

IZA DP No. 6285

**Unequal Access to Higher Education in
the Czech Republic: The Role of
Spatial Distribution of Universities**

Michal Franta
Martin Guzi

January 2012

Unequal Access to Higher Education in the Czech Republic: The Role of Spatial Distribution of Universities

Michal Franta
Czech National Bank

Martin Guzi
IZA

Discussion Paper No. 6285
January 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Unequal Access to Higher Education in the Czech Republic: The Role of Spatial Distribution of Universities*

We explore a potential source of human capital spatial disparities: the unequal access to tertiary education caused by the absence/presence of a local university. Because the entrance to a university is a sequential process in the Czech Republic we model both a student's decision to apply to a university and the admission process. Two possible sources of unequal access to university are distinguished: cost savings and informational advantages for those residing close to a university. Estimation results suggest that the presence of a university per se is not driving student's decision to apply. Further we find that information advantage due to university proximity plays a significant role in the admission process. However this advantage is specific to the field of study, and becomes stronger in the case of highly oversubscribed study fields. To equalize the chance of admission, policy makers should consider geographical expansion of the system of universities accompanied by the expansion of university programs.

JEL Classification: I20, I21, J24

Keywords: human capital, spatial distribution, access to tertiary education

Corresponding author:

Martin Guzi
IZA
P.O. Box 7240
53072 Bonn
Germany
E-mail: guzi@iza.org

* This research was supported by grant No. 322/2006 from the Grant Agency of Charles University. We thank Daniel Münich, Štěpán Jurajda, Jan Kmenta, Peter Katuščák, Richard Stock and Bart Cockx for comments and suggestions.

NON-TECHNICAL SUMMARY

In the paper we analyze the demand for tertiary education depending on university proximity. In the Czech Republic the tertiary level of education is provided solely by universities. Moreover the entrance to a university is a sequential process: first the student decides which university to apply and then his/her admittance depends on the performance in the admission test. We identify the most influential factors affecting both the probability of applying and the probability of admission to a university.

The study explores the difference in the probability of entering a tertiary education for two secondary school graduates who differ only in that the first one resides within commuting distance to university while the other one does not. We presume that the advantage of having access to a local university follows in two ways. First, cost savings from being able to live at home with their parents and not having to move. In this manner the costs of acquiring a diploma at a local university are lower. Second, informational advantages from a university milieu in the neighborhood are beneficial in the admission process. The information advantage to a potential applicant transfers through face-to-face contacts with university students, possibility to use university facilities during secondary school study or easier access to preparatory courses organized by local faculties, etc. Study detects the presence of the two described causes of unequal access to tertiary education and assesses their relative importance.

Employing logit model regressions, we show that individual characteristics (parents' education, computer ownership, test score) as well as class characteristics (class size, average class test score) and district characteristics (unemployment rate, regional GDP growth) are significant influential factors affecting a graduate's decision to apply for tertiary education. Estimation results suggest that the presence of a university per se is not driving the student's decision to apply.

Furthermore, in our econometric analysis, we find that information due to the presence of local university plays a significant role in the admission to university. However the advantageous information is specific to the field of study. Therefore living near a university that provides an applicant's preferred study field increases the admission probability to that field even if the applicant applies to a different university. The effect is stronger in the case of highly oversubscribed study fields.

The analysis shows that the system of the financial support of university students (e.g. dormitories, student loans) is sufficient to equalize the differences in the probability of applying caused by the presence/absence of a local university. To equalize the chance of admission policy makers should consider geographical expansion of the system of universities accompanied by the expansion of university programs.

1 Introduction

A growing interest in the various consequences of human capital geographical distribution can be noticed in the recent economic literature. Research focuses mainly on productivity spillovers, regional economic and population growth, and regional unemployment.¹ In general, significant effects of human capital spatial disparities have been found. The impact of the unequal spatial distribution of human capital can be even more pronounced in post-transition countries since they exhibit a higher geographical variation of the tertiary-educated population than countries in Western Europe.²

Given the importance of private and social returns to human capital and the implications of its geographical distribution especially in post-transition countries, a natural question arises: What determines the spatial distribution of human capital? In this paper we deal with a fundamental source of human capital: university education.³

The literature distinguishes three principal determinants of the geographical distribution of the tertiary-educated population: the general migration of the tertiary-educated population,⁴ the post-university migration of graduates,⁵ and unequal access to tertiary education. We examine the third potential reason, exploring the determinants of access to university study. In particular, we focus on the unequal access to tertiary education caused by the absence/presence of a local university. We ask whether students living close to a university are more likely to continue with tertiary education after they graduate from secondary school.

¹ Productivity spillovers are discussed in Moretti (2004) and Lange and Topel (forthcoming). The link between the local level of human capital and regional economic and population growth is examined, for example, in Glaeser et al. (1995) and Kaldewei and Walz (2001). Regional unemployment is discussed in Overman and Puga (2002) for developed countries and in Jurajda and Terrell (2008) for post-transition countries.

² See Table 1 that reports the coefficient of the variation of the tertiary-educated population share for selected post-transition and western European countries.

³ Human capital can be acquired via education, training, experience, or medical treatment (Becker 1964).

⁴ See Giannetti (2002, 2003).

⁵ See Bound et al. (2004) and Makovec (2005).

Basically, we consider two ways how the local university affects a secondary school graduate in the admission process.⁶ First, students living near a university face a cheaper option for obtaining tertiary education. They can continue to live with their parents, saving on rent and moving costs, as opposed to graduates residing away from a university.⁷ Second, secondary school graduates residing close to a university can be better informed about the admission process, university study, job prospects related to a university degree, the local university's open days, etc. The second reason, based on heterogenous information, is realized through various channels: information provided by secondary schools, face-to-face contacts with university students, the possibility to use university facilities during secondary school study, etc. In general, secondary school students living near a university can benefit from the university environment they live in and consequently can be more advantaged in the admission process.⁸

Our goal is to detect the presence of the two mentioned causes of unequal access to tertiary education and assess their relative importance. The identification is possible due to a unique data set. It enables us to scrutinize the process of entrance to a university in detail; we distinguish two stages—application and admission. So, our research goes beyond the existing studies that usually deal with university enrollees only.

We conduct the analysis for the Czech Republic, a country with the high geographical variation of tertiary-educated population typical for post-transition countries (Table 1). Graph 1 depicts the distribution of the tertiary-educated population by district (NUTS4-level) in the Czech Republic for 1991 and 2001. This implies that the population with a university degree is concentrated mainly in a few areas. There were two districts with

⁶ Note that we elaborate the effect of a local university on the demand for tertiary education. The supply of tertiary education has remained unchanged since the 1960s when the current system of universities were established in the Czech Republic.

⁷ Matějů (2007) presents the results of an income and expenditure survey of university students (Eurostudent 2005). According to this survey, 40% of university students live with their parents. Moreover, expenditures for transportation and accommodation account for the one third of total students' expenditures. So, attending a local university represents a large savings.

⁸ In general, the two reasons are a) the set of factors that affect performance on the admission test and influence the decision on application and b) the set of factors that do not affect performance on the admission test and do influence the decision on application. The related economic literature discussed in the next section points out the access to information as a prominent representative of the first set and the costs of study as a prominent representative of the second set. Therefore, in the paper we consider only these two reasons instead of the broadly defined set of factors.

15% and 17% of university degree holders in 1991 (23% and 25% in 2001). The vast majority of districts, however, reports only 5% of university graduates in 1991 (10% in 2001). Moreover, Graph 2 suggests that the highest increase in the tertiary-educated population during 1991–2001 occurred in regions (NUTS3-level) already endowed with a high portion of tertiary-educated population. There are two possible explanations: either the highly-educated population has migrated towards the highly-educated regions or secondary school graduates residing in regions without a university have faced worse prospects regarding university study.

The Czech Republic experiences not only the heterogenous spatial distribution of the tertiary-educated population but also highly unequal spatial distribution of universities. According to the Czech Institute for Information on Education (2002), 94% of public university departments are concentrated in 13% of districts. In terms of enrolled students the problem of the unequal distribution of the sources of tertiary education is even more obvious. The majority (over 70%) of all university students enrolled in the academic year 1997/98 attended universities located in three districts only (Prague, Brno, and Ostrava).

We exploit the high spatial variation in the number of institutions providing tertiary education to estimate the effect of a local university on a prospective university applicant. We set up a structural model of the post-secondary schooling decision—whether the secondary school graduate applies to a university or not. We estimate the reduced form of the model and discuss the relationship between the reduced form model parameters and structural parameters. Estimation results suggest that study costs do not drive a graduate’s application decision. Rather there is a positive effect of the district’s characteristics, represented by the share of the tertiary-educated population in the district, on the prospective applicant. The impact is larger for a graduate from a specialized secondary school than from a gymnasium. Moreover, we observe different application strategies followed by applicants from districts with and without a local university.

Next, we discuss a model of admission (conditional on application). The econometric analysis reveals that applicants are, in terms of their admission probabilities, positively

affected by the presence of a local university. Moreover, we find that the influence of a local university is “program specific”, i.e. a local university brings an advantage to applicants applying to programs provided by the local university even if they apply to that program in a non-local university. So, living close to any university need not necessarily help in the admission process but living near the “right” university helps. We also briefly discuss the nature of information flows that underlie the positive effects of local universities.

The identification of the source of unequal access to a university is of high policy relevance. If secondary school graduates are constrained by the cost of study then policies should support those living far from a university. On the other hand, if the source lies in the distance-dependent distribution of information, policies should expand tertiary education to other regions or improve the access to information for disadvantaged graduates.

The structure of the paper is the following. In the next section we discuss the research dealing with the effects of local universities and the post-secondary schooling decision. We also describe the schooling system and the admission process to a university in the Czech Republic. Section 3 introduces the reduced form models of applying and admission. In Section 4, we introduce the data set employed, point out its uniqueness and carry out a preliminary analysis. We also discuss the variables entering our econometric analysis. Estimation results are reported in Section 5. The structural and reduced form models of applying and the model of being admitted (conditional on application) are introduced in Appendices A, B, and C. Graphs and Tables can be found at the end of the paper in Appendix D.

2 Literature review

The analysis carried out in this paper connects two branches of economic research: the research focusing on the various effects of university presence (or university proximity) and the research exploring the determinants of the post-secondary schooling decision.

Models of the post-secondary schooling decision usually do not take into account the effects of local universities or the distance to the nearest university.

The research on the effects of university presence is mainly focused on regional and individual economic outcomes—local productivity and output (Andersson et al. 2001), and the local labor market (Beeson and Montgomery 1993). University proximity is used as a source of exogenous variation in educational attainment in studies on returns to education (Card 1995, Card 1999).

The effect of a local university on the prospective university applicant is considered in Do (2004), Frenette (2002, 2003), Sá et al. (2004) and Eliasson (2006). All of these authors recognize the effect of the lower cost of study for the population living near a university and in general find that those living near a university participate in university study more often. In addition, Do (2004) takes into account a “knowledge spillover” that influences the choice of the quality of university where university applicant enrolls. He finds that the quality of the university in which the applicant enrolls is positively linked to the quality of a local university. Frenette (2002) also mentions that the dependence of the probability to enroll on the distance to the nearest university could be caused by fact that “students in outlying areas simply don’t see the benefits from a university education since fewer people hold a degree”. The research approach he employs, however, cannot distinguish between these two determinants of behavior dependent on the distance to the nearest university.

Regarding the post-secondary schooling decision there are three main theoretical approaches established in the literature. First, human capital theory views the decision on taking another period of schooling as an investment decision (Becker 1964, structural model in Willis and Rosen 1979). An individual compares the present value of future benefits based on expected future earnings with the costs related to continuing education. Another approach assumes that education is also a consumption good and the decision on post-secondary education is a current consumption choice (e.g. Gullason 1989). Finally, the third approach views schooling as an indicator of an individual’s capabilities that has nothing to do with the individual’s productivity. Therefore, the decision on post-secondary education reflects the individual’s

willingness to provide the signal (Spence 1973). The structural model of application derived in Appendix A is based on the human capital approach.

Long (2004), Sá et al. (2004), Whitehead et al. (2006), and Brooks (2002) represent recent empirical studies on the determinants of the post-secondary schooling decision. Individual capabilities, a student's socio-economic background, the characteristics of universities, regional economic conditions, and local social interactions are found to be influential for the outcome of secondary school graduates' decision processes. Brooks (2002) focused on these determinants from the point of view of the information they convey, discussing possible inequalities among students. She points out, for example, the importance of interactions with current university students in the prospective applicant's decision process. Details on the determinants of the post-secondary schooling decision are introduced in Appendix A where we derive an economic model.

2.1 Czech educational system

In this section, we describe the educational system in the Czech Republic circa 1998. Primary education lasts nine years. Afterwards students apply to various types of secondary school depending on their future career plans and ability. The lower level of secondary education involves two years of vocational education. The higher level of secondary education takes four years and comprises three types of schools: (i) vocational education leading to a certification exam, (ii) specialized secondary (technical) education in professional fields, and (iii) general secondary education in academic high schools known as gymnasiums.⁹

The tertiary level of education is provided solely by universities. The admission process to a university is a sequential process: first the student decides where to apply, then the university decides who to admit, and then the students decides where to enroll. In the first step, the student carries out a simultaneous decision whether to apply to a university and where. In order to apply the student must hold a final general exam

⁹ Gymnasiums and specialized schools correspond to the ISCED level 3A, while vocational schools to the ISCED level 3C.

credential known as *maturita* from secondary school. The *maturita* exam is a comprehensive examination that comprises Czech language and an additional two or three subjects chosen by the student. However, the examinations are not standardized and therefore universities do not take into account the performance on the *maturita* exam during the admission process. All gymnasiums and 95 percent of specialized secondary school programs, but only 12 percent of vocational school programs, lead to *maturita* and the possibility of university admission. Vocational programs finishing with a *maturita* exam represent only a small minority of all vocational programs, therefore we exclude them from the analysis. In 1998, 60 percent of secondary school students finished with *maturita* and thus were eligible to apply for tertiary education.

In 1998, there were 111 university departments in 23 universities in the Czech Republic. Universities provide education in 41 program specializations.¹⁰ Altogether there were 225 different university programs available in 1998. Government-owned universities did not charge tuition and only performance stipends were in place. The law to accredit private universities in the Czech Republic was passed in 1999, therefore they do not enter our analysis. Finally, a system of support for university students was in operation in 1998 in the Czech Republic. It involves, for example, the possibility of living in dormitories for those residing out of commuting distance from a university, meal tickets, and public transportation discounts. The system was intended to lower the study costs especially for those living far from the university they were enrolled in.

Each university charged a fee for each application, which is not a limiting factor but it made applicants ration their decisions. The number of applications per person is not limited and on average applicants sent two applications. Some university programs were highly oversubscribed.¹¹

Students sent applications to preferred university programs and all of them were invited to participate in admission tests. The decision on admission was solely on the side of the university and was based on the performance on admission tests and/or oral interviews. Universities did not take into account the applicant's residence. In 1998

¹⁰ The classification of program specialization corresponds to the two-digit JKO classification.

¹¹ In 1998, the probability of admission to psychology programs was only 5% and 10% to law programs.

about 50 percent of graduates who applied to a university were admitted to at least one program. If the applicant succeeds in admission to several university programs¹² he chooses where to enroll.

3 Models

As described in the previous section, the admission process to university comprises the graduate's decision on applying to the program, the university's decision on admission, and the applicant's decision on enrollment.

The derivation of a structural model of the application decision and its reduced form is discussed in Appendix A. The reduced form model takes the following form

$$Apply_{icd} = \alpha_0 + \alpha I_{icd} + \beta C_{cd} + \phi R_d + \delta d_d + \sum_{j=1}^{8(7)} \mu_j SD_{icd}^j + \varepsilon_{icd}, \quad (1)$$

where $Apply_{icd}$ is a dummy variable that indicates whether a graduate i from secondary school class c and district d applies to a university. In addition, we control for individual characteristics (vector I_{icd}), class (secondary school) characteristics (vector C_{cd}), and regional (district) characteristics (vector R_d). The coefficient on the dummy variable d_d is our primary interest. The dummy equals 1 if a graduate resides within commuting distance to the nearest university and 0 otherwise. Finally, the set of eight (seven for specialized secondary schools) dummy variables SD_{icd} accounts for subjects taken at the maturita exam for gymnasiums and the field of study in specialized secondary schools. The set of variables included on the right hand side is a compromise between data availability and variables suggested by research on the post-secondary schooling decision discussed in the literature review and Appendix A. Variables are described in Section 4.

¹² 27% of those admitted were admitted to two programs in 1998, 9% to 3 programs.

The dummy variable indicating the presence of a local university d_d in (1) captures both of the effects of a local university on the prospective applicant: the lower costs of study and heterogenous information. However in the application regression we cannot distinguish these two effects. We resolve the problem in the admission regression where the applicant's potential cost of study does not play any role and the coefficient on the local university dummy captures the effect of heterogenous information only.¹³

A non-zero coefficient on the local university dummy variable in the admission regression equation reveals whether those living near a university are advantaged in the admission process because of information spread within the university neighborhood. To examine the nature of heterogeneous information in detail we add an additional dummy variable f_j that identifies applicants according to whether they live close to a university offering the program they apply to. So, f_j equals one if an applicant applies to a program that is provided by a local university even if the applicant applies to that program at other (possibly non-local) universities.

So, we divide all the applications into three subgroups. We distinguish applications by applicants living far from any university ($d_d = 0, f_j = 0$), applicants living near a university that does not provide the applicant's preferred program ($d_d = 1, f_j = 0$), and applicants living near a university providing the applicant's preferred program ($d_d = 1, f_j = 1$).

The estimated coefficients on the two dummy variables indicate whether either proximity to any university or proximity to a university with the desired program can provide an advantage in the admission process at any other university.¹⁴

¹³ Note that information affecting the decision on application and information providing an advantage in the admission test (and/or oral interview) need not necessarily be of the same nature. Therefore, by econometric analysis we identify whether heterogenous information along with the costs of study play a role in the application decision and whether heterogenous information affects admission to university. Then, based on an assumption about the common nature of heterogenous information we can discuss the relative role of costs of study and heterogenous information in the application decision.

¹⁴ So we ask, for example, whether an applicant residing near a university providing medical programs is more informed and thus advantaged in the admission test to a medical program compared to an applicant living near a university that does not provide medical programs or another applicant living far from a university.

Our baseline specification of the model of admission (conditional on application) derived in Appendix C is

$$Admitted_{icd}^j = \beta_0 + \alpha I_{icd} + \beta C_{cd} + \varphi R_d + \delta_1 d_d + \delta_2 f_j + \rho F_j + v_{icd}^j, \quad (2)$$

where $Admitted_{icd}^j$ is a dummy variable that indicates whether application i has successfully passed through the admission procedure at a program j (conditional on applying).¹⁵ We control for individual characteristics (I_{icd}), secondary school (class) characteristics (C_{cd}), and district characteristics (R_d). Finally, the admission regression equation includes the vector of university characteristics F_j that control for university-specific characteristics relevant to the admission process—the self-selectivity of applicants and differences in admission tests (see discussion in Appendix C).

4 Data

Our empirical analysis is based on the following two anonymized data sets collected by the Institute for Information on Education (IIE) in 1998: (i) the data set of all university applicants (Uchazec) and (ii) the data set of all secondary school graduates (Maturant). Additional district descriptions are taken from the Czech Statistical Office.

The database Uchazec contains data on all individuals who applied to universities in 1998. Specifically, it provides information on all applications sent to universities for a given year together with the result of the admission process. The data allow distinguishing applications sent to a university by program specialization. The database Maturant is the result of a project that tested all graduates at every secondary school that finishes with the maturita exam in the Czech Republic in 1998. The database provides data on student characteristics, performance and subjects taken at the maturita exam. The graduate's performance is measured by the average test score computed

¹⁵ The unit of observation is an application. In the case of the application decision regression the unit of observation is a secondary school graduate.

from four tests taken in Czech, one foreign language, mathematics and study aptitude and this variable is called the composite score. We merged the database Maturant to Uchazec on an individual basis to obtain a set of information on the cohort of secondary graduates finishing their study with the maturita exam in 1998 augmented with graduates' revealed preferences for tertiary education. The uniqueness of the resulting data set allows us to provide a detailed analysis of graduates' education track depending on university accessibility.

Estimating models (1) and (2) we employ the following variables. Vector I_{icd} includes individual characteristics: a female dummy, a dummy for if born before 1980, composite score rank,¹⁶ the level of parental education, and a dummy for computer ownership. Computer ownership is included as a proxy for the missing information on family income. We distinguish the three levels of parental education (i) basic or vocational (reference category), (ii) secondary, (iii) tertiary; the highest level of the two parents is taken. Vector C_{cd} includes class (secondary school) characteristics: class size, composite score class average and private secondary school indicator.

For regression equation (1) vector R_d consists of district (NUTS4-level) unemployment rate and regional (NUTS3-level) GDP growth in 1998. For both regression equations (1) and (2) vector R_d includes the share of tertiary-educated population in the district¹⁷ and the relative excess demand for gymnasiums in the district. Relative excess demand for gymnasiums in a district is computed as the demand for gymnasium seats predicted using the share of tertiary-educated population in a district, a university presence dummy and the share of seats at gymnasiums taken by 6- and 8-year gymnasium programs subtracted by the supply of gymnasium seats (relative to all secondary school seats in a district).¹⁸ Basically, relative excess demand serves as a proxy for the average level of non-cognitive skills for secondary school graduates.¹⁹ A high district relative

¹⁶ Composite score rank expresses the rank of each graduate in the whole cohort of graduates, comparing their composite scores. The variable is normalized so that a rank of 100 is the best graduate and the rank of 0 the worst.

¹⁷ Data on the tertiary-educated population are taken from Census 2001.

¹⁸ The computation procedure for relative excess demand for the districts in the Czech Republic is introduced and thoroughly discussed in Drnakova (2006). She kindly provided us with data for 2002/2003; earlier data are not available.

¹⁹ The effect of non-cognitive skills on various outcomes is discussed in Heckman et al. (2006).

excess demand for gymnasiums implies that students entering gymnasiums exhibit on average a higher level of non-cognitive skills in comparison with districts with low relative excess demand.²⁰ Specialized secondary schools are often viewed as the second best option—those not admitted to gymnasiums enter specialized secondary schools. Thus, districts with high relative excess demand for gymnasiums are assumed to exhibit a high level of non-cognitive skill even for students of specialized secondary schools. So, in the econometric analysis we control for individual cognitive skills using the composite score of an individual and for non-cognitive skills incorporating relative excess demand for gymnasiums in a district.

The set of dummies SD_{icd} denotes subjects taken at the maturita exam for gymnasiums (mathematics, biology, physics, chemistry, history, geography, social sciences, and foreign language) and the field of secondary school for specialized secondary schools (agriculture, manufacturing, light manufacturing, health care, social sciences, art, and business as a reference category).

The dummy variable d_d equals 1 if a graduate resides within commuting distance of the nearest university. We have collected data on travel time between the district capital of a graduate's secondary school and the nearest university.²¹ We determine the district of a graduate's residence based on the address of her secondary school because data does not provide any information about a graduate's residence.²² The travel time threshold for the dummy variable d_d is set to 30 minutes. It is important to note that the travel time is computed for travel by car and that travel within cities (from home to university) is not considered. The same journey by bus or train usually takes a longer time. We illustrate this by Graph 3 which depicts the kernel density of commuting time by public transportation for secondary school graduates as of 2007. Moreover, additional time is needed for inner city travel so the overall commuting time (door to

²⁰ Non-cognitive skills involve e.g. motivation, persistence. We presume that a higher level of non-cognitive skills helps in admission to the secondary level of education.

²¹ The information about traveling time is computed using the software Kilometrovnik taken from the webpage www.trans.cz. We compute the time of a car driving from all 76 district capitals to each of 11 university centers.

²² The sources of heterogenous information considered are related to the secondary school location rather than place of residence. Our data confirms that more than 90% of students attend gymnasium in their district of residence. In the analysis we use the secondary school's district instead of the place of residence.

door) is higher. Therefore, we consider the threshold of 30 minutes as reasonable. Graph 4 shows the map of the Czech Republic with marked districts within commuting distance to the nearest university.

The vector of university program characteristics F_j includes variables that control for differences in admission tests over university programs, for the self-selectivity of applicants and for differences in excess demand for particular programs. As discussed in Appendix C we employ the following variables: dummy variables for university departments, program specializations and university programs. Further we include marginal composite admission score variable that is defined as the lowest composite score of the applicant who succeeds in the admission process (assuming that the university admits entirely according to the applicant's composite score).

4.1 Descriptive analysis

In this section we describe the population of secondary school graduates and university applicants. We also inspect the application strategies of secondary school graduates in terms of the chosen university programs and the admission probabilities of the chosen programs.

Table 2 provides a basic summary of the characteristics of gymnasium and specialized secondary school graduates and applicants to university in 1998. Graduates and applicants are further divided according to their residence type, i.e. whether there exists a university within commuting distance (local university) or not (no local university). It follows that more than 90% of gymnasium graduates choose to apply to university and approximately two-thirds of them are admitted regardless of the location. The corresponding figures for specialized secondary schools are 50% and 40%.

The considerable difference between gymnasiums and specialized secondary schools in the shares of applicants (and of those being admitted conditional on application) has two origins. First, study programs at gymnasiums are intended to prepare students for university study. Gymnasium graduates, therefore, generally perform better in

admission tests than applicants from specialized secondary schools. Second, the difference is also supported by the sorting process at the level of entry to secondary school. Students enter a gymnasium presuming they will continue their study at a university and therefore the population entering the secondary level of education is sorted according to interest in (and ability for) tertiary education. So, we expect different behavior of graduates regarding entrance into the tertiary level of education for the two types of secondary school and thus we carry out separate analyses for gymnasiums and specialized secondary schools.

Table 2 also indicates a decreasing pattern in the shares of applicants and those being admitted (conditional on application) when comparing graduates with and without a local university. For example, 53% of graduates from specialized secondary schools living near a local university apply in comparison with 47% of those living far away. Restricting our attention to admission, we observe a lower share of admission for applicants living far from a university than the share for those living near a university (0.36 vs. 0.41 for specialized secondary schools and 0.64 vs. 0.66 for gymnasiums). The admission decision does not depend on the level of potential study cost and therefore heterogenous information can be a reason for admission shares depending on the distance to the nearest university. The reported differences in admission shares, however, need not prove the presence of heterogenous information since the shares are not conditional on other characteristics. Differences can result also from differences in ability, socio-economic background, etc. for the two residence types. Note that a worse socio-economic background (parental education, computer ownership) and a lower level of observable cognitive skills (composite score) appear for those graduates and applicants living far from a university (see Table 2). The effect of heterogenous information and other observable characteristics is examined by the econometric analysis in Section 5.

Application strategies—program specializations

The costs of study and heterogenous information can influence not only the decision about application but also the choice of university programs. Examining differences in applicants' revealed preferences for university programs according to residence type

can shed some light on the effect of a local university on the secondary school graduate's application decision.

Table 3 presents the shares of applicants who apply to one program specialization only (not necessarily at one university). Lower study costs affect the choice of the programs of applicants with local university who apply locally. To filter out the study cost effect and examine the effect of heterogeneous information we focus on those applying to a non-local university(ies) only (i.e. applicants with a local university applying to a non-local university only and applicants without a local university). All applicants in this subgroup face high potential costs of university study. Applicants living near a university tend to stick to one program specialization more than those living far away (28% vs. 20% of applicants from gymnasiums). The difference is even higher for specialized secondary school graduates (63% vs. 53%). The result suggests that applicants from a university neighborhood are better informed about university programs and thus have a more concrete idea about what to study. The reason could also be that applicants living far from a university prefer programs that demand broader knowledge so they can also apply to related programs. The preferences for various programs are explored in Tables 4a and 4b.²³

Table 4a shows that some programs are demanded more by applicants living far from a university (e.g. Education, Technical Chemistry) and by applicants living near a university (e.g. Theory and History of Art, Philosophy, Engineering). The total share of applicants without a local university is 0.45 for applicants from gymnasiums and 0.46 for applicants from specialized secondary schools. The column *Difference I* in Table 4a indicates the difference between the total share of applicants without a local university and the share of applicants without a local university for a particular program. So, positive figures suggest more than the average share of applicants from locations without universities and vice versa. The table also provides the number of programs and districts where it is possible to study a particular program together with the probability of admission to that program.

²³ Note that all the results presented in this section are unconditional.

In Table 4a, we compare the preferences of all applicants and applicants applying to a non-local university(ies) only. Similarly to previous paragraphs we argue that the second group of applicants faces the same (high) potential costs of study and therefore heterogenous information as a reason for differences can be identified. According to Table 4a, the variation of differences decreases markedly when we focus on applicants that apply only to non-local universities (see columns *Difference I* and *Difference II*). Therefore, the revealed preferences for university programs are affected by the costs of study and consequently by programs provided by local universities.

Table 4b provides a comparison of cognitive skills (composite scores) and non-cognitive skills (relative excess demand for gymnasiums in a district) of applicants to selected programs by residence type and by the location of the university they apply to. It follows that gymnasium graduates living near a university and applying only non-locally are, in the terms of composite scores, smarter than their colleagues who apply only locally. Interestingly, the opposite is true in the case of specialized secondary school graduates.

Furthermore, the table suggests that the relation between the composite scores of applicants living near a university and applying only locally and those living far from a university relates to differences in the revealed preferences of applicants (column *Difference I*). Programs demanded relatively more by those without local university (positive numbers in column *Difference I*) are demanded more on average by smarter graduates who live far in comparison with those living near a university (higher composite score for applicants from locations without a university than the score for applicants applying locally only) and vice versa. This result suggests different perceptions of various programs from different groups of graduates.

Finally, Table 4b implies that applicants residing far from a university exhibit a higher district level of non-cognitive skills for almost all programs. Therefore, applicants living far from a university are on average more motivated and persistent than those living near a university. Again, this could be a result of information heterogeneity.

Application strategies—probability of admission

Table 5 suggests that the average applicant applying only to a local university faces a higher probability of being admitted than an average applicant living far from a university (the difference is around 2% for applicants from gymnasiums and 5% from specialized secondary schools). This result holds even if we condition on cognitive skills (composite score) and the district level of non-cognitive skills (relative excess demand for gymnasiums in a district).²⁴ The different strategies regarding the probabilities of admission correspond to the different strategies with respect to university programs discussed in the previous subsection.

5 Estimation results

Our empirical analysis proceeds in two steps. First, we explore whether individuals residing in close proximity to a university are more likely to apply to any university and we discuss the main determinants that influence the application decision. In the second step, we estimate the probability of admission conditional on application.

We start estimating the reduced form model of applying to a university (1) separately for gymnasium and secondary school graduates. We estimate equation (1) as a logit model clustering data by class.²⁵ Table 6 reports estimation results.²⁶ Reported are marginal effects.

The local university dummy variable is not significant for either type of secondary school graduate,²⁷ indicating the decision about application is influenced neither by the

²⁴ We divided the population of applicants into 16 groups defined by quartiles of composite score and demand for gymnasiums in a district and compare the average probability of admission for applicants living near and far from a university.

²⁵ Moulton (1990) argues that individuals from the same socio-economic background (secondary school, class) could share the same unobservable characteristics. The disturbances of such groups of individuals are then correlated and we take the possibility of clustering into account.

²⁶ Logistic regression diagnostics: we find that the model is correctly specified (specification error test—*linktest* in Stata) and it fits the data well in the case of specialized secondary schools (Hosmer and Lemeshow's test—*lfit* in Stata). The model's performance is worse for gymnasiums. Finally, we do not detect any multicollinearity problems (command *collin* in Stata).

²⁷ The result is robust to changes in the travel time threshold that defines the local university dummy.

direct costs of study nor by heterogeneous information affecting the application decision. This conclusion assumes that the lower cost of study and heterogeneous information act in the same direction, i.e. both lower the probability of applying for graduates living far from a university. Both the system of financial support of university students (e.g. dormitories) and the information availability seem to be sufficient to equalize the differences in the probability of applying caused by the presence/absence of a local university.

Furthermore, Table 6 suggests that in addition to individual skills also gender and individual socio-economic background (parental education, computer ownership) are significant determinants of the application decision. For example, the average female graduate from a specialized secondary school faces a 10% lower probability of applying to a university than the average male. This is a consequence of gender segregation on the level of secondary school field specialization. For example, only 4% (128 out of 3131) of graduates from secondary schools specialized in health care are male. The study programs of such secondary schools are not intended to prepare students for university study. Therefore, graduates from such schools usually do not go on to a tertiary level of education (only one third of them apply).

The estimated impact of regional (district) characteristics suggests links between the local economic and living conditions and graduates' behavior regarding application to university. Higher unemployment leads to a higher probability of applying, which is in line with the lower opportunity costs of university study in districts exhibiting high unemployment. The effect is stronger for graduates from specialized secondary schools. Our interpretation is that they have specific skills and thus are more sensitive to unemployment changes. Similarly, higher regional economic growth lowers the incentives to go on with study. Again, higher growth increases opportunity costs and graduates (especially from specialized schools) tend to enter the labor market immediately after graduation from secondary school.

Moreover, Table 6 shows that the propensity to apply is positively affected by the local share of the tertiary-educated population. Each percentage point of the share of tertiary-

educated population in a district²⁸ accounts for at least a 0.31 percentage point higher probability of applying for gymnasium graduates (1.14 percentage points for graduates from specialized secondary schools). This result suggests that the source of heterogenous information we attempt to detect by the local university dummy is not the distance to a university but rather the local environment created by the highly educated population.²⁹

Finally, Table 7 reports the marginal effects of dummy variables for subjects taken for the maturita exam (gymnasiums) and the field of school (specialized secondary schools) on application decision. Subjects taken for the maturita exam and the fields of secondary school determine the choice of the prospective field, which subsequently determines the costs and benefits of the schooling decision. Therefore, almost all the dummies are significant.³⁰

Derivations presented in Appendices A and B yield that structural parameters are in absolute value lower than the reduced form model estimates (a similar conclusion holds for marginal effects). Basically, this is a consequence of the fact that the prospective applicant considers whether to apply under a certain probability of being admitted to a particular university department. Thus the magnitudes of the effect of a particular determinant are lower (in absolute value) than estimated. So, for example, a 10% decrease in the probability of applying for the average female from a specialized secondary school is the upper bound of the real effect.

In Appendix B, we also derive that the standard error of the estimated coefficient of the local university dummy variable is lower in the reduced form model than when we are able to estimate a structural model. The local university dummy variable is, therefore, insignificant also in the structural model.

²⁸ The standard deviation of the percentage share of the tertiary-educated population in a district is 2.1.

²⁹ Correlation coefficient between local university dummy variable and the share of tertiary-educated population in a district is 0.45.

³⁰ For each field of specialized secondary school approximately half of the applicants apply to the corresponding program at a university (e.g. 57% applicants from secondary schools specialized in light manufacturing apply to technical programs.).

In the second part of the econometric analysis we present the estimation results of the admission regression equation (2). Here, the unit of observation is an application. We estimate equation (2) as a logit model clustering data by individuals.³¹ The model of admission derived in Appendix C suggests several specifications of the regression equation (2). The estimated marginal effects of the considered specifications are reported in Table 8 for gymnasiums and in Table 9 for specialized secondary schools. Marginal effects that are robust to specification are highlighted. The robustness is understood in the sense that the estimated marginal effects do not change signs or significance under various specifications of the regression equation. In the following we discuss only the robust estimates that are shown in Table 10.

The coefficient on the dummy variable indicating the applicant's residence within commuting distance to a university that provides the program an applicant is applying to, f_j , is positive and significant. For specialized secondary schools the dummy variable for the presence of a local university d_j is not significant. This implies that living near an arbitrary university need not necessarily help in the admission process. Rather living near a university that provides the applicant's preferred program increases the admission probability to that program even if the applicant applies to that program at a different university. Information that brings an advantage in the admission process is, therefore, "program specific". For example, face-to-face contacts with older students who passed the admission process successfully, easier access to preparatory courses organized by local universities or extra information provided by secondary school teachers or counselors who are experienced with programs provided by local universities can be beneficial for local graduates.

Universities which do not experience high excess demand³² usually admit the vast majority of applicants and thus one cannot expect that some applicants are advantaged on the grounds of access to information. On the other hand, in the case of highly oversubscribed university programs³³ additional information can provide an advantage to an applicant. To test the dependence of the impact of heterogeneous information on

³¹ Post-estimation diagnostics suggests model specification problems for applications from specialized secondary schools.

³² 10 out of 42 programs admitted more than 70% of applicants from secondary schools in 1998.

³³ 4 out of 42 programs admitted less than 10% of applicants from secondary schools in 1998.

relative excess demand for a university we interact a dummy variable indicating the presence of the desired program at a local university, f_j , with the probability of admission to that program. The interaction term is significant for gymnasium graduates. So, living near a university with the preferred program increases the probability to be admitted for the average applicant from a gymnasium by 9.9%. Moreover, the lower the probability of admission to a university program (i.e. higher excess demand for the program) an average applicant applies to, the higher the advantage premium provided by a local university.

Other estimated marginal effects suggest the importance of individual characteristics (parental education, composite score) for the probability of being admitted. For the selected marginal effects of individual, class, and regional characteristics see Table 10.

Heterogenous information that underlies relative success in the admission process is not distributed through the official study program at secondary schools. Gymnasiums provide general education not specialized in any particular fields of study. This need not be the case for the specialized secondary schools (transportation, health, teaching, engineering, etc.). Specialized secondary schools are, however, very evenly distributed across the Czech Republic and there are more than five schools in 70% of the districts. Thus, the geographical distribution of specialized secondary schools does not drive differences in admission probabilities according to residence type for graduates from specialized secondary schools.³⁴

³⁴ We distinguish 25 different study fields of specialized secondary schools. Almost 50% of schools are of business specialization, followed by engineering and electronic specializations with shares of 10% and 8%, respectively. Focusing on specialized secondary schools with very narrow specialization we look at whether these schools are established close to a university with the same specialization and whether graduates of such secondary schools apply for these faculties. We find that highly specialized secondary schools are not usually established near a university with the same program specialization. For example, veterinary medicine at the tertiary level can be studied only in Brno, however there are six secondary schools (212 students/60% apply) of veterinary specialization, none of them located in Brno. Only 36% of graduates apply to Brno. Some graduates (25% of applications) apply to a similar program like biology or agriculture. Next, there are four schools with a specialization in silicate chemistry (Tábor, Karlovy Vary, Jablonec nad Nisou, Česká Lípa). Almost 90% of all graduates who choose to apply (40% apply), decide for tertiary education with a specialization in silicate chemistry. The universities with this program are located in Prague (65% of applications) and Pardubice (35% of applications). None of secondary schools specialized in silicate chemistry are located near Prague or Pardubice.

In our econometric analysis, we find that heterogenous information due to the presence/absence of a local university does not influence the application decision but does influence the admission decision. The information spread in a university neighborhood is relevant for the applicant's performance in admission. This kind of information does not contribute to a graduate's decision about application. Results suggest that it is a different kind of information that positively affects the secondary school graduate. Graduates living in an environment that is characterized by a high share of tertiary-educated population tend to apply more. So, information that is disseminated by a highly educated environment constitutes heterogeneity in information for secondary school graduates with regard to their application decision.

7 Conclusion remarks

We analyze the demand for tertiary education in the Czech Republic. We develop a structural model of the post-secondary decision process of a secondary school graduate and a model of admission to a university. The reduced form models are estimated employing data from 1998. The estimation results of both application and admission equations suggest a significant role of the environment around a secondary school graduate in the process of entering the tertiary level of education. In the first stage, applying, those living in conditions that are characterized by high local shares of tertiary-educated population exhibit a higher probability of applying to a university. Living near a university does not influence the potential applicant's decision about application—neither because of lower study costs nor because of more information regarding university study available. In the second stage, admission, we observe that for certain program specializations there is a higher probability of being admitted for applicants with a local university that provides programs with the same program specialization. The effect is stronger in the case of highly oversubscribed programs.

The policy question that follows our findings is how to ensure equal conditions regarding the process of entering tertiary education by secondary school graduates in the Czech Republic. Equal conditions usually mean that all secondary school graduates can apply for university study—so there are no constraints based, for example, on sex,

race, or application fee that would prevent a particular group of potential university students from entering the tertiary level of education. However, our analysis uncovered some factors that underlie differences between secondary school graduates with respect to their chance of applying and being admitted. These factors relate to individual characteristics, socio-economic background, regional economic and social characteristics, and characteristics of the secondary school.

The question arises here what factors should be considered as those reflecting some inequality. So, for example, different probabilities of applying and being admitted for students with different levels of ability to study should not be viewed as a problem of unequal access to tertiary education. On the other hand, some factors have nothing to do with the individual ability to study and still influence the admission procedure. For example, we found that female students from specialized secondary schools face a remarkably lower (by 10%) probability of applying, which is caused by gender segregation by fields of secondary schools. We should therefore examine why some fields of secondary school prepare students for university study less than others.

In this paper, we focused on the effect of the presence of a local university on a secondary school graduate's prospects regarding the university study. We found that a local university can constitute an advantage in the admission process for those living near the local university. To equalize the chance of admission policy makers should consider expanding the system of universities. Moreover, we found that the advantage concerns the university programs that are offered by the local university. The expansion of universities should be, therefore, accompanied by the expansion of university programs. Alternatively, equal chances of entering tertiary education could be achieved also by the improvement of information spread since we detected that it is the information emitted by a local university that provides the advantage. We suggested that such information can be spread, for example, by face-to-face contacts with university students. In this paper, we do not examine the nature of such information in detail. So, future research is needed to elaborate the essence of such information and to answer the question whether information availability is an adequate alternative for an expansion of the system of institutions providing tertiary education.

References

- Andersson R., Quilgey J.M., Wilhelmsson M. (2001), University decentralization as regional policy: The Swedish experiment, *Journal of Economic Geography* 4, 371-388.
- Bound J., Groen J., Kezdi G., Turner S. (2004), Trade in university training: cross-state variation in the production and stock of college-educated labor, *Journal of Econometrics* 121, 143-173.
- Becker G. S. (1964), *Human Capital*, Columbia University Press.
- Beeson P., Montgomery E. (1993), The effects of colleges and universities on local labor markets, *The Review of Economics and Statistics* 75, 753-761.
- Brooks R. (2002), Edinburgh, Exeter, East London – or employment? A review of research on young people's higher education choices, *Educational Research*, 44(2), 217-227.
- Card D. (1995), Using geographic variation in college proximity to estimate the return to schooling, in *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, Toronto, ON: University of Toronto Press.
- Card D. (1999), The causal effect of education on earnings, in *Handbook of Labor Economics*, Volume 3, O. Ashenfelter and D. Card, eds., North-Holland, Amsterdam.
- Do C. (2004), The effects of local colleges on the quality of college attended, *Economics of Education Review*, 23, 249-257.
- Drnáková L. (2006), Determinants of secondary school choice in the Czech Republic, CERGE-EI WP number 341.
- Eliasson K. (2006), The effects of accessibility to university education on enrollment decisions, Geographical Mobility, and Social Recruitment, *Scandinavian Working Papers in Economics*, number 690.
- Fidrmuc J. (2004), Migration and regional adjustments to asymmetric shocks in transitions economies, *Journal of Comparative Economics*, 32, 230-247.
- Frenette M. (2002), Too far to go on? Distance to school and university participation, Statistics Canada, 191.

- Frenette M. (2003), Access to College and University: Does Distance Matter? Statistics Canada, 201.
- Giannetti M. (2002), The effects of integration on regional disparities: Convergence, divergence, of both? *European Economic Review*, 46, 539-567.
- Giannetti M. (2003), On the mechanics of migration decisions: skill complementarities and endogenous price differentials, *Journal of Development Economics*, 71, 329-349.
- Glaeser E.L., Scheinkman J.A., Shleifer A. (1995), Economic growth in a cross-section of cities, *Journal of Monetary Economics* 36, 117-143.
- Green W.H. (2000), *Econometric Analysis* (Fourth Edition), Prentice-Hall, Inc.
- Gullason E. T. (1989), The consumption value of schooling: An empirical estimate of one aspect, *Journal of Human Resources*, 24(2), 287-298.
- Heckman J. J., Stixrud J., Urzua S. (2006), The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior, *Journal of Labor Economics*, 24(3), 411-482.
- Institute for Information on Education (2002), Institute for Information on Education, Ministry of Education, Youth and Sport in the Czech Republic, Towards the society of knowledge, in Czech.
- Jurajda Š., K. Terrell (2008), Regional unemployment and human capital in transition economies, CERGE-EI WP number 345.
- Kaldewei C., Walz U. (2001), The determinants of regional growth in Europe: An empirical investigation of regional growth models.
- Lange F., Topel R. (in press), The social returns to education and human capital, *Handbook of Education Economics*.
- Long B.T. (2004), How have college decisions changed over time? An application of the conditional logistic choice model, *Journal of Econometrics* 121, 271-296.
- Makovec M. (2005), Does it pay to study far from home? Explaining the returns to geographic mobility of Italian college graduates, unpublished manuscript
- Matějů P. (2007), Public-private funding of higher education: A social balance – The Czech case study on public and private costs for higher education: Results from the first phase, Institute of Sociology of the Academy of Sciences of the Czech Republic.

Moretti E. (2004), Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data, *Journal of Econometrics*, 121, 125-212.

Moulton B.R. (1990), An illustration of a pitfall in estimating the effects of aggregate variables on micro units, *The Review of Economics and Statistics*, Vol. 72(2), 334-338.

Overman H. G., Puga D. (2002), Unemployment clusters across Europe's regions and countries, *Economic Policy*, 117-147.

Sá C., Florax R.J.G.M., Rietveld P. (2004), Does accessibility to higher education matter? Choice behavior of high school graduates in the Netherlands, *Tinbergen Institute Discussion Paper*, 04-061/3.

Spence M. (1973), Job Market Signaling, *Quarterly Journal of Economics*, 87, 355-374.

Whitehead J.M., Raffan J., Deany R. (2006), University Choice: What Influences the Decision of Academically Successful Post-16 Students?, *Higher Education Quarterly*, 60(1), 4-26.

Willis R. J., Rosen S. (1979), Education and self-selection, *Journal of Political Economy*, 87(5), S7-36.

Appendix A: Structural and reduced form models of applying

A.1 Structural model of applying

In this section we introduce the model of the application decision. The model combines standard models of schooling decision based on human capital theory and expected utility theory.

A secondary school graduate makes a decision whether to apply to a university or not by comparing the expected utility of those two actions. We denote the individual's expected utility of applying as U^a and the expected utility of non-applying as U^{na} . The secondary school graduate chooses to apply if $U^a > U^{na}$.

The unit a graduate can apply to is a university program, in the following text denoted simply as program. Suppose there are $F > 0$ programs. Further, an applicant spends one whole day participating in an admission procedure for a particular university (usually a written/oral test) and thus it is not possible to participate in another admission test if they are organized in the same day. All admission procedures (tests) are undertaken in a few days $T > 0$ during the year. Furthermore, an applicant to university program j bears admission costs $AC_j > 0$ (admission fees, courses for preparing the student for admission tests, travel costs concerning the admission procedure, etc.). Each student has a certain amount of money $M > 0$ that can be used for financing her admission process.

If a graduate applies to just one program then the expected utility from applying to that program equals $U^a = p_1(B_1 - C_1) + \sigma(M - AC_1)$, where p_1 is the probability of being admitted to the program conditional on application, B_1 and C_1 denote the present values of benefits and costs, respectively, of being admitted to the program. Finally, the coefficient σ represents how much disutility is related to admission costs with respect to the utility given by the expected benefits and costs.

Similarly, if an individual applies to two programs the expected utility is given by:

$$U^a = p_1(1-p_2)(B_1 - C_1) + p_2(1-p_1)(B_2 - C_2) + p_1p_2 \max_{j \in \{1,2\}} \{B_j - C_j\} + \sigma(M - AC_1 - AC_2).$$

We basically divide the situation of the individual applying to two programs into three mutually exclusive events. An applicant is admitted either to the first program only (with probability $p_1(1-p_2)$), or to the second program only (with probability $p_2(1-p_1)$), or to both programs (with probability p_1p_2).³⁵ However, we assume that an individual can enroll in only one program. Therefore, if she is admitted to both programs, the program with the higher present value of the expected benefits net of costs is preferred.

In general, a secondary school graduate chooses to apply to such programs to maximize his expected utility taking into account time and budget constraints. So, an applicant decides whether to apply to a program j ($t_j = 1$) or not ($t_j = 0$). The optimization problem takes the following form:

$$U^a = \max_{\{t_1, \dots, t_F\}} \left\{ \sum_{k=1}^F \left[\sum_{A \in C(N,k)} \left(\left(\prod_{j \in A} t_j p_j \right) \left(\prod_{j \in N \setminus A} (1-t_j p_j) \right) \right) \max_{j \in A} \{B_j - C_j\} \right] + \sigma \left(M - \sum_{j=1}^F t_j AC_j \right) \right\} \quad (\text{A1})^{36}$$

$$\sum_{j=1}^F t_j \leq T \quad (\text{A2})$$

$$\sum_{j=1}^F t_j AC_j \leq M \quad (\text{A3})$$

$$t_j \in \{0, 1\}, j = 1, \dots, F. \quad (\text{A4})$$

The maximized function is a generalization of the case for one or two programs discussed above. The time constraint (A2) captures the fact that the admission process takes place during a few days during the year and it is possible to take part in one

³⁵ The form of compounded probabilities implicitly involves a reasonable assumption that the probability of being admitted to the first program does not affect the probability of being admitted to the second program. The two events are statistically independent.

³⁶ $N \equiv \{1, \dots, F\}$, $C(N, k)$ is a set of combinations of size k from the set N .

admission process a day only.³⁷ The budget constraint (A3) captures the limitations given by admission costs and the individual's disposable income.

The present values of benefits and costs related to university program j are denoted by B_j and C_j , respectively. The variable p_j denotes the graduate's subjective estimate of the probability of being admitted to program j conditional on application.³⁸ An individual can infer the probability in various ways, such as from admission probabilities in previous years (published every year) and from her performance at secondary school in comparison with schoolmates.

The model (A1)-(A4) captures several phenomena observed in reality. Table 5 reports a high negative correlation (almost -0.9) between the average number of applications of an applicant and the average probability of admission (conditional on application) to programs an applicant applies to.³⁹ In the model framework, if the individual's probability of being admitted to a program is close to one, he/she will not apply for another program since the marginal utility from sending another application could be lower than the disutility related to admission costs (i.e. the probability of being admitted to the additional program and not to being admitted to already considered programs could be so low that admission cost multiplied by σ could exceed the *expected* net benefits). So, if the graduate's estimate of the probability of admission is low, she sends an additional application. This yields a negative relationship between the average number of applications and the probability of being admitted.

³⁷ Here we ignore the possibility that the admission tests of an individual's desired programs may take place on the same day.

³⁸ We assume that the probability of being admitted to program j conditional on application (p_j) is a primitive of the problem. If we assume that the primitive is a probability of being admitted to a program along with not being admitted at another university program conditional on application (\tilde{p}_j) then the maximized function would take the simple form:

$$\sum_{j=1}^F t_j \tilde{p}_j (B_j - C_j) + \sigma \left(M - \sum_{i=1}^F t_i A C_i \right).$$

³⁹ In Table 5, we divide the applicants according to secondary school type (gymnasium, specialized secondary school), residence type (with/without local university), and the location of the university they apply to (local only/non local only/both). This yields eight subgroups for the comparison of the number of applications and the probability of admission.

So far, we have discussed the expected utility U^a of applying. In the case of the expected utility of non-applying U^{na} an individual compares the expected benefits and the costs of not applying. Non-apppliers can enter the labor market or stay out of the labor market. We put these possibilities together into one outside option. The expected utility of non-applying is related to regional labor market prospects and individual and secondary school characteristics.

A.2 Reduced form model of applying

The optimization problem of a prospective university applicant (A1)-(A4) implies that the individual's expected utility of applying equals

$$U^a = \sum_{k=1}^F \left[\sum_{A \in C(N,k)} \left(\left(\prod_{j \in A} t_j p_j \right) \left(\prod_{j \in N \setminus A} (1 - t_j p_j) \right) \right) \max_{j \in A} \{B_j - C_j\} \right] + \sigma \left(M - \sum_{j=1}^F t_j A C_j \right) \quad (\text{A5})$$

for some nonzero t_j 's $j \in \{1, \dots, F\}$.

We denote the individual's present value of benefits from being admitted to a program j net of the present value of costs as u_j , i.e.

$$u_j = B_j - C_j. \quad (\text{A6})$$

The literature suggests many factors that influence the present value of costs and benefits related to participation in university study. A detailed discussion of the variables used can be found in Sá et al. (2004), Long (2004), Whitehead et al. (2006), and Brooks (2002).

Besides individual characteristics, also university characteristics significantly influence the benefits and costs of study at a university (Long 2004). According to the human capital approach, benefits reflect the present value of extra earnings that an individual earns from additional year(s) of schooling. We use dummy variables for the university programs an applicant applies to FS_k $k \in \{1, \dots, S\}$ as a proxy for future extra earnings. Regarding the costs of study that are driven by university, distance to university is a

principle determinant. Since there is no tuition paid at universities in the Czech Republic, moving (travel) costs are the substantial determinant of study costs. Those who live far from a university have to bear them. Those who live near a university face a low-cost option of university study. We therefore use a dummy variable indicating the presence of a local university d as a proxy for study costs. The distance dummy is of primary interest since it reveals several phenomena present in the process of entering a university. In addition to the effect of study costs, it captures also the effect of heterogenous information on the prospective applicant's behavior.

Finally, Sá et al. (2004) note that the benefits and costs of university study are related to expected employability and expected earnings in the region of parental household (spatial heterogeneity). Therefore we include also regional characteristics R into our analysis. Furthermore, Sá et al. (2004) point out that localized social interaction among secondary school students plays an important role. Therefore we add secondary school (class) characteristics C into the analysis.

Taking into account the above mentioned literature on the determinants of post-secondary schooling decisions, we model the present value u_j by a linear predictor containing the following set of variables:

$$u_j = \alpha I + \beta C + \sum_{k=1}^K \eta_k FS_k + \delta d + \varphi R, \quad (A7)$$

where vector I includes individual characteristics, vector C includes secondary school (class) characteristics, FS_k $k \in \{1, \dots, K\}$ are dummy variables for the university program an applicant applies to, and d is a dummy variable for local university. Vector R includes regional characteristics.

Combining (A5)-(A7) we obtain the general formula for the expected present value of applying to a tertiary level of education:

$$U^a = \sum_{k=1}^F \left[\sum_{A \in C(N,k)} \left(\left(\prod_{j \in A} t_j p_j \right) \left(\prod_{j \in N \setminus A} (1 - t_j p_j) \right) \right) \max_{j \in A} \{ \alpha I + \beta C + \sum_{k=1}^K \eta_k FS_k + \delta d + \varphi R \} \right] + \sigma \left(M - \sum_{j=1}^F t_j AC_j \right) \quad (\text{A8})$$

for some nonzero t_j 's $j \in \{1, \dots, F\}$. For instance, if an applicant maximizes her utility by applying to just one program (with the program indexed by 1) the formula takes the following form:

$$U^a = p_1 (\alpha I + \beta C + \delta d + \varphi R) + p_1 \eta_1 FS_1 + \sigma (M - AC_1). \quad (\text{A9})$$

If an applicant sends two applications (programs 1 and 2) the formula takes the form:

$$U^a = (p_1 + p_2 - p_1 p_2) (\alpha I + \beta C + \delta d + \varphi R) + p_1 (1 - p_2) \eta_1 FS_1 + p_2 (1 - p_1) \eta_2 FS_2 + p_1 p_2 \max_{j \in \{1,2\}} \{ \eta_j FS_j \} + \sigma (M - AC_1 - AC_2) \quad (\text{A10})$$

If an applicant applies to two programs the formula takes the form:

$$U^a = (p_1 + p_2 - p_1 p_2) (\alpha I + \beta C + \delta d + \varphi R) + (p_1 + p_2 - p_1 p_2) \eta_1 FS_1 + \sigma (M - AC_1 - AC_2). \quad (\text{A11})$$

In general, the first part of the formula containing vectors I, C, d, R does not change (up to the multiplier representing the probability of being admitted). The difference between formulas consists of the part that contains program dummies and the sums of admission costs.

If we drop out admission costs (admission costs are low with respect to disposable income and very similar across universities) the expected utility of applying takes the following form:

$$U^a = \hat{p}_{(i)} (\alpha I + \beta C + \delta d + \varphi R) + \sum_{k=1}^K \gamma_{k,(i)} \eta_k FS_k^* + \sigma M, \quad (\text{A12})$$

where $\hat{p}_{(i)}$ is the individual's subjective probability of being admitted at least to one program and $\gamma_{k,(i)}$ are multinomial of probabilities.

In formula (A12) we have to deal with two econometric issues. First, the variables denoted by an asterisk are not observable for graduates who choose not to apply. Therefore, we proxy the program dummies by dummies for subjects that students take for the maturita exam in gymnasium or by dummies for the field of their specialized secondary schools denoted by $SD_s, s \in \{1, \dots, S\}$.⁴⁰

The second econometric issue concerns the problem of parameter heterogeneity. Coefficients denoted by subscript (i) differ across individuals. Moreover, the coefficients $\hat{p}_{(i)}$ and $\gamma_{(i)}$ are unobservable and they are, in general, correlated with the variables they are multiplied with. On the other hand, all the coefficients are numbers between zero and one. In Appendix B, we derive the relationship between the maximum likelihood estimates of a logit model when we take the individual coefficients $\hat{p}_{(i)}$ and $\gamma_{(i)}$ into account and when we don't. We use the results derived in the Appendix in the discussion on the relationship between structural parameters and the reduced form parameters in the section describing estimation results.

The reduced form of the model for the utility of applying is

$$U_{red}^a = \alpha^* I + \beta^* C + \delta^* d + \varphi^* R + \sum_{s=1}^S \mu_s SD_s + \sigma M + \varepsilon^a, \quad (\text{A13})$$

where the disturbance term captures, for example, the different preferences of individuals.

⁴⁰ In the main text, we discuss how the subjects taken for the maturita exam (fields of specialized secondary school) determine the program specialization at a university.

The non-applying options consist of entering the labor market, or staying out of the labor market. We assume that the econometric model of the expected utility of not applying (U^{na}) takes the following form:

$$U_{red}^{na} = \tilde{\alpha}I + \tilde{\beta}C + \tilde{\varphi}R + \sum_{s=1}^S \tilde{\mu}_s SD_s + \varepsilon^{na}. \quad (A14)$$

If the disturbances $\varepsilon^a, \varepsilon^{na}$ have an extreme value distribution of type I then the difference of disturbances yields a logistic distribution. So, the probability of applying is modeled as:

$$\begin{aligned} \Pr(\text{Applying}) &= \Pr(U^a > U^{na}) = \Pr(U^a - U^{na} > 0) = \\ &= \Pr\left(\varepsilon^{na} - \varepsilon^a < \alpha^*I + \beta^*C + \delta^*d + \varphi^*R + \sum_{s=1}^S \mu_s SD_s + \sigma M - \tilde{\alpha}I - \tilde{\beta}C - \tilde{\varphi}R\right), \quad (A13) \\ &= \Lambda\left(I(\alpha^* - \tilde{\alpha}) + C(\beta^* - \tilde{\beta}) + R(\varphi^* - \tilde{\varphi}) + \delta^*d + \varphi^*R + \sum_{s=1}^S \mu_s SD_s + \sigma M\right) \end{aligned}$$

where Λ is a logistic distribution function.

Appendix B: MLE estimates based on two dependent samples

In this section we derive the relationship between the maximum likelihood estimates of coefficients of the logit model in the case of a systematic change of the size of the explanatory variables.

Let us assume a logit model:

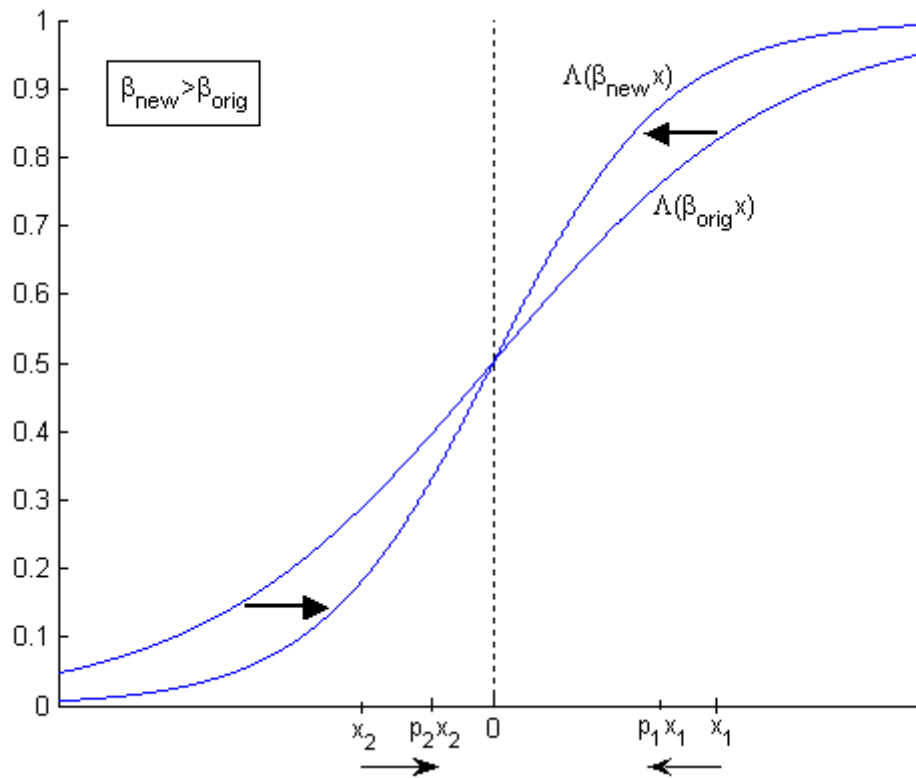
$$P(y = I) = \Lambda(\alpha + \beta_1 x_1 + \dots + \beta_k x_k). \quad (\text{B1})$$

Let there be two random samples:

$$S_1 = \{ [y_i, x_{1i}, \dots, x_{ki}], i \in 1, \dots, n \} \text{ and } S_2 = \{ [y_i, p_{1i} x_{1i}, \dots, p_{ki} x_{ki}], i \in 1, \dots, n \},$$

where $p_{li} \in (0, 1)$, $l \in 1, \dots, k$, $i \in 1, \dots, n$. Our goal is to derive the relationship between the maximum likelihood estimates of coefficients (or marginal effects) of the logit model (B1) employing samples S_1 and S_2 . The aim of this exercise is to find out the relative sizes of the MLE coefficients (and consequently marginal effects) and their standard errors. More precisely, we want to derive the change in the size of a coefficient caused by the systematic change in the size of an explanatory variable (both the variable that corresponds to the coefficient and that doesn't).

The case of one explanatory variable is depicted by the following figure. We consider a logit model estimated by MLE using two samples S_1 and S_2 . The movement of x 's towards the origin causes the movement of the logit curve in the direction depicted by arrows. The coefficient in the linear term underlying the new logit curve increases.



For the multi-dimensional case we derive the relationship using the maximum likelihood conditions of the logit model resulting from the first order conditions of the likelihood maximization problem (see e.g. Green 2003):

$$\sum_{i=1}^n (y_i - \Lambda_i) = 0$$

$$F_1(\beta; x_1, \dots, x_k, y) \equiv \sum_{i=1}^n (y_i - \Lambda_i) x_{1i} = 0$$

...

$$F_k(\beta; x_1, \dots, x_k, y) \equiv \sum_{i=1}^n (y_i - \Lambda_i) x_{ki} = 0,$$

where $\Lambda_i \equiv \Lambda(\alpha + \beta_1 x_{1i} + \dots + \beta_k x_{ki})$ is a logistic distribution function. The implicit function theorem gives

$$\frac{\partial \beta_l}{\partial x_l} = -\frac{\frac{\partial F_l}{\partial x_l}}{\frac{\partial F_l}{\partial \beta_l}} \quad \text{and} \quad \frac{\partial \beta_l}{\partial x_m} = -\frac{\frac{\partial F_l}{\partial x_m}}{\frac{\partial F_l}{\partial \beta_l}}. \quad (\text{B2})$$

The following calculations are mainly for illustrative purposes; we do not intend to provide an exact mathematical derivation. Therefore, we suppress the discrete property of the original conditions and we take standard partial derivatives.⁴¹

Some rearrangements of the terms in the fractions (B2) yields:

$$\frac{\partial F_l}{\partial \beta_l} = \sum_{i=1}^n -\Lambda_i x_{li}^2 < 0 \quad (\text{B3})$$

$$\frac{\partial F_l}{\partial x_l} = \sum_{i=1}^n (y_i - \Lambda_i) - \sum_{i=1}^n x_{li} \lambda_i \beta_l = -\beta_l \sum_{i=1}^n x_{li} \lambda_i \quad (\text{B4})$$

$$\frac{\partial F_l}{\partial x_m} = -\beta_m \sum_{i=1}^n x_{li} \lambda_i, \quad (\text{B5})$$

where λ_i is a logistic density function.

So, the signs of the derivatives in (B2) are solely driven by the signs of the terms (B4) and (B5). The direction of the change of the coefficient caused by the systematic change in the explanatory variables ($\partial \beta_l / \partial x_{l(m)}$) is, therefore, determined by the signs of the original coefficients ($\beta_{l(m)}$) and explanatory variables. If all explanatory variables are positive and the original coefficient is positive then derivatives in (B4) and (B5) are negative and derivatives in (B2) are also negative. So, if a new sample arises from the original one by a systematic lowering of the values of explanatory variables (the explanatory variable does not change) then the new coefficient is larger than the original coefficient. A similar conclusion holds for marginal effects since the marginal effect for the logit model can be expressed as a coefficient multiplied by a nonnegative number (the probability density function evaluated in means of variables).

⁴¹ On the other hand, we still use sums and not integrals.

In addition to the change in marginal effects, we also derive the change in standard errors. Basically, we focus on the change of the diagonal terms of the Hessian matrix since these terms determine the standard errors of MLE estimates. The diagonal terms of the Hessian matrix of the logit model (B1) take the following form:

$$[H]_{ll} = -\sum_{i=1}^n \Lambda_i (1 - \Lambda_i) x_{il}^2, \quad l \in \{1, \dots, k\}.$$

Then

$$\begin{aligned} \left[\frac{\partial H}{\partial x_{\bullet l}} \right]_{ll} &= -\sum_{i=1}^n (\lambda_i (1 - \Lambda_i) x_{il}^2 - \Lambda_i \lambda_i x_{il}^2 + 2\lambda_i x_{il}) = \\ &= -\sum_{i=1}^n (\lambda_i x_{il}^2 - 2\Lambda_i \lambda_i x_{il}^2 + 2\lambda_i x_{il}) = -\sum_{i=1}^n \lambda_i x_{il} (x_{il} - 2\Lambda_i x_{il} + 2) \end{aligned},$$

where we use the fact that in the logistic distribution function holds $\lambda_i = \Lambda_i (1 - \Lambda_i)$.

Since $\Lambda_i \in (0, 1)$ we can conclude that

$$\text{if } x_{il} \in [0, 1] \text{ then } \left[\frac{\partial H}{\partial x_{\bullet l}} \right]_{ll} < 0.$$

So, if the explanatory variable decreases (new explanatory variable still satisfies $x_{il} \in [0, 1]$), the corresponding diagonal term of the Hessian matrix increases. Since the asymptotic variance-covariance matrix of the MLE is the inverse of Hessian matrix, the standard error decreases. So, if the estimation employing sample S_2 yields an insignificant variable (with a range between 0 and 1), then this variable is also insignificant in the estimation employing the original sample S_1 .

Appendix C: Model of admission

Applying to a university is a decision made by a secondary school graduate. The admission of those who applied is entirely the decision of a university. Universities base their decision solely on the results of admission tests. There are *no other criteria* for admission than test scores. Therefore, the model of admission (conditional on application) is a model of admission test scores.

The result of the admission test depends mainly on individual capabilities. Moreover, some additional information may be provided by the applicant's secondary school or by her schoolmates and therefore secondary school characteristics may play a role as well. The performance in the admission test should not depend on regional economic characteristics; other regional characteristics that characterize the environment of an applicant can influence her performance at the admission test.

Finally, university characteristics should be taken into account. First, admission tests are different and also the test score threshold necessary for admission differs across university programs. Second, there are differences in demand for a university. An applicant applying to an oversubscribed program has a lower probability of admission than an otherwise similar applicant applying to a program that is not oversubscribed. Third, we encounter the self-selection problem, i.e. students with different abilities for university study apply to different programs. An applicant applying to a university whose pool of applicants has an overall higher ability faces a lower probability of admission than an otherwise similar applicant applying to a university whose pool of applicants has an overall lower ability. University characteristics employed in the model should control for these three differences among universities.

Econometric model of admission

The unit of observation is an application opposed to the model of applying where the unit is a secondary school graduate. An individual's admission test score S^* is an unobservable (latent) variable. An individual is admitted to a program j if

$$S^* \geq T_j,$$

where T_j is a test score threshold necessary for admission to a particular program j . We model the latent variable S^* in the following way:

$$S^* = \beta_0 + \alpha I + \beta C + \delta d + \varphi R + \varepsilon,$$

where I denotes the set of individual characteristics, vector C includes the set of secondary school (class) characteristics, d stands for the dummy variable indicating the presence of a local university in the applicant's place of residence and vector R contains regional characteristics. An individual is admitted to a program j if

$$S^* - T_j = \beta_0 + \alpha I + \beta C + \delta d + \varphi R - T_j + \varepsilon \geq 0, \quad (C1)$$

where the outcome (admitted or not) is given as

$$Admitted = I[S^* - T_j \geq 0].$$

If we assume logistic distribution for the disturbance then the individual's probability of being admitted to program j (conditional on application) is given as

$$P(admitted = 1 | I, C, d, R, F_j) = \Lambda(\beta_0 + \alpha I + \beta C + \delta d + \varphi R + \tau F_j), \quad (C2)$$

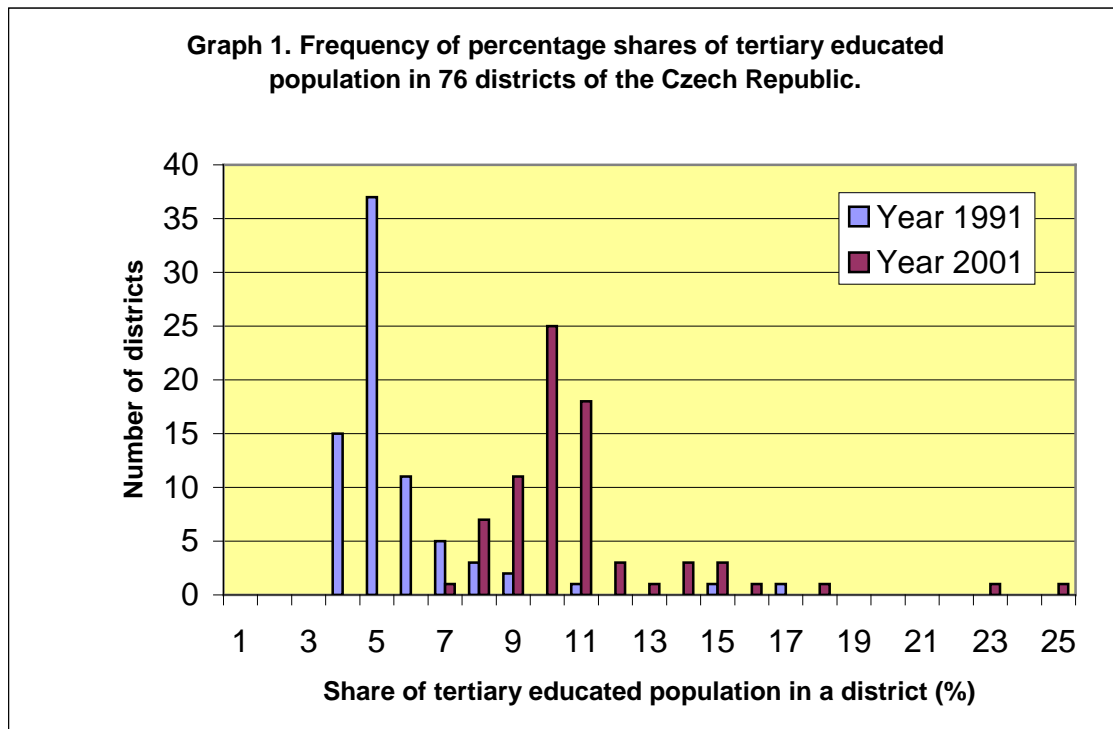
where F_j is a vector of university characteristics that serves as a proxy for the admission test score threshold T_j .

What are appropriate university characteristics F_j that are linked to the threshold T_j ? Dummy variables for university (university program, program specialization) are used to control for differences in admission tests across universities. Furthermore, to control for the average quality of applicants at program j and excess demand for program j we employ a composite score rank of a marginal applicant who is admitted. More precisely, we sort applicants into program j by their composite score. Using the

information on the probability of being admitted we find the lowest composite score that ensures an individual of being admitted if admission test scores duplicate the order of the applicants' composite score achievements.⁴²

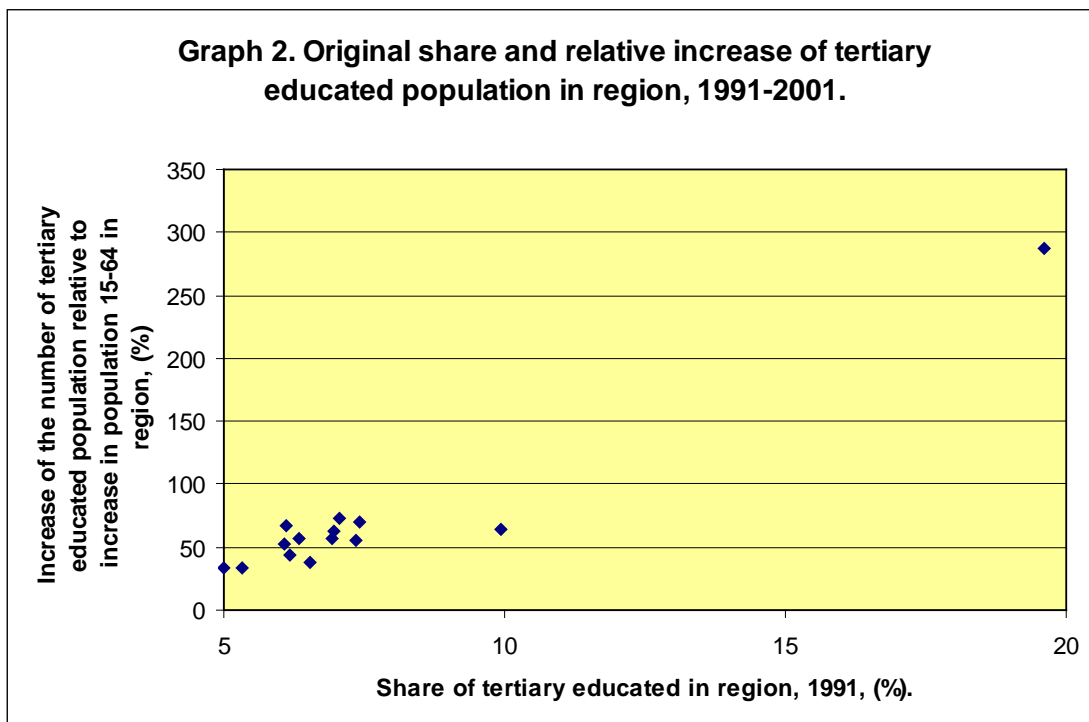
⁴² The use of composite score rank makes sense if we employ dummy variables for university programs. For faculties we already control for the average quality of applicants and excess demand by employing university dummies.

Appendix C: Graphs and Tables



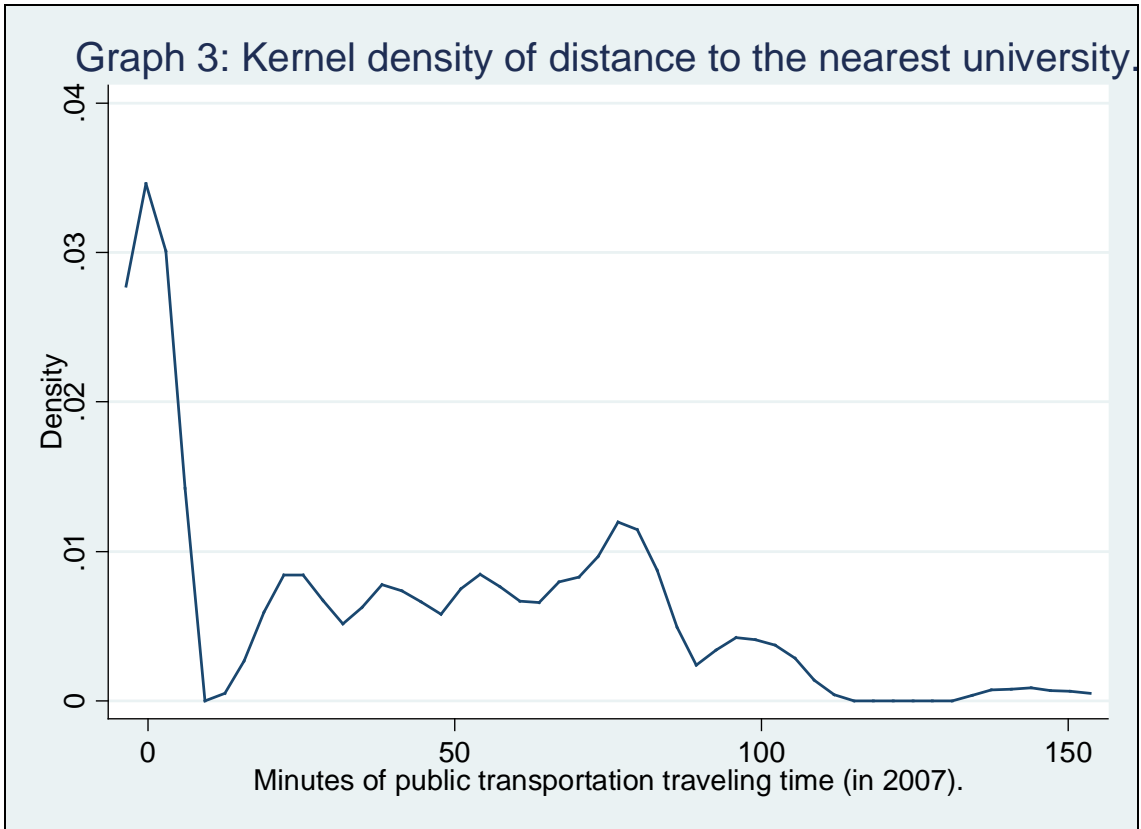
Source: Census 1991, 2001.

Note: Districts correspond to NUTS4-level.



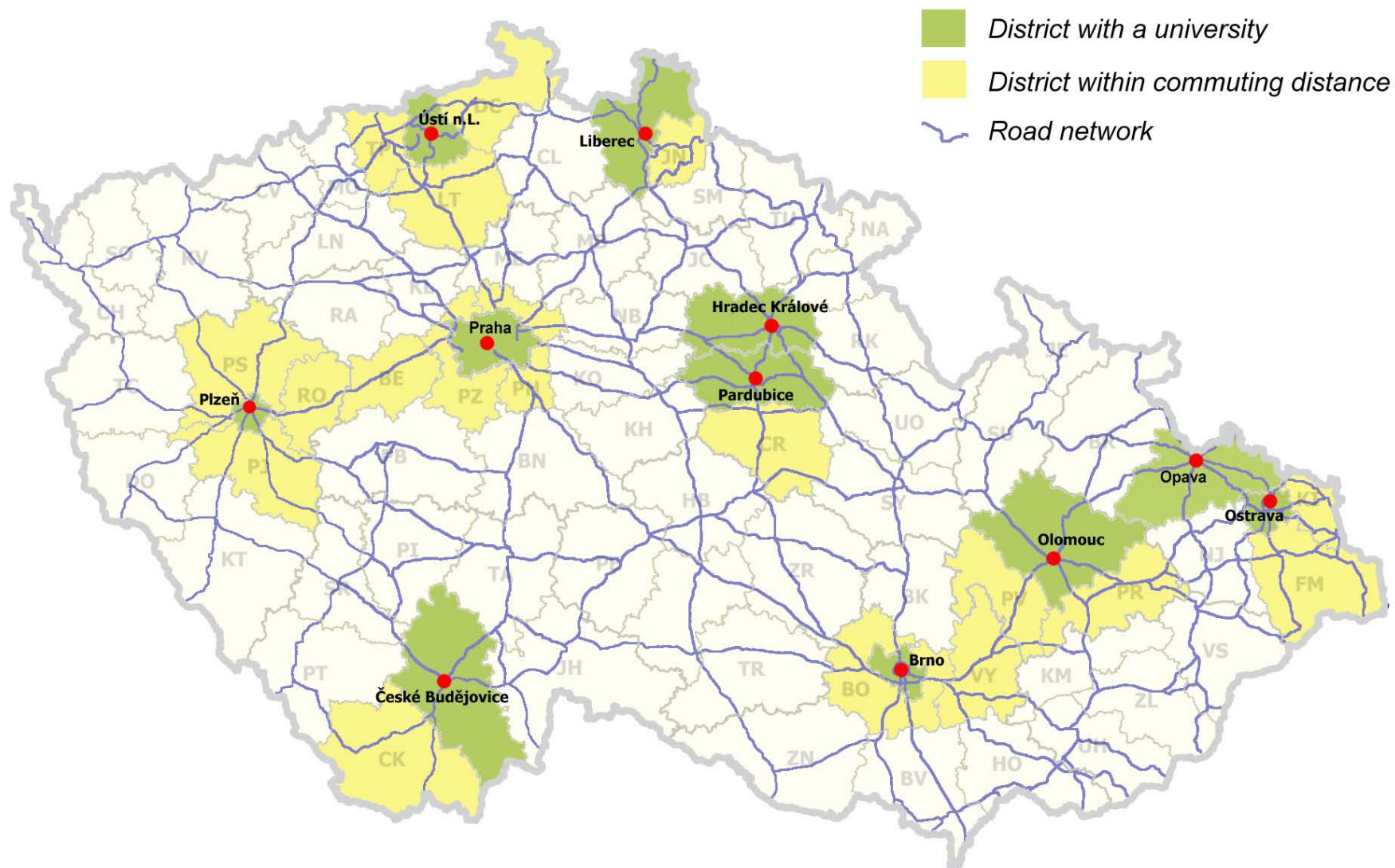
Source: Own calculations based on Census 1991 and 2001.

Note: Region corresponds to NUTS3-level.



Source: Own calculations based on Maturant 1998 and Uchazec 1998.

Graph 4: Districts with universities and districts within a commuting distance to the nearest university, Czech Republic, 1998.



Source: Own calculations.

Table 1. Coefficient of variation of tertiary-educated population share in region.

	Bulgaria	Czech R.	Hungary	Ukraine	Sweden	Greece
w/ capital city	0.35	0.33	0.34	0.3	0.15	0.25
w/out capital city	0.19	0.13	0.16	0.19	0.13	0.1

Source: Jurajda and Terrell (2008), SSO, GSO.

Notes: Numbers for Bulgaria, Czech Republic, Ukraine, and Hungary come from 2001. Numbers for Sweden and Greece come from 2005. Regions are defined by NUTS3 classification.

Table 2. Descriptive characteristics of gymnasium (G) and specialized secondary school (S) graduates and applicants.

	G		S	
Residence type:	local university	no local university	local university	no local university
Secondary school graduates	n=10208	n=8617	n=18894	n=18393
<i>share of applicants</i>	0.92	0.90	0.53	0.47
Individual characteristics				
<i>share of women</i>	0.58	0.61	0.60	0.63
<i>computer at home</i>	0.60	0.47	0.46	0.37
<i>born before 1980</i>	0.52	0.49	0.59	0.58
<i>composite score</i>	64.7	63.9	49.4	48.8
Shares of parental highest education:				
<i>basic&vocational</i>	0.09	0.16	0.25	0.36
<i>secondary</i>	0.34	0.43	0.49	0.48
<i>tertiary</i>	0.57	0.40	0.26	0.16
Class characteristics				
<i>class size (number of students)</i>	28.38	28.16	26.08	26.81
Regional characteristics				
<i>unemployment rate (%)</i>	6.58	7.87	6.88	8.01
<i>GDP growth (1997=100)</i>	99.28	97.22	98.97	97.27
<hr/>				
Applicants	n=9425	n=7784	n=9892	n=8599
<i>share of those admitted</i>	0.66	0.64	0.41	0.36
Individual characteristics				
<i>share of women</i>	0.57	0.60	0.51	0.54
<i>computer at home</i>	0.61	0.48	0.54	0.46
<i>born before 1980</i>	0.51	0.48	0.56	0.54
<i>composite score</i>	65.33	64.67	52.60	52.56
Shares of parental highest education:				
<i>basic&vocational</i>	0.08	0.15	0.19	0.26
<i>secondary</i>	0.33	0.43	0.49	0.52
<i>tertiary</i>	0.58	0.42	0.32	0.22
Class characteristics				
<i>class size (number of students)</i>	28.48	28.25	26.50	27.37
<i>private school</i>	0.10	0.02	0.21	0.18
Regional characteristics				
<i>unemployment rate (%)</i>	6.54	7.87	6.84	7.99
<i>GDP growth (1997=100)</i>	99.31	97.20	98.98	97.13

Source: Own calculations based on Maturant 1998, Uchazec 1998 and data provided by CSO.

Notes: G denotes gymnasium graduates and S specialized secondary schools graduates.

Composite score is the average test score computed from four tests taken from Czech and one foreign language, mathematics and study aptitude.

Table 3. Share of applicants from gymnasiums (G) and specialized secondary schools (S) applying to one program specialization only.

	G		S	
	Local university	Non-local university	Local university	Non-local university
Applying to:				
<i>local university only</i>	0.21		0.52	
<i>non-local university/ies only</i>	0.28	0.20	0.63	0.53
<i>both</i>	0.13		0.30	
<i>total</i>	0.19	0.20	0.48	0.53

Source: Own calculations based on Maturant 1998, Uchazec 1998 and data provided by CSO.

Table 4a. Difference between the average share of applicants without local university and share of applicant without local university in selected programs by type of secondary school.

G	University program	Number of universities	districts	Prob. of admission	All applicants	Applicants to non-local universities only
					Difference I	Difference II
	Technical chemistry	5	2	0.51	1142	727
	Architecture	4	3	0.18	419	201
	Veterinary medicine	2	1	0.25	421	243
	Medicine	11	7	0.29	1925	930
	Philosophy, Theology	11	6	0.27	1084	410
	Social sciences	6	5	0.13	951	395
	Education	10	5	0.12	685	431
	Theory and history of art	5	4	0.12	398	160
S						
S	University program	Number of universities	districts	Prob. of admission	All applicants	Applicants to non-local universities only
					Difference I	Difference II
	Engineering	6	6	0.59	2215	1042
	Construction	5	4	0.51	1230	687
	Transportation	4	2	0.31	555	244
	Agriculture	8	3	0.3	2822	1621
	Education	10	5	0.12	702	435

Source: Own calculations based on Maturant 1998 and Uchazec 1998.

Notes: The share of applicants from gymnasiums (specialized secondary schools) without local university is 0.45 (0.46).

The share of applicants applying to non-local universities only is 0.86 for applicants from gymnasiums and 0.83 for applicants from specialized secondary schools.

G denotes gymnasiums, S specialized secondary schools.

Table 4b. Cognitive and non-cognitive skills of applicants from gymnasiums (G) and specialized secondary schools (S) in selected university programs.

G		Cognitive skills - composite score			Non-cognitive skills		Difference I (taken from Table 4a)	
		Local university		Non-local university	Local university	Non-local university		
Residence type:	Applying to:	Local university only	Non-local university only	Both				
University program								
Technical chemistry		66.66	68.49	66.55	66.29	0.012	0.019	0.08
Architecture		68.44	69.20	69.07	66.45	0.014	0.021	-0.07
Veterinary medicine		63.19	63.57	64.10	64.40	0.009	0.016	0.02
Medicine		66.78	67.11	66.58	67.16	0.012	0.020	-0.03
Philosophy, Theology		67.53	68.69	66.39	65.52	0.018	0.021	-0.13
Social sciences		68.18	68.84	66.41	65.71	0.014	0.017	-0.10
Education		58.88	60.65	62.15	60.07	0.010	0.018	0.11
Theory and history of art		64.87	64.76	64.52	62.90	0.020	0.014	-0.11
S								
Engineering		51.63	50.37	51.28	49.97	0.014	0.022	-0.08
Construction		51.78	49.56	50.14	47.90	0.019	0.023	-0.07
Transportation		52.11	50.78	53.48	52.49	0.018	0.020	-0.11
Agriculture		55.32	49.68	55.19	54.31	0.019	0.021	0.04
Education		45.62	45.29	47.04	47.50	0.016	0.020	0.11

Source: Own calculations based on Maturant 1998 and Uchazec 1998.

Non-cognitive skills are measured using the relative excess demand for a gymnasium in a district.

Table 5. University applicants from gymnasiums (G) and specialized secondary schools (S): descriptive statistics.

Residence type:	G				S			
	Local university			No local university	Local university			No local university
	Applying to: Only local university n=3636	Only non-local university/ies n=1427	Both n=4362	n=7784	Only local university n=5318	Only non-local university/ies n=1966	Both n=2608	n=8599
Composite score	65.11	65.82	65.35	64.67	52.37	50.82	54.41	52.56
Avg number of applications sent	2.82	3.60	4.53	4.09	1.85	2.30	3.44	2.63
Avg prob of being admitted conditional on application	26.99	26.54	22.54	24.66	29.63	29.64	24.25	25.33
Share of admitted to at least one program	0.62	0.65	0.69	0.64	0.38	0.40	0.47	0.36
Avg number of faculties an applicant is admitted to	0.99	1.14	1.24	1.16	0.47	0.52	0.68	0.50

Source: Own calculations based on Maturant 1998 and Uchazec 1998.

Notes: G denotes gymnasium graduates and S specialized secondary schools graduates.

Table 6. Application equation: estimated marginal effects.

(G - gymnasiums, S - specialized secondary schools)

	G	S
Local university dummy		
Living within commuting distance to a university	0.000 (0.03)	0.003 (0.19)
Individual characteristics		
Female	-0.009** (2.18)	-0.101*** (11.67)
Highest level of parental education: secondary	0.022*** (5.70)	0.109*** (14.33)
Highest level of parental education: tertiary	0.046*** (10.02)	0.210*** (21.66)
Computer at home	0.010*** (2.71)	0.096*** (14.17)
Born before 1980	-0.015*** (4.68)	-0.081*** (11.57)
Composite score rank	0.001*** (14.60)	0.006*** (36.16)
Estimated marginal effects for subjects taken at the maturita exam (gymnasiums) and field of secondary school (for specialized secondary schools) are presented in Table 7.		
Class (school) characteristics		
Class size	0.001** (2.37)	0.005*** (4.14)
Class average composite score rank	0.000** (1.96)	0.002*** (5.07)
Private secondary school	-0.007 (0.83)	0.025* (1.65)
Regional (district) characteristics		
District unemployment rate	0.002*** (2.67)	0.007*** (3.51)
Regional GDP growth	-0.001* (1.77)	-0.013*** (6.26)
Share of tertiary-educated population in district	0.314*** (3.37)	1.135*** (4.79)
Relative excess demand for gymnasiums in a district	0.187** (2.14)	0.590** (2.57)
Constant	0.019 (0.23)	0.528** (2.57)
Observations	15809	31637

Absolute value of z statistics in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: The dependent variable is a dummy that equals 1 if the graduate applies to at least one university and 0 otherwise.

Reference individual is male, born in 1980, parents finished basic or vocational education, no computer at home, attends state secondary school and resides outside of commuting distance to a university.

The field reference category is business for specialized secondary school graduates.

Table 7. Application equation: estimated marginal effects for subjects taken for the maturita exam and the subject of secondary school.

	Gymnasiums	Specialized secondary schools	
Subject taken for the maturita exam		Specialization	
Foreign language	-0.012 (0.59)	Agriculture	0.090*** (4.43)
Mathematics	0.053*** (9.74)	Manufacturing	0.120*** (8.38)
Biology	0.021*** (3.75)	Light manufacturing	0.121*** (5.88)
Physics	0.019** (2.39)	Health care	0.062*** (3.10)
Chemistry	0.052*** (6.89)	Social sciences	0.263*** (9.70)
History	0.030*** (6.48)	Art	0.230*** (5.71)
Geography	-0.003 (0.74)		
Social sciences	0.004 (0.98)		

Absolute value of z statistics in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8. Admission equation for gymnasiums: estimated marginal effects.

	1	2	3	4	5	6	7
Local university dummy							
Living within a commuting distance to a university	-0.004 (0.42)	-0.018** (1.98)	-0.015 (1.57)	-0.057*** (3.32)	-0.028 (1.60)	-0.006 (0.32)	-0.050*** (2.82)
Interaction with probability of admission	-	-	-	0.141*** (3.68)	0.034 (0.85)	-0.013 (0.33)	0.051 (1.32)
Presence of preferred program at local university	0.067*** (7.47)	0.077*** (8.40)	0.061*** (6.57)	0.111*** (6.20)	0.105*** (5.83)	0.094*** (5.16)	0.099*** (5.43)
Interaction with probability of admission	-	-	-	-0.108*** (2.58)	-0.137*** (3.24)	-0.091*** (2.16)	-0.098** (2.35)
Individual characteristics							
Female	-0.031*** (4.49)	-0.028*** (3.83)	-0.018** (2.44)	-0.043*** (6.23)	-0.029*** (4.06)	-0.014* (1.93)	-0.016** (2.14)
Highest level of parental education: secondary	0.017 (1.48)	0.026** (2.18)	0.026** (2.22)	0.015 (1.31)	0.024** (2.00)	0.030*** (2.58)	0.029** (2.46)
Highest level of parental education: tertiary	0.080*** (7.07)	0.088*** (7.49)	0.092*** (7.69)	0.076*** (6.74)	0.089*** (7.50)	0.092*** (7.84)	0.089*** (7.56)
Computer at home	0.007 (1.10)	0.006 (0.81)	0.003 (0.43)	0.012* (1.80)	0.010 (1.39)	0.001 (0.19)	0.004 (0.50)
Born before 1980	0.007 (1.06)	0.008 (1.21)	0.008 (1.25)	0.005 (0.79)	0.007 (0.99)	0.007 (0.99)	0.006 (0.87)
Composite score rank	0.008*** (30.33)	0.009*** (30.98)	0.009*** (30.99)	0.008*** (30.18)	0.010*** (31.52)	0.009*** (30.82)	0.009*** (31.24)
Subjects at maturita (dummies)	-	-	-	-	-	Included	Included
Class (school) characteristics							
Class size	0.003*** (2.64)	0.002** (2.21)	0.002** (2.00)	0.002** (2.42)	0.002 (1.49)	0.001 (1.44)	0.000 (0.09)
Class average composite score rank	0.004*** (9.74)	0.004*** (9.87)	0.005*** (10.64)	0.004*** (9.93)	0.005*** (10.69)	0.004*** (10.24)	0.003*** (5.37)
Private secondary school	-0.053*** (3.19)	-0.047*** (2.67)	-0.049*** (2.76)	-0.055*** (3.29)	-0.054*** (3.07)	-0.069*** (3.94)	-0.044** (2.41)
% of admitted (gymnasium)	-	-	-	-	-	-	0.294*** (9.00)
Regional (district) characteristics							
Share of tertiary educated population in district	-0.599*** (5.25)	-0.372*** (2.98)	-0.299** (2.37)	-0.687*** (6.04)	-0.188 (1.61)	-0.282** (2.40)	-0.161 (1.36)
Relative excess demand for gym.	-0.011 (0.08)	0.028 (0.18)	0.028 (0.18)	0.050 (0.34)	0.068 (0.44)	-0.043 (0.28)	-0.322** (2.05)
University characteristics							
Program specialization dummies	Included	-	-	Included	Included	Included	Included
University dummies	-	Included	-	-	-	-	-
University program dummies	-	-	Included	-	-	-	-
Marginal rank of admittance to program	-	-	-	-	-0.015*** (30.07)	-0.015*** (27.62)	-0.015*** (30.28)
Constant	-1.556*** (33.16)	-1.277*** (28.35)	-1.323*** (29.23)	-1.290*** (31.01)	-0.143** (2.50)	-0.213*** (2.83)	-0.213*** (3.08)
Observations	43073	42897	42771	42833	42833	43073	43073

Absolute value of z statistics in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Reference individual is male, born in 1980, parents finished basic or vocational education, no computer at home, attends state secondary school and resides outside of commuting distance to a university.

Pseudo R2 of reported specifications between 0.21-0.27. Misspecification problems detected for some specifications.

Notes: Included – indicates group of dummies used in estimations but not reported here because of number of dummies.

Table 9. Admission equation for specialized secondary schools: estimated marginal effects.

	1	2	3	4	5	6	7
Local university dummy							
Living within a commuting distance to a university	0.009 (1.15)	-0.003 (0.43)	-0.007 (0.89)	0.016 (1.31)	0.014 (1.09)	0.015 (1.22)	-0.004 (0.33)
Interaction with probability of admission	-	-	-	-0.017 (0.59)	-0.051* (1.71)	-0.057* (1.90)	-0.062** (2.05)
Presence of preferred program at local university	0.041*** (4.91)	0.048*** (5.97)	0.049*** (5.86)	0.039*** (2.97)	0.040*** (3.15)	0.045*** (3.53)	0.062*** (4.87)
Interaction with probability of admission	-	-	-	0.007 (0.25)	0.016 (0.54)	0.008 (0.26)	-0.045 (1.48)
Individual characteristics							
Female	-0.040*** (6.24)	-0.029*** (4.44)	-0.019*** (2.84)	-0.046*** (7.14)	-0.025*** (3.86)	-0.024*** (3.58)	-0.017*** (2.62)
Highest level of parental education: secondary	0.017** (2.23)	0.016** (2.16)	0.017** (2.31)	0.016** (2.14)	0.019** (2.43)	0.019** (2.45)	0.018** (2.41)
Highest level of parental education: tertiary	0.045*** (5.48)	0.047*** (5.74)	0.048*** (5.70)	0.044*** (5.31)	0.049*** (5.80)	0.048*** (5.70)	0.046*** (5.61)
Computer at home	0.001 (0.18)	0.006 (0.97)	0.004 (0.72)	0.003 (0.54)	0.004 (0.68)	0.004 (0.69)	0.005 (0.83)
Born before 1980	0.003 (0.49)	-0.003 (0.52)	-0.001 (0.26)	0.001 (0.15)	0.001 (0.20)	0.002 (0.30)	-0.003 (0.48)
Composite score rank	0.003*** (20.78)	0.004*** (22.25)	0.004*** (22.51)	0.003*** (20.28)	0.004*** (22.46)	0.004*** (22.89)	0.004*** (23.59)
Field of secondary school (dummies)	-	-	-	-	-	Included	Included
Class (school) characteristics							
Class size	-0.001 (1.17)	-0.001 (1.27)	-0.000 (0.54)	-0.001 (1.30)	-0.001 (1.26)	-0.001 (0.74)	-0.001 (1.21)
Class average composite score rank	-0.000 (0.13)	0.001*** (3.76)	0.001*** (4.10)	0.000 (0.06)	0.001*** (3.22)	0.001*** (4.54)	-0.000 (1.35)
Private secondary school	-0.101*** (11.38)	-0.083*** (9.53)	-0.080*** (8.76)	-0.104*** (11.70)	-0.085*** (9.45)	-0.074*** (7.24)	-0.046*** (4.55)
% of admitted (gymnasium)	-	-	-	-	-	-	0.416*** (16.43)
Regional (district) characteristics							
Share of tertiary educated population in district	-0.536*** (5.53)	-0.165 (1.63)	-0.165 (1.56)	-0.579*** (5.97)	-0.308*** (3.18)	-0.332*** (3.46)	-0.274*** (2.80)
Relative excess demand for gymnasiums in a district	-0.135 (0.94)	-0.149 (1.05)	-0.205 (1.38)	-0.127 (0.88)	-0.115 (0.78)	-0.091 (0.62)	-0.204 (1.43)
University characteristics							
Program specialization dummies	Included	-	-	Included	Included	Included	Included
University dummies	-	Included	-	-	-	-	-
University program dummies	-	-	Included	-	-	-	-
Marginal rank of admittance to program	-	-	-	-	-0.009*** (28.26)	-0.008*** (27.40)	-0.008*** (26.52)
Constant	-0.671*** (11.69)	-0.501*** (19.11)	-0.550*** (19.93)	-0.250*** (9.06)	0.275*** (8.55)	-0.040 (0.64)	-0.051 (0.83)
Observations	31060	31048	30259	30850	30850	31060	31060

Absolute value of z statistics in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Reference individual is male, born in 1980, parents finished basic or vocational education, no computer at home, attends state secondary school and resides outside of commuting distance to a university.

Pseudo R2 of reported specifications between 0.25-0.30. Misspecification problems detected for some specifications.

Notes: Included – indicates group of dummies used in estimations but not reported here because of number of dummies.

Table 10. Admission equation - selected marginal effects.
(G - gymnasiums, S - specialized secondary schools)

	G	S
Local university dummy		
Presence of study field at local university	0.099*** (5.43)	0.062*** (4.87)
Interaction with probability of admission field	-0.098** (2.35)	-0.045 (1.48)
Individual characteristics		
Female	-0.016** (2.14)	-0.017*** (2.62)
Highest level of parental education: secondary	0.029** (2.46)	0.018** (2.41)
Highest level of parental education: tertiary	0.089*** (7.56)	0.046*** (5.61)
Computer at home	0.004 (0.50)	0.005 (0.83)
Born before 1980	0.006 (0.87)	-0.003 (0.48)
Composite score rank	0.009*** (31.24)	0.004*** (23.59)
Class (school) characteristics		
Private secondary school	-0.044** (2.41)	-0.046*** (4.55)
Regional (district) characteristics		
Share of tertiary educated population in district	-0.161 (1.36)	-0.274*** (2.80)
Relative excess demand for gym.	-0.322** (2.05)	-0.204 (1.43)
Faculty characteristics		
Marginal rank of admittance to program	-0.015*** (30.28)	-0.008*** (26.52)
Constant	-0.213*** (3.08)	-0.051 (0.83)
Observations	43073	31060

Absolute value of z statistics in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%

Reference individual is male, born in 1980, parents finished basic or vocational education, no computer at home, attends state secondary school, resides outside of commuting distance to a university.