional on parents' field of education

Schooling choice	Gymnas		Natural Resources		Tech and Transport		Admin and Commerce		Hotel, Catering and Home Econ		Social and Health Care Services		Culture		Humanist and Teaching	
Boys																
Mother's field of education	73.5	+++	8.2	+++	45.8	+++	9.7	+++	4.0		0.8	+++	2.6		0.3	
Not mother's field of education	38.9		2.0		34.6		7.0		3.8		0.5		1.4		0.2	
Father's field of education	83.3	+++	9.6	+++	40.3	+++	11.3	+++	7.3	+++	1.7	+++	0.9		0.8	+
Not Father's field of education	42.4		2.7		33.2		7.2		3.8		0.5		1.4		0.2	
Girls																
Mother's field of education	87.0	+++	4.5	+++	4.2	+++	4.7		11.7	+++	12.4	++	6.6	+++	1.7	
Not mother's field of education	59.4		1.6		2.3		4.4		7.9		11.1		2.7		1.0	
Father's field of education	91.1	+++	2.3	++	2.7		3.5		9.0		10.6		2.7		1.9	
Not Father's field of education	63.2		1.6		2.9		4.4		8.3		11.3		2.7		1.0	

Table 4 Characteristics of individuals and their parents conditional on the schooling choices

Schooling choices	Gymnasium		Natural Resources		Technology ann Transport		Administration and Commerce		Hotel, Catering and Home Economics		Social and Health Care Services		Culture		Humanist and Teaching	
Characteristics of the individuals																
Share of men (%)	43.3	---	57.5	+++	92.8	+++	63.4	+++	32.6	---	4.7	---	35.9	---	15.3	---
Average grade of theoretical subjects	8.21	+++	6.78	---	6.52	---	6.91	---	6.71	---	6.91	---	6.99	---	6.91	---
Average grade of all subjects	8.27	+++	7.11	---	6.86	---	7.19	---	7.07	---	7.28	---	7.32	---	7.26	---
Average number of pupils in comprehensive school	135	+++	112	---	125	---	135	+++	123	---	133	+	129	-	128	
Family income	57,097	+++	37,093	---	43,443	---	47,657	---	42,629	---	43,572	---	46,029	---	43,525	---
Mother's level of education (%)																
Secondary education	33.43	---	45.85	+++	43.72	+++	41.49	+++	39.82	+++	42.43	+++	41.48	+++	44.94	+++
College- level	25.14	+++	13.46	---	13.29	---	16.53	---	11.05	---	11.45	---	16.78	---	15.06	---
Higher education	19.44	+++	5.16	---	4.43	---	6.19	---	4.03	---	3.68	---	7.58	---	6.35	---
Father's level of education (%)																
Secondary education	27.66	---	38.89	+++	34.29	+++	30.79		29.19		30.98	+	29.66		36.94	+++
College- level	13.43	+++	6.28	---	6.71	---	9.40	---	4.83	---	5.55	---	8.66	--	8.24	
Higher education	21.36	+++	2.84	---	3.79	---	5.95	---	3.48	---	3.42	---	7.05	---	4.47	---
Mother's sosioeconomic status (%)																
Self-employed persons	8.43		21.62	+++	8.35		7.19	---	6.82	---	7.44	--	8.72		8.00	
Upper-level employees	19.53	+++	4.41	---	4.82	---	7.67	---	4.28	---	4.13	---	7.11	---	6.59	---
Lower-level employees	37.49	+++	22.66	---	28.49	---	32.47		24.73	---	26.66	---	30.60	--	29.41	-
Manual workers	12.42	---	18.70	++	22.80	+++	19.23	+++	22.85	+++	22.36	+++	18.19	++	21.18	+++
Students, pensioners and others	9.48	---	12.86		14.83	+++	13.14	+++	15.19	+++	14.57	+++	13.62	++	13.88	
Socioeconomics status unknown	12.66	---	19.75	++	20.72	+++	20.30	+++	26.14	+++	24.84	+++	21.74	+++	20.94	++
Father's sosioeconomic status (%)																
Self-employed persons	13.00	+++	28.20	+++	12.39		11.35	---	9.58	---	10.13	---	12.48		9.88	-
Upper-level employees	21.18	+++	3.37	---	4.51	---	7.65	---	4.30	---	4.20	---	8.26	---	5.41	---
Lower-level employees	15.22	+++	5.98	---	9.09	---	11.81	-	7.59	---	8.74	---	9.26	---	10.82	
Manual workers	19.86	---	24.76		30.80	+++	26.26	+++	25.39	+++	27.44	+++	25.77	++	33.41	+++
Students, pensioners and others	6.24	---	10.70	+++	9.01	+++	7.30		8.67	+++	8.17	+++	6.24		5.88	
Socioeconomics status unknown	24.50	---	27.00	--	34.20	+++	35.63	+++	44.46	+++	41.32	+++	37.99	+++	34.59	++

Table 5 Distance to schooling, and characteristics of regions and alternatives conditional on the schooling choices

Schooling choice	Gymnasium	Natural Resources	Technology and Transport	Administration and Commerce	Hotel, Catering and Home Economics	Social and Health Care Services	Culture	Humanist and Teaching
Distance to the nearest municipality								
Gymnasium	2.7	3.4	3.1	2.7	2.8	2.8	2.8	3.0
Natural Resources	23.5	28.0	25.0	24.8	24.8	25.3	24.7	25.9
Technology and Transport	9.4	17.8	10.6	8.6	10.1	10.0	11.5	12.0
Administration and Commerce	13.2	25.0	16.1	11.2	15.0	15.6	15.2	16.8
Hotel, Catering and Home Economics	9.9	18.7	11.7	8.9	10.5	11.1	11.9	12.5
Social and Health Care Services	17.2	32.0	20.3	16.4	19.8	18.9	19.1	21.9
Culture	20.6	36.4	23.9	20.6	24.2	23.4	18.7	26.0
Humanist and Teaching	46.7	63.0	53.1	47.3	53.1	51.7	50.7	36.9
Characteristics of the region								
Size of the municipality (km^2)	1,220	1,538	1,280	1,278	1,343	1,299	1,246	1,317
Population density (inhabitants/km^2)	209	65	139	213	167	153	147	119
Municipal tax revenues (€)	3,265	2,885	3,103	3,246	3,165	3,139	3,114	3,031
Youth unemployment rate (%)	15.09	17.93	16.61	15.30	16.35	16.19	16.48	16.75
Jobs in agriculture (%)	5.46	10.70	6.80	5.34	6.12	6.31	6.71	7.31
Jobs in manufacturing (%)	26.21	27.31	27.74	26.71	26.84	27.59	27.99	28.69
Jobs in services (%)	67.21	60.72	64.30	66.86	65.90	64.96	64.17	62.79
Inward mobility (%)	4.90	4.31	4.66	4.81	4.77	4.74	4.72	4.42
Outward mobility (%)	4.94	4.56	4.76	4.86	4.83	4.82	4.85	4.58
Characteristics of alternatives								
Mean income of male	38,177	12,453	23,256	24,610	18,398	21,845	16,571	22,426
Mean income of female	24,228	10,571	14,744	18,036	14,232	17,731	11,994	17,040
Unemployment rate of male	5.3	9.4	11.6	10.1	14.6	6.4	18.3	6.9
Unemployment rate of female	4.9	14.3	16.0	10.4	13.4	6.2	18.7	11.5

Table 6. Estimated coefficients, conditional and mixed logit models, distance km

	Boys		Girls	
Distance to the nearest municipality	**Clogit**	**Mixed logit**	**Clogit**	**Mixed logit**
Mean	-0.010*** (0.001)	-0.014*** (0.002)	-0.005*** (0.001)	-0.005*** (0.001)
Std.E	-	0.011*** (0.003)	-	0.003 (0.003)
Number of obs	132,984	132,984	123,408	123,408
Log likelihood	-11812.06	-11802.61	-9797.64	-9795.40

Notes: Standard errors in parentheses, *** significant at 0.1 per cent
Estimations contain 16,623 boys and 15,426 girls.
All estimations Include mother's and father's socioeconomic status, level and field of education, family income, mother tongue, individual's average grade in comprehensive school, peer's average grade in comprehensive school, size of comprehensive school, urbanization of the municipality and youth unemployment

Table 7. Marginal effects, conditional logit, distance 100km

Boys	**High school**	**Natural Resources**	**Technology and Transport**	**Administration and Commerce**	**Hotel, Catering and Home Economics**	**Social and Health Care Services**	**Culture**	**Humanist and Teaching**
High school	-0.193*** (0.037)	0.004*** (0.001)	0.125*** (0.024)	0.044*** (0.008)	0.011*** (0.002)	0.001 (0.056)	0.008*** (0.002)	-
Natural Resources	0.004*** (0.001)	-0.006*** (0.001)	0.001*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000 (0.000)	0.000*** (0.000)	-
Technology and Transport	0.125*** (0.024)	0.001*** (0.000)	-0.139*** (0.019)	0.009*** (0.002)	0.002*** (0.001)	0.000 (0.012)	0.002*** (0.000)	-
Administration and Commerce	0.044*** (0.008)	0.000*** (0.000)	0.009*** (0.002)	-0.055*** (0.008)	0.001*** (0.000)	0.000 (0.004)	0.001*** (0.000)	-
Hotel, Catering and Home Economics	0.011*** (0.002)	0.000*** (0.000)	0.002*** (0.001)	0.001*** (0.000)	-0.015*** (0.003)	0.000 (0.001)	0.000*** (0.000)	-
Social and Health Care Services	0.001 (0.056)	0.000 (0.000)	0.000 (0.012)	0.000 (0.004)	0.000 (0.001)	-0.001 (0.074)	0.000 (0.001)	-
Culture	0.008*** (0.002)	0.000*** (0.000)	0.002*** (0.000)	0.001*** (0.000)	0.000*** (0.000)	0.000 (0.001)	-0.010*** (0.002)	-
Humanist and Teaching	0.000 (0.007)	0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-
Pr(choice = y|1 selected)	0.753	0.005	0.160	0.056	0.014	0.001	0.010	0.000
Girls								
High school	-0.036 (0.028)	0.002 (0.013)	0.003** (0.001)	0.007*** (0.002)	0.006*** (0.002)	0.012*** (0.003)	0.005 (0.028)	0.001 (0.009)
Natural Resources	0.002 (0.013)	-0.002 (0.014)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Technology and Transport	0.003** (0.001)	0.000 (0.000)	-0.003*** (0.001)	0.000** (0.000)	0.000** (0.000)	0.000** (0.000)	0.000 (0.000)	0.000 (0.000)
Administration and Commerce	0.007*** (0.002)	0.000 (0.000)	0.000** (0.000)	-0.007*** (0.002)	0.000** (0.000)	0.000*** (0.000)	0.000 (0.000)	0.000 (0.000)
Hotel, Catering and Home Economics	0.006*** (0.002)	0.000 (0.000)	0.000** (0.000)	0.000** (0.000)	-0.007*** (0.002)	0.000*** (0.000)	0.000 (0.000)	0.000 (0.000)
Social and Health Care Services	0.012*** (0.003)	0.000 (0.000)	0.000** (0.000)	0.000*** (0.000)	0.000*** (0.000)	-0.013*** (0.003)	0.000 (0.001)	0.000 (0.000)
Culture	0.005 (0.028)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.001)	-0.005 (0.030)	0.000 (0.000)
Humanist and Teaching	0.001 (0.009)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.001 (0.009)
Pr(choice = y|1 selected)	0.918	0.004	0.006	0.015	0.015	0.028	0.011	0.002

Notes: Standard errors in parentheses, ** significant at 1 per cent, *** significant at 0.1 per cent

Table 8. Marginal effects, conditional logit, boys

	High school	Natural Resources	Technology and Transport	Admin and Commerce	Hotel, Catering and Home Economics	Social and Health Care Services	Culture	Humanist and Teaching
Mother's field of education								
	0.024**	-0.001**	-0.015**	-0.005**	-0.001**	-0.000	-0.001**	0.000
General/high school	(0.008)	(0.000)	(0.005)	(0.002)	(0.000)	(0.007)	(0.000)	(0.000)
	-0.001*	0.001**	-0.000*	-0.000*	-0.000*	-0.000	-0.000**	0.000
Natural Resources	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	-0.016**	-0.000**	0.018**	-0.001**	-0.000**	-0.000	-0.000**	0.000
Technology and Transport	(0.006)	(0.000)	(0.006)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)
	-0.006**	-0.000**	-0.001**	0.007**	-0.000**	-0.000	-0.000**	0.000
Administration and Commerce	(0.002)	(0.000)	(0.000)	(0.002)	(0.000)	(0.001)	(0.000)	(0.000)
	-0.001**	-0.000*	-0.000**	-0.000**	0.002**	-0.000	-0.000**	0.000
Hotel. Catering and Home Economics	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000
Social and Health Care Services	(0.007)	(0.000)	(0.002)	(0.001)	(0.000)	(0.010)	(0.000)	(0.000)
	-0.001**	-0.000*	-0.000**	-0.000**	-0.000**	-0.000	-0.001**	0.000
Culture	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000
Humanist and Teaching	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Father's field of education								
	0.071***	-0.002***	-0.046***	-0.016***	-0.004***	-0.000	-0.003***	-0.000
General/high school	(0.013)	(0.000)	(0.007)	(0.003)	(0.001)	(0.020)	(0.001)	(0.002)
	-0.002***	0.003***	-0.000***	-0.000***	-0.000***	-0.000	-0.000***	-0.000
Natural Resources	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	-0.049***	-0.000***	0.054***	-0.004***	-0.001***	-0.000	-0.001***	-0.000
Technology and Transport	(0.008)	(0.000)	(0.005)	(0.001)	(0.000)	(0.005)	(0.000)	(0.001)
	-0.020***	-0.000***	-0.004***	0.025***	-0.000***	-0.000	-0.000***	-0.000
Administration and Commerce	(0.003)	(0.000)	(0.001)	(0.003)	(0.000)	(0.002)	(0.000)	(0.000)
	-0.005***	-0.000***	-0.001***	-0.000***	0.007***	-0.000	-0.000***	-0.000
Hotel. Catering and Home Economics	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	-0.000	-0.000
Social and Health Care Services	(0.027)	(0.000)	(0.006)	(0.002)	(0.001)	(0.035)	(0.000)	(0.000)
	-0.004***	-0.000***	-0.001***	-0.000***	-0.000***	-0.000	0.005***	-0.000
Culture	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000
Humanist and Teaching	(0.003)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.004)
Mother's level of education	baseline: upper secondary education							
	0.052***	0.000	-0.031***	-0.016***	-0.003	-0.000	-0.002	-0.000
College	(0.011)	(0.001)	(0.008)	(0.005)	(0.002)	(0.009)	(0.002)	(0.004)
	0.079***	-0.000	-0.055***	-0.023***	0.001	-0.000	-0.002	-0.000
Higher degree	(0.015)	(0.001)	(0.012)	(0.006)	(0.003)	(0.001)	(0.003)	(0.003)
Father's level of education	baseline: upper secondary education							
	0.051+	-0.000	-0.032	-0.011*	-0.005*	-0.001	-0.002	0.000
College	(0.026)	(0.001)	(0.011)	(0.005)	(0.002)	(0.035)	(0.002)	(0.002)
	0.080**		-0.050***	-0.021**	-0.008**	-0.001	0.001	0.000
Higher degree	(0.024)	-0.002	(0.014)	(0.007)	(0.003)	(0.031)	(0.003)	(0.008)
Mother's socioeconomic status	baseline: self-employed							
	0.016	-0.004**	-0.029	0.016	0.002	-0.001	-0.001	0.000
Upper-level employees	(0.049)	(0.001)	(0.018)	(0.011)	(0.004)	(0.058)	(0.003)	(0.013)
	0.026	-0.006***	-0.025	0.004	0.003	-0.001	-0.000	0.000
Lower-level employees	(0.046)	(0.001)	(0.016)	(0.008)	(0.003)	(0.059)	(0.003)	(0.005)
	-0.044*	-0.002*	0.028+	0.014	0.007	-0.000	-0.002	0.000
Manual workers	(0.021)	(0.001)	(0.015)	(0.009)	(0.005)	(0.002)	(0.002)	(0.006)
	-0.008	-0.002	-0.007	0.012	0.002	-0.000	0.002	-0.000
Students and pensioners	(0.032)	(0.001)	(0.020)	(0.014)	(0.006)	(0.014)	(0.005)	(0.013)
	-0.021	-0.003***	0.009	0.010	0.001	-0.000	0.004	0.000
Other	(0.024)	(0.001)	(0.017)	(0.011)	(0.004)	(0.002)	(0.005)	(0.001)
	0.037	-0.004**	-0.044+	0.013	-0.005	-0.002	0.005	-0.000
Information missing	(0.067)	(0.001)	(0.027)	(0.020)	(0.005)	(0.073)	(0.008)	(0.010)
Father's socioeconomic status	baseline: self-employed							
	0.002	-0.003**	-0.023	0.014	0.009+	0.001	-0.000	0.000
Upper-level employees	(0.039)	(0.001)	(0.014)	(0.009)	(0.005)	(0.045)	(0.003)	(0.010)
	0.017	-0.005***	-0.020	0.004	0.001	0.001	0.001	0.000
Lower-level employees	(0.055)	(0.001)	(0.014)	(0.008)	(0.003)	(0.049)	(0.003)	(0.048)
	-0.030	-0.001	0.016	0.009	0.004	0.001	0.000	0.000
Manual workers	(0.056)	(0.001)	(0.018)	(0.009)	(0.003)	(0.074)	(0.003)	(0.013)
	0.014	-0.001	-0.021	0.010	0.000	0.001	-0.003	-0.000
Students and pensioners	(0.062)	(0.001)	(0.020)	(0.014)	(0.005)	(0.071)	(0.004)	(0.013)
	0.020	-0.002+	-0.011	-0.012	0.006	0.003	-0.003	-0.000
Other	(0.128)	(0.001)	(0.029)	(0.011)	(0.007)	(0.160)	(0.003)	(0.012)
	0.008	-0.005***	-0.025	0.016	0.009	-0.002	-0.001	-0.000
Information missing	(0.054)	(0.001)	(0.025)	(0.018)	(0.009)	(0.053)	(0.005)	(0.010)
Average grade of individual	0.387***	-0.009***	-0.262***	-0.074***	-0.025***	-0.002	-0.014***	-0.000 (.)

Notes: Standard errors in parentheses, + significant at 10 per cent, * significant at 5 per cent, ** significant at 1 per cent, *** significant at 0.1 p