Admission to Selective Schools, Alphabetically

Štěpán Jurajda and Daniel Münich
CERGE-EI

December 2, 2005

Abstract

One's position in an alphabetically sorted list may be important in determining access to rationed goods or oversubscribed public services. Motivated by anecdotal evidence, we investigate the importance of the position in the alphabet of the last name initial of Czech students for their admission chances into oversubscribed schools. Empirical evidence based on the population of students applying to universities in 1999 suggests that, among marginal applicants, moving from the top to the bottom of the alphabet decreases admission chances by over 2 percent. The implication of such admission procedures for student ability sorting across differently oversubscribed schools is then confirmed by evidence based on a national survey of secondary students' test scores.

Acknowledgements Both authors are Research Affiliates at CEPR, London; Jurajda is also Research Fellow at IZA, Bonn. The help of Petr Matějů and Jindřich Krejčí from the Sociological Institute of the Czech Academy of Sciences and Vladimír Burda (formerly of the Czech National Institute of Technical and Vocational Education) in acquiring and processing the Sonda Maturant data is gratefully acknowledged, as are comments from Peter Katuscak and Peter Orazem. This research has been supported by the Grant Agency of the Czech Republic (grant No.403/03/340) and a grant from the CERGE-EI Foundation under a program of the Global Development Network. All opinions expressed are those of the author(s) and have not been endorsed by CERGE-EI or the GDN.
Address CERGE-EI, POB 882, Politických vězňů 7, Prague, Czech Republic. E-mail: `stepan.jurajda@cerge-ei.cz`, `daniel.munich@cerge-ei.cz` CERGE-EI is a joint workplace of the Center for Economic Research and Graduate Education, Charles University, and the Economics Institute of the Academy of Sciences of the Czech Republic.

1 Introduction

Sorting based on 'alphabetical order' is a fact of everyday life. Team members are listed in this order, including co-authors of scientific papers; students may be seated in classroom according to their last name initials' position in the alphabet; competing firms are displayed alphabetically in phone and other directories.

Could this systematic and omnipresent sorting provide an advantage to those positioned high in the alphabet? This question is often the object of popular discussions.[1] Customers may choose their service provider from the top of an alphabetically sorted directory; students seated in front rows of classrooms may achiever higher learning outcomes; employers using the apparently non-discriminatory alphabetical order may be more attentive to job applicants who are interviewed first; there may be citation bias against authors whose last names begin with letters that occur late in the alphabet (McCarl, 1993); etc. Yet, so far there is little evidence on the issue, thanks in large part to lack of data with individual initials.

The question of non-discriminatory sorting is particularly important when allocating a prize or distributing a rationed good or oversubscribed public service, even when the allocation mechanism is based on applicants' characteristics. For example, van Ours and Ginsburgh (2003) show that the (randomly assigned) order in which musicians play in a competition affects their success, both in the competition and in their whole career. In this paper, we study an allocation mechanism that affects entire population cohorts: We ask whether students with last names sorted high in the alphabet enjoy higher chances of being admitted to oversubscribed (selective) schools.

Why would one expect such effect to take place? Alphabetical sorting can be applied in school admission procedures when lists with multiple student characteristics (including test scores) are

[1] For example, The Economist (2001) suggests such effect may be present in politics by pointing out the high fraction of U.S. presidents and U.K. prime ministers with last names sorted high in the alphabet.

prepared for admission committees or when students are called to oral exams in alphabetical order. When applications are evaluated based on multiple criteria in absence of a clear summarizing measure, marginal cases at the top of the list may obtain a more favorable treatment compared to marginal applicants toward the bottom of the list where constraints on total number of possible admissions become more binding. Similarly, it is plausible that examiners are more attentive (approving) and applicants more rested during oral exams scheduled in the morning of an exam day.

In this paper, we empirically test several implications of alphabet-based admission procedures. We focus on the case of the Czech Republic, which features a highly selective admission process at both secondary and tertiary schooling level, and thus provides a good example of the many European selective education systems.[2] At 12%, the country has one of the lowest tertiary attainment rates in the OECD (OECD, 2004) and students entering the university system typically come from selective general (academic) secondary programs serving less than 15% of each cohort of secondary-school students.[3] Furthermore, anecdotal evidence suggests that alphabetical sorting is being used in various ways in school admission procedures, both at the secondary and tertiary level. In addition to the general mechanisms described above, Czech universities also occasionally use the alphabetical order to break ties when several students achieve similar admission-test results.

Our empirical analysis is based on the experience of the whole population of secondary-school graduates in 1999. We start by studying the success of their applications to Czech universities and

[2] Admission standards for tertiary education are applied in Belgium, Denmark, Germany, and the Netherlands. Stricter admission standards are used in the UK, Sweden and in some of the French universities. See Jacobs and van der Ploeg (2005).

[3] In 1999—the year our data come from—55% of all applicants to Czech universities were not able to enroll in any program (as 72% of all individual-university-specific applications were turned down). In contrast, almost 70% of the applicants who graduated in the same year from the selective academic secondary programs did manage to enter a university.

find a significant effect of one's last-name-initial position in the alphabet on admission chances. The presence of alphabet-affected admission practices implies that among students admitted to selective schools, those with last names in the bottom part of the alphabet have on average higher ability. We test this implication using a national study-achievement test administered to our student population graduating from secondary schools in 1999 and find evidence fully consistent with the alphabet-based sorting hypothesis. Throughout our analysis we also test for the importance of the alphabetical position of the first-name initial, thus providing a natural check on our main results. It is reassuring that we do not find the first-name-initial position in the alphabet to play any important role.

The paper is organized as follows: We describe the Czech education system and our student data in Section 2. Here, we also outline our testing strategy. Sections 3 and 4 present the college-admission and test-score analysis, respectively. Finally, we provide some tantalizing wage analysis based on a 1996 household survey in Section 5. The last section summarizes our findings.

2 The Czech Education System and our Data

Although the structure of the Czech educational system parallels those of other European countries, it differs dramatically in the relative magnitude of education provision across specific degrees and school types: While the secondary school completion rate is very high, only a small proportion of the Czech population has traditionally completed university. After the collapse of communism in 1989, total enrollment in Czech public colleges doubled,[4] but college-program completion rates decreased and, given the large size of cohorts graduating from secondary schools during the 1990s, the tertiary attainment rate of the Czech population aged 25-34 remains starkly low at about 12% as of 2002 (OECD, 2004).

[4] All of the universities in our data are public and tuition-free. Enrollment in private colleges emerged only after 1999. Even today, private tuition-based tertiary education remains miniscule in the Czech Republic.

The low tertiary attainment rate is not surprising given that a major group of secondary-level students attends apprenticeship programs, which offer only dismal prospects of continuing on to higher education degrees. Most of the apprenticeship programs do not lead to a school-leaving comprehensive examination ('Maturita' in Czech), which is a pre-requisite for tertiary education. These exams, administered at the end of most four-year secondary programs are prepared by each school individually based on national guidelines; they approximately correspond to the U.K. General Certificate of Secondary Education (GCSE) or the German 'Abitur' exam.[5]

Our study is based on administrative data capturing the experience of the cohort of students graduating from all types of secondary programs with the school-leaving exam in 1999. For each student in this cohort, we observe the school-leaving-exam test scores as well as all applications to universities together with the admission decision. Below, we describe the data in more detail and use them to offer several stylized facts about the Czech education system.

2.1 Student Test Scores in Secondary Schools

In 1999 the first (and so-far the last) nation-wide study achievement test—a national 'Maturita' exam—was administered at all programs with the school-leaving exam. The testing, conducted independently of the traditional school-specific 'Maturita' exams, thus targeted approximately 60 percent of the entire age cohort of twelve-graders, i.e. over 100 thousand students in over 1,642 schools. Exams were held simultaneously and the results were processed centrally.[6]

The tested students come from three types of Czech secondary 4-year programs: apprenticeship,

[5] In terms of the OECD classification of education levels, the apprenticeship programs without a 'Maturita' exam correspond to the ISCED 2 level (and a small group of workers with ISCED 3C). These programs serve about 40% of the cohort. Secondary-school education with 'Maturita' then correspond to ISCED 3A. All students taking the 'Maturita' exam have completed at least 12 years of education.

[6] While pilot testing of standardized 'Maturita' exams started in 1997, the 1999 program, called 'Sonda Maturant 99,' was the first to cover the whole population of students taking the 'Maturita' exam and was the largest school testing program administered in the Czech Republic to date.

specialized and general academic. Apprenticeship programs typically focus on craft skills, while examples of specialized secondary programs include construction or nursing schools. Finally, the general academic programs are typically strong in both humanities and mathematics skills and give highest value added in terms of study achievements.[7]

Our test data provide standardized test scores (on a 0 to 100 scale) corresponding to students' mathematics skills and to their command of the native Czech as well as one foreign language. Because different students choose different foreign languages, we focus on the mathematics and Czech results. Besides test scores, the data include students' gender, school type and district identifier. A unique feature of these data is that they contain the first and last name initials of tested students.[8] Out of the total of 105,979 tested students, we observe name initials for over 97 thousand students and among these, 91,599 have valid test scores from mathematics and Czech language tests available.[9]

We checked whether the last-name-initial distribution in the student data is similar to that based on the population register. The correlation across the two data sources in each letter's share is high (0.95). Figure 1 presents the distribution of last name initials in our data.[10]

[7] See Filer and Münich (2000) or Matějů and Straková (2005) for a detailed description of the Czech education system and Münich (2004) for evidence on the value added of each type of secondary schools.

[8] Full student names were recorded as part of the student-specific information supplied by the schools. We were able to obtain only the initials.

[9] There are 16 thousand students who selected a different, reduced form of the mathematics test. We prefer to keep these reduced tests and control for the different mean of the test by including a dummy variable indicating the choice of the reduced test. In an alternative approach, we drop the reduced-test scores. The results we obtained were both qualitatively and quantitatively highly similar to those we present below.

[10] In the Czech Republic, there are no types of last names related to a history of family wealth (such as "van" or "von" names; see, e.g., Moldanová, 2004). The country is also highly ethnically homogenous, with only one sizeable minority, the Roma. There are no nationally representative studies identifying the minority members, let alone their names. We were able to compare the national last-name-initial statistics with those based on a 2002 data set of older Roma respondents collected by a Czech NGO, 'People in Need'. The Roma last-name distribution appears highly similar to the national one with a letter-share correlation of 0.76. In any case, there are very few Roma in selective

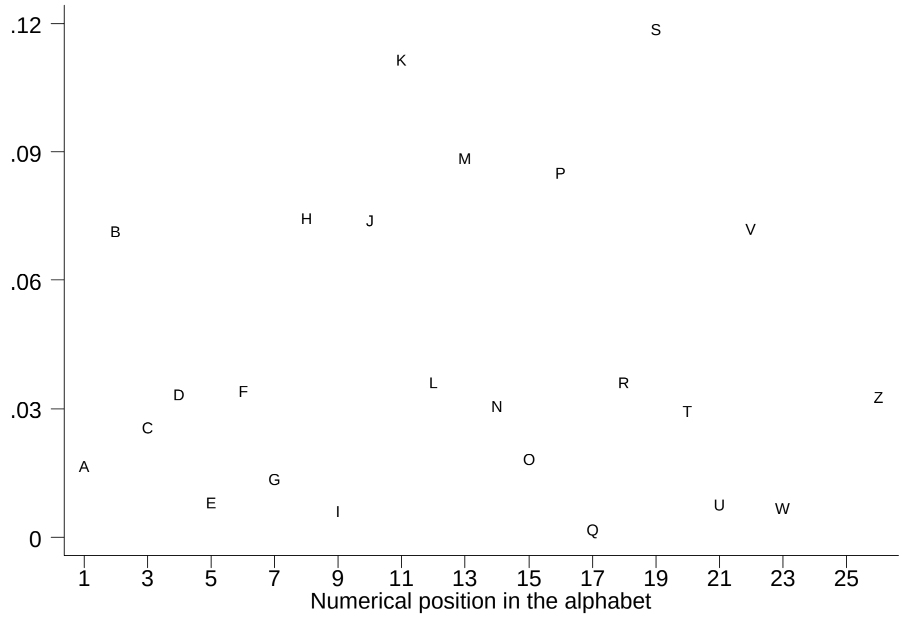

Figure 1: Distribution of Last Name Initials

Next, Table 1 provides a summary of the test score data by school type and supports the typical ordering of study achievement with academic programs at the top and apprenticeship programs at the bottom.[11] Students graduating from academic programs also have the highest chance of being admitted to highly selective universities. Further, Filer et al. (1999) show that in 1997 wages of workers with specialized degrees were about 20% higher than wages of otherwise comparable workers with apprenticeship degrees. There was a similar wage premium for those workers with academic secondary degrees who did not achieve university degrees.[12]

Czech schools as they are very likely to end up with only the lowest level of compulsory education and, given Czech language inefficiency, are often redirected to schools for mentally handicapped children (Šimíková et al., 2004).

[11] The number of students reported in the table (N) reflects all students in the data, irrespective of whether they have a valid mathematics or Czech test score.

[12] The Czech Republic also features one of the highest college/high-school wage gaps in the EU (Jurajda, 2005).

Table 1: Mean Test Scores and Excess Demand by School Type

School Type	Academic	Specialized	Apprenticeship
Mathematics test score	46.3	26.6	22.7
	(16.2)	(15.0)	(10.5)
Czech language test score	74.0	58.8	51.9
	(11.8)	(12.3)	(11.4)
Share of female students	0.58	0.59	0.43
N	19,448	50,922	26,699
District Excess Demand*	0.31	0.24	0.15
	(0.15)	(0.19)	(0.24)

Note: Standard deviations in parentheses.

*The ratio of the number of rejected applications over the number of admitted ones.

Not surprisingly, general academic and specialized high schools are in high excess demand.[13] The last row of Table 1 shows the average of an excess demand measure available for each school type in each of Czech Republic's 76 districts (NUTS-4 territorial units).[14] Our measure of excess demand is defined as the ratio of the number of applicants rejected over the number of students

[13] The total size of the academic programs has not increased sufficiently during the first post-communist decade to meet the excess demand. The Czech Ministry of Education kept strict and rather stable limits on the maximum number of students admitted to public academic programs that it would finance. A partial adjustment came from the establishment of private secondary schools in the early 1990s (Filer and Münich, 2000). It is also important to note that there are administrative limits on the number of applications submitted by each student. These limits lower the observed excess demand for the most sought-after programs as students judge the low chance of admission to a general academic program against the higher admission probability of an alternative application submitted to a less selective school.

[14] We pool the districts falling within the capital city of Prague into one district because of the high mobility of students across districts within the city of Prague. The average district population size (excluding Prague) is approximately 100 thousand.

admitted to schools of a given type in a given district in 1998.[15] We note that on top of the school-type related differences in over-subscription, there is also substantial school-district-specific variation in excess demand for secondary schooling due in part to the shrinking of the youth Czech population, which occurs at different rates across districts (Münich, 2004). The student admission process is governed independently by individual schools, which base their admission decision on their entrance-exam results combined with other student-background information including grades from elementary education.

2.2 College Admissions

Secondary-school graduates can submit an unlimited number of university applications (as long as they manage to be physically present for the admission exams at all schools they apply to). An application process typically consists of a written exam and in a subset of faculties includes also oral exams. While some faculties have a common admission procedure, many departments organize their own admission tests. Applicants are then informed of their admission status and choose schools they want to enroll in.

We have merged the secondary-school-graduates data described above with the administrative register of individual applications to Czech universities in 1999. The college-application data report the success or failure of each individual application (whether a given student was admitted to a particular school as well as eventual enrollment), but falls short of providing admission-test scores and does not give name initials. We therefore focus our analysis on college applicants who have graduated from secondary programs in 1999 (for whom we have available name initials as well as 'Maturita' scores) and omit those who did so earlier. Applications by "fresh" secondary-school graduates constitute 55% of all applications and 61% of university admissions in 1999.

[15] Thus, this measure of excess demand does not exactly correspond to the year in which our cohort of 1999 graduates was admitted. This school information was collected in 1998 as part of the 'Sonda Maturant' testing program.

The merged data provide information on a total of 116,479 applications submitted by 41,486 1999-secondary-school graduates to 116 distinct faculties of Czech public universities.[16] In total, 29% of these applications were admitted and as a result 49% of our applicants eventually did enroll in a university program. Looking across the 116 faculties our data distinguish, the fraction of applications admitted varies widely around the median of 0.29, but is fairly low even at the 90th percentile of average faculty-level admission probability, which equals 0.60. Hence, almost all universities are highly selective.

2.3 The Use of Alphabet and Testing Strategy

There is anecdotal evidence suggesting that admission procedures to Czech secondary schools and universities often use the 'alphabetical order' to sort applications before final evaluation by an admission committee or to schedule applicants' oral exams. Alphabetical sorting of applications characterized by multiple evaluation criteria may lead to an effect of the sorting on admission outcome if constraints on total number of possible admissions become more binding towards the bottom of a list. Similarly, it is plausible that examiners are more attentive during oral exams scheduled in the morning of an exam day and that the applicants are tired during the afternoon exam sessions.[17] In some cases Czech universities openly use the alphabetical order to break ties among applicants with identical admission test scores.[18]

[16] This means that about 57% of 'Maturita'-exam takers in 1999 did not apply for a university. Some of these secondary-school graduates can continue their education in the so-called higher vocational schools, which are outside of the university system.

[17] See van Ours and Ginsburgh (2003) and the references therein for evidence suggesting that order of performance matters in musical competitions.

[18] Such practice was recently featured in the official description of admission procedures of a department in a prestigious local university. We quote from the official specification of the admission procedure at one department of the Philosophical Faculty of the Charles University: "After sorting applicants based on test score, the first 30 will be admitted (should more applicants reach the same test score level, the list will be sorted alphabetically based on last name initial)." The announcement, in Czech, was originally posted at `http://prijimacky.ff.cuni.cz` and is now

Assuming that ability and last-name initials are independent and that students do not adjust their application strategy based on their position in the alphabet, such an alphabet-based admission process would lead to a negative correlation between being admitted to selective (higher) schools and one's position in the alphabet, conditional on applying. Furthermore, such admission processes would also lead to a positive correlation between ability and one's numerical position in the alphabet among students already admitted to highly selective schools as well as among students admitted to easily accessible schools.

To see this, suppose that students are of three ability types (high, medium, and low) and the distribution of ability is independent of one's position in the alphabet. Suppose further that all high-ability students, irrespective of their last name initial, are admitted to highly selective programs and that all of the low-ability students end up studying in least selective programs. For the medium types, however, given the limited supply of educational services, being sorted low in the alphabet leads to lower chances of access to selective schools. Therefore, there will be a higher-than-average ability of students with last names sorted low in the alphabet in less selective schools (thanks to medium-ability Zs) and a higher-than-average ability of students with last names sorted low in the alphabet admitted to selective schools (thanks to medium-ability As).

Our education data thus allow us to test two immediate predictions of the alphabet-related-admissions hypothesis, albeit each on a different level of schooling. First, we can use our detailed data on the college-admission process to focus on marginal applicants (medium ability types) and ask about the direct admission effect of the alphabet.[19] Second, we can use secondary-school test scores

available at `http://home.cerge.cuni.cz/munich/alpha1.html`.

[19] We would ideally like to measure the alphabetical effect on admission in faculties (or departments) that are using alphabetical sorting in their admission procedures. Here, we face three fundamental difficulties. First, our data only tell us what faculty a given student applied to while there are often department-specific admission procedures in place. Second, the unique student data we have been able to access in 2005 comes from 1999 and schools do not keep records of admission organizational practices. Third, our preliminary testing revealed that it is often difficult to

to ask about ability sorting across differently selective schools.[20] While we do not observe students in the least-selective no-exam apprenticeship programs (which cover about 40% of students), there is substantial variation in excess demand among the 1,642 secondary schools with the 'Maturita' exam that are covered in our data (see Section 4).

3 College Admission Analysis

Our hypothesis, based on the anecdotal evidence discussed above, is that one's position in the alphabet according to last name initial affects one's chances of college admission for applications that are on the margin of being admitted. Ideally, we would identify marginal applications using admission test scores from each department or faculty. In the absence of this information, we predict students' admission chances using their test score from the end-of-secondary-school national 'Maturita' exam, to control for student ability, and the average success rate in college admission of all students from each secondary school, to control for school-quality and reputation effects. In addition, we also use students' gender and age to predict admission chances. Our strategy is then to ask about the effect of the alphabet on admission decisions in different parts of the distribution of predicted admission chances. We expect the effect to be present for those in the central part of such faculty-specific distribution, i.e. for those who are neither highly likely nor highly unlikely to be admitted to a given faculty, given our measures of their ability and secondary school quality.

Specifically, we first estimate admission equations separately for each faculty using a linear

ask department or faculty officials specific questions about the use of the alphabetical order in a manner that does not reveal our research question and therefore does not lead to possibly selected response rate. Below, we therefore analyze the whole population of Czech university faculties keeping in mind that our results will reflect the likely mix of schools that do and do not use alphabet-based admission procedures.

[20] We do not have information on student quality before entering secondary schooling and do not observe applicants who were not successful in getting admitted; hence, we cannot focus on the marginal applicants to secondary schools in order to ask about the alphabet's potential direct effect on secondary-school admission.

probability model and controlling for our student/school quality measures.[21] Next, we assign each application a within-faculty percentile ranking according to its predicted probability of admission. Such percentile rankings are comparable across schools in the sense that they allow us to separately analyze groups of applications that are close to the admission margin[22] or are very likely or very unlikely to be admitted. In a second step-regression, we then re-estimate the admission equation, this time on the pooled sample of all faculties/universities and while controlling for one's position in the alphabet, but we conduct this second-step regression separately for parts of the percentile ranking distribution. In short, the second-step regression asks about the predictive power of one's position in the alphabet on admission chances, given that the application in question is likely to be in the marginal-acceptance group (at a given school). In this pooled-sample specification, we do not include our applicant/secondary-school quality measures, but we additionally control for the overall level of excess demand at a given tertiary school.[23]

An important aspect of our specification choice is how we control for one's position in the alphabet. The simplest approach is to include the numerical position (1 to 26) of one's first- and last-name initial. However, this could be a very noisy measure of one's position in the alphabet in a school-specific admission process. First, each letter in the alphabet represents a population group of different size (Figure 1). One should therefore construct an alternative measure reflecting the fraction of population with last (first) name initial sorted higher in the alphabet. Second, to the extent that school-specific groups of applications do not closely mimic the population-wide

[21] The two test scores and the school average success were positive and statistically significant in the vast majority of these school-specific regressions. The analysis in this section is not materially affected by using a Logit specification in place of the linear model.

[22] Note that the average predicted probability of admission is equal to the ratio of admitted cases to all applications to a given faculty. Hence, the median predicted probability corresponds to the margin of admission.

[23] Our excess demand measure is the ratio of rejected to accepted applications. It helps to predict admission chances and it could not be used in the faculty-specific first-stage regressions. We enter this measure in a flexible non-parametric way—as a step function corresponding to quantiles of school-specific excess demand.

alphabetical structure, one should also construct such distribution-based alphabet-position measure separately for each school.[24] To suggest the extent of measurement error bias, we compare results

Table 2: University Admission Regressions

	All Applications	Percentile Rank $\leqq$ 40	Pct Rank 40-60	Pct Rank $\geqq$ 60
	(1)	(2)	(3)	(4)
	Alphabet Position Measure Based on Letters' Numerical Order			
Last Initial	-0.006	-0.001	**-0.026**	-0.009
	(0.002)	(0.000)	(0.012)	(0.009)
First Initial	-0.002	-0.006	-0.006	-0.001
	(0.006)	(0.0003)	(0.016)	(0.015)
	Alphabet Position Measure Based on Population-Distribution Order			
Last Initial	-0.005	-0.001	**-0.022**	-0.008
	(0.005)	(0.006)	(0.009)	(0.008)
First Initial	-0.005	-0.007	-0.004	-0.005
	(0.005)	(0.005)	(0.011)	(0.011)
	Alphabet Position Measure Based on Faculty-Specific-Distribution Order			
Last Initial	-0.005	-0.002	**-0.022**	-0.007
	(0.005)	(0.005)	(0.009)	(0.008)
First Initial	-0.006	-0.008	-0.005	-0.007
	(0.005)	(0.005)	(0.009)	(0.008)
N	116,479	46,059	22,890	47,530

Note: Robust standard errors allowing for clustering of unobservables at the last-initial level are in parentheses. Bolded coefficients are statistically significant at the 5% level. Columns (2) to (4) correspond to different parts of the predicted-admission-probability distribution, which is based on student and school characteristics. Each regression additionally controls for a step function in faculty-specific excess demand (oversubscription).

based on all three types of measures, all scaled to give one's alphabetical percentile position between

[24] Since our data contain name initials only for those applicants who have graduated from secondary schools in the same year, even this most detailed measure can be noisy.

0 and 1.

Table 2 shows the complete set of second-stage coefficients of one's position in the alphabet, both in terms of first and last name initial. The three horizontal panels distinguish between the three different types of position measures we use. Each column corresponds to a different part of the predicted admission distribution: the first column gives estimates of interest based on the complete sample of all applications. Column (2) then provides alphabetical parameters from regressions based on the sub-sample of applications which fall below the 40th percentile of school-specific predicted admission chances. Next, columns (3) and (4) correspond to percentile ranges 40-60 and over 60, respectively. Our hypothesis is that marginal cases (those in the middle of the predicted admission distribution) should be affected by one's last-name, but not first-name, initial.

It is clear that there is a statistically significant negative effect of being sorted low in the alphabet on admission chances for those applications that are close to the center of the predicted admission distribution.[25] The results are robust to the use of alternative alphabetical-position measures and they are also not sensitive to additionally conditioning on applicants' quality controls (student test scores and school average success rate—these results are available upon request). Finally, in none of the estimated specifications did the first-name initial position play any role, which is reassuring for our interpretation of the estimates.

The size of the effect implies that among marginal applications, moving from A to Z reduces admission chances by over 2 percent. This is not a negligible effect, especially given that it likely reflects a mix of schools which do and do not use alphabetical sorting in their admission procedures.

[25]Statistical inference is only little affected by the clustering of unobservables at the level of last-name initial, motivated by the grouped variation of this regressor; we obtain qualitatively similar results when clustering at the level of individual students, which reflects the likely correlation of unobservables across applications submitted by the same student. We also note that the reported standard errors for the first-name initial are also clustered at the last-name initial level and are thus not correct. A more conservative (and in this case unnecessary) approach would be to base first-name-initial inference on re-estimating each specification with first-name clustering.

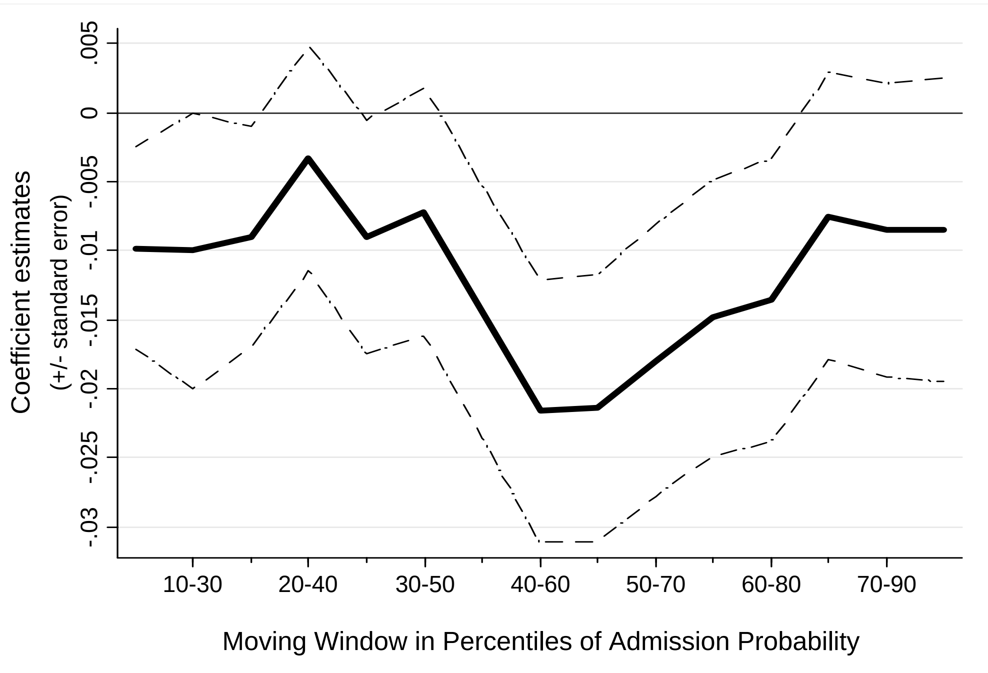

Figure 2: Last-Name Initial Coefficients Across Predicted Admission Distribution

For comparison, increasing one's 'Maturita' mathematical test score by one standard deviation leads to increasing the admission chances by 1 percent.[26]

Our choice of the 40-60 percentile range is obviously arbitrary. Next, we have therefore re-estimated the second-stage regression (with the most detailed third alphabet position measure) for a set of double-decile (moving) windows in predicted percentile position. We display the estimated last-name-initial coefficients in Figure 2. It is clear that the negative impact of last-name initial is strongest in the middle of the predicted admission-chances distribution, i.e. for the marginal cases, while it is close to 0 both for those applications that are very likely and those that are very unlikely to get accepted.

We have conducted several additional sensitivity checks (using the third, most detailed measure

[26]Both the standard deviation and the coefficient estimate of the mathematical test score correspond to the 22,890 applications in the 40-60 range.

of one's position in the alphabet). First, we noted that the second step of our analysis, based on all individual applications, implicitly weights school-specific admission practices by the size of each school-specific pool of applications. This is an optimal strategy to the extent that the first-stage faculty-specific prediction regressions, which we use to identify marginal applications, are more precisely estimated for larger application groups. As a robustness check, we have re-estimated the second-stage regressions for the marginal applications using 100 cases on each side of the median predicted admission probability of each school. This way, we work with approximately the same number of marginal applications as in column (3), but each faculty has the same weight in the regression. We again obtained a statistically significant last-name-position coefficient of -0.022 and an insignificant parameter estimate for the first-name initial position.

We have also alternatively identified marginal applications using a range based not on the percentile ranking of applications, but based on the predicted probabilities of admission themselves. Specifically, we have re-estimated our second-stage regression on the sub-sample of 35,213 (18,179) applications with predicted admission chances ranging within 0.1 (0.05) of the average predicted probability at each school. We have again obtained small and insignificant first-name coefficients while the last-name parameter was -0.013 (-0.020) with a corresponding p value of 0.08 (0.02). In sum, it appears that the main finding is very robust to the way we identify the marginal group of applications.[27]

Finally, in order to illustrate the importance of the so-far maintained parsimonious linear-effect assumption, we have also estimated the second-step regression with a step function in last-name

[27] We have also separately estimated our specifications using the third, most detailed position measure for the Philosophical Faculty of Charles University—the faculty, which openly features alphabet-based tie-breaking practices (see n. 18). Using all 2,962 applications we obtain a large negative last-name initial coefficient of -0.037 with a corresponding p value of 0.028. This coefficient becomes somewhat smaller as we "zoom in" on the marginal applications, but remains above the population-wide estimates presented in this Section. The first-name coefficients estimated for this particular faculty are invariably small and statistically insignificant.

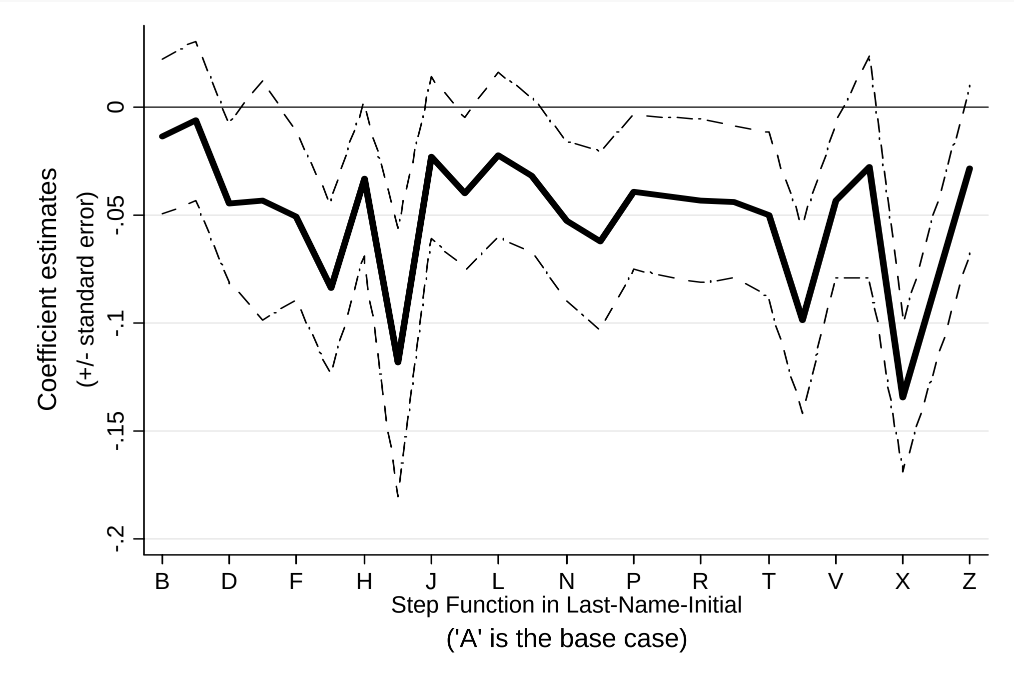

Figure 3: Non-Parametric Specification for the 40-60 Percentile Range

initial. In Figure 3 we present the step-function coefficients estimated off the 40-60 predicted-admission-probability percentile region and therefore corresponding to the linear coefficient of -0.026 in column (3) of Table 2. While there are strong 'spikes' for specific letters, the displayed pattern is broadly consistent with the linear-effect assumption; replacing one linear term with 25 dummy variables increases the R2 of the regression very little: from 0.1820 to 0.1829.[28]

4 Test Score Analysis

As we argue in Section 2.3, alphabet-based admission procedures would lead to a positive correlation between ability and one's numerical position in the alphabet among students admitted to highly

[28] The figure may suggest a special role for the first three letters (A, B and C); however, introducing a dummy variable for the first three letters into the regressions in column (2) results in an insignificant dummy coefficient and affects neither the magnitude nor the significance of the last-name-initial linear coefficient.

selective schools as well as among students admitted to the most accessible schools. While we cannot directly test for the importance of the alphabet for secondary-school admission decisions (due to our inability to observe all applications and identify marginal applicants), our data do allow us to assess the presence of ability-alphabet sorting among students already admitted to secondary schools. In particular, assuming that 'Maturita' test scores reflect ability as of the time of admission,[29] we can ask whether among students admitted to highly or little over-subscribed schools those with last names sorted low in the alphabet have higher test scores. We do not observe students in the least selective apprenticeship programs, but our data do cover 60% of the entire student cohort in all 1,642 4-year schools with school-leaving examination. These schools display dramatic variation in the degree of student selection: The school with the median value of excess demand rejects 19 students for every 100 admitted ones. In comparison, a school at the 90th (10th) percentile of the school-specific excess demand distribution rejects 43 (3) students.[30]

We first ask whether there is a positive relationship between test scores and one's numerical position in the alphabet in the whole sample of 1999 test scores and by school type. Next, we offer a stronger test of our hypothesis by interacting our excess demand measure with one's position in the alphabet. If the reason why one's alphabetical position plays a role for one's study results is admission policy in selective schools, then such interaction should be strong and positive.

We find that more selective schools indeed do display higher test scores (and presumably ability) for those of their students that have last names sorted low in the alphabet. Tables 3, 4, and 5 bear out this claim. Specifically, Table 3 presents regression coefficients of interest from the basic

---

[29] This assumption is problematic to the extent that different schools improve students' test scores differently. However, this problem is diminished when we estimate our regressions for each school type separately. Furthermore, all of our alphabet-related estimates reported below are robust to the inclusion of school fixed effects, both in terms of statistical significance and coefficient magnitude.

[30] These statistics are based on assigning each individual school the district and school-type specific value of the excess demand measure (see Section 2.1).

Table 3: Mathematics Test Score Regressions

School Type	All	Academic	Specialized	Apprenticeship
	(1)	(2)	(3)	(4)
Alphabet Position Measure Based on Letters' Numerical Order				
Last Initial	**0.748**	**2.514**	0.032	0.049
	(0.241)	(0.678)	(0.293)	(0.228)
First Initial	-0.040	0.692	-0.515	0.274
	(0.168)	(0.523)	(0.300)	(0.293)
Alphabet Position Measure Based on Population-Distribution Order				
Last Initial	**0.566**	**2.033**	0.198	0.006
	(0.231)	(0.597)	(0.259)	(0.228)
First Initial	-0.167	0.233	**-0.566**	-0.273
	(0.146)	(0.417)	(0.237)	(0.213)
N	91,599	19,174	48,594	23,829

Note: Robust standard errors allow for clustering at the last-initial level. Bolded coefficients are statistically significant at the 5% level. Each regression also controls for students' gender, a Prague dummy, and school-type dummies.

mathematics-test-score regressions, while Table 4 replicates this analysis for the Czech language test scores. The two panels of each table correspond to the two population-wide measures of one's position in the alphabet introduced in Section 3. In the first column of each table, we present the name-initials coefficients estimated off the entire sample of tested students. The estimates, which are not sensitive to the use of alternative measures of alphabetical position, suggest that having a last name initial sorted low in the alphabet is correlated with high test scores in both mathematics and Czech language tests. Columns (2) to (4) of each table then ask the same question separately for each school type. We see that it is in the most selective schools, i.e. in the general academic programs, where the data show a strong relationship between test scores and last-name-initial alphabetical position. When using the population-based position measures, we obtain a puzzling negative estimate of the first-name-initial position in the specialized schools, representing the sole violation of our natural specification test.[31] The last-name-initial effects are not only statistically, but also economically significant: Moving from 'A' to 'Z' increases the predicted mathematics test score in the general (academic) programs by 2 to 2.5 points on the 0 to 100 test score scale, corresponding to a rise from the median to the 55th percentile on the score distribution. The size of the Czech-language effect is similar.

Table 4: Czech Language Test Score Regressions

School Type	All	Academic	Specialized	Apprenticeship
	(1)	(2)	(3)	(4)
Alphabet Position Measure Based on Letters' Numerical Order				
Last Initial	**0.465**	**0.940**	0.263	0.474
	(0.204)	(0.482)	(0.247)	(0.356)
First Initial	-0.312	-0.303	-0.548	0.173
	(0.209)	(0.328)	(0.315)	(0.448)
Alphabet Position Measure Based on Population-Distribution Order				
Last Initial	**0.381**	**0.869**	0.190	0.363
	(0.185)	(0.375)	(0.147)	(0.311)
First Initial	-0.340	-0.341	**-0.537**	-0.061
	(0.265)	(0.209)	(0.237)	(0.310)
N	91,599	19,174	48,594	23,829

Note: See notes to Table 3.

Finally, in Table 5 we offer the last test of our sorting hypothesis by interacting the school-type- and district-specific excess demand measure with one's position in the alphabet. Excess demand likely proxies for ability of admitted students and is therefore also separately controlled for. Our preferred specification, presented in the table, is one where we impose no effect of last (or first)

[31] Clustering at the level of first-name initials does not lead to a loss of statistical significance for these coefficients.

name initial in schools with zero excess demand.[32] There are strong positive coefficients on all of the estimated interaction terms between excess demand and last name initial position in the alphabet. The higher the extent of student selection at school entry, the stronger the test scores of those situated towards the bottom of the alphabet. In contrast, none of the first-name interactions are important. The interaction terms are stronger in the sub-sample of students of the 320 Czech academic secondary programs[33] (in columns 3 and 4).[34]

Overall, we find that students with surnames sorted low in the alphabet do achieve higher test scores and that this "effect" is stronger in more over-subscribed schools. As we can think of no alternative explanation, we find these results strongly consistent with the ability-alphabet sorting hypothesis and therefore suggestive of the presence of alphabet-based admission procedures at the secondary school level.[35]

[32] In unreported regressions where we enter not only the interaction between last name initial and excess demand, but also one's last name position separately, both of the last-initial coefficients end up below conventional levels of statistical significance with one exception of a positive interaction coefficient in the mathematics regression based on the population-order measure.

[33] Our excess demand measure varies within this school type from 0.10 at the 10th percentile to 0.53 at the 90th percentile.

[34] To illustrate the size of the interaction term for the first alphabetical-position measure, consider the mathematics 'effect' of moving from 'A' to 'Z' in a typical (median) academic program (with 3 students rejected for every 10 admitted) and in a highly selective program featuring a 90th-percentile student rejection rate (5 rejected for 10 admitted): the 'effect' is 2.0 in the first school and 3.4 in the latter.

[35] We also note that this secondary-school evidence makes the finding of a negative effect of being sorted low in the alphabet on college admission chances stronger. Selective secondary schools are more likely to see their students apply for university admission. If the 'Z' students in selective secondary schools are indeed more able than the 'A' students, as this section suggests, than they should have higher, not lower chances of being admitted to colleges. The finding of a significant negative effect on college admission of being low in the alphabet is therefore strong evidence of alphabet-sorting admission procedures.

Table 5: Test Score Regressions with Excess-Demand Interactions

Test Type	Mathematics	Czech	Mathematics	Czech
School Type	All	All	Academic	Academic
	(1)	(2)	(3)	(4)
Alphabet Position Measure Based on Letters' Numerical Order				
Last Initial * Excess Demand	**2.339**	**1.272**	**6.817**	**2.630**
	(0.708)	(0.568)	(1.637)	(1.122)
First Initial * Excess Demand	1.010	-1.181	2.241	-1.267
	(0.823)	(0.909)	(1.175)	(1.634)
Alphabet Position Measure Based on Population-Distribution Order				
Last Initial * Excess Demand	**1.738**	**1.075**	**5.421**	**2.358**
	(0.687)	(0.481)	(1.519)	(0.876)
First Initial * Excess Demand	0.236	-1.185	0.965	-1.129
	(0.456)	(0.703)	(1.445)	(1.190)
N	91,599	91,599	19,174	19,174

Note: See notes to Table 3. The excess demand measure is the ratio of the number of rejected applications over the number of admitted ones.

5 Wage Analysis

An interesting question is whether the consequences of such admission procedures can be detected in labor-market outcomes of the adult population. In particular, our sorting hypothesis predicts that one's position in the alphabet is correlated with one's ability within groups of workers defined by the degree of selectivity of their schooling (within group, those workers with last-name initial high in the alphabet should have lower ability).[36] This obviously depends on the extent to which

[36] Whether one's position in the alphabet is a predictor of wages *on average,* beyond its effect through educational attainment, is a more complicated question. If wages rise with ability the same way for workers with different education degrees (levels), one would not expect any average wage effect of the alphabet after controlling for educational attainment.

wages reflect ability and also presumes the existence of alphabet-based admission procedures in history—affecting all age groups in the labor force. In this section, we use household survey data to ask this question empirically.

Specifically, we use a Czech retrospective survey data collected from over 3 thousands households in December 1996.[37] We note that while our education-attainment analysis is based on detailed administrative population data, our wage sample is small and likely affected by non-response and wage mis-reporting.[38] Our wage analysis focuses on males because of the complications that marriage (change of last name) brings to the analysis of adult females' alphabetical position on labor market outcomes.[39] Our wage measure consists of a monthly gross salary adjusted for daily hours worked. We observe 1,852 employed male workers aged 16 to 60 in 1996.[40] The mean log CZK (Czech Crown) monthly wage rate is 8.86.

---

[37] The survey's 3,157 households were randomly selected based on the sample frame of the official Czech Labor Force Survey. The sample is representative of the 1996 population in terms of major demographic characteristics. These data have been used in, e.g., Münich et al. (2005) and we refer the reader to the more detailed data description provided there.

[38] Focusing on wages, we ignore the potential effect that one's position in the alphabet has on participation, both directly through a potential effect on hiring (from a sorted list) and indirectly through schooling attainment. This omission is driven by our household survey data where distinguishing unemployment from being out of the labor force is difficult. We also do not report estimates of the direct effect that one's position in the alphabet could have on educational attainment. Our college-admission analysis points out that the alphabetical order matters only for marginal applicants. Similarly, our secondary-school analysis highlights that alphabetical effects are strong in only a subset of the schools. Hence, it is unlikely that there would be an average effect on educational attainment. Indeed, we are unable to estimate any alphabet-related parameters with any degree of precision using our small household data.

[39] We compared the last-name initial distribution from our wage data to that derived based on the population register. The correlation of each letter's share in the Czech population and our sample is 0.96. We conclude that the omission from our sample of males who did not engage in any employment in our sample frame does not affect the 'alphabetical' composition of our data.

[40] Our cross-section of employed workers is somewhat larger than that based on the same household data and used in Münich et al. (2005). We drop fewer missing-value observations because our analysis requires fewer variables and we do not impose constrains on typical hours worked.

Our strategy is to estimate simple log-wage regressions where in addition to standard Mincerian controls (education and a quadratic in potential experience and a dummy for the capitol city of Prague), we also condition on two name-initial variables indicating one's position in the alphabet. We again rely on two alternative measures: (i) the numerical position (1 to 26) in the alphabet of one's first and last name initial, and (ii) a population-distribution based measure which gives the fraction of the population with first or last names sorted higher in the alphabet for each worker.[41] Again, we view the coefficient on the first-name initial as a natural check of our identification strategy.

The sorting hypothesis implies a positive effect of the alphabet within both highly and least selective schools. Given the small size of our wage data, however, we cannot afford to separately estimate our wage regression for detailed school types. The simplest approximation of school type, related to the degree of student selection, is to distinguish between the highly selective general secondary programs and universities and all other education programs (elementary, apprenticeship and specialized). We follow this division in Table 6, where column (1) refers to estimates based on all of our our wage observations, while columns (2) and (3) divide the sample based on education type.

While the estimates based on the small group of workers with selective education in column (3) are very noisy, we find a positive last-name-initial coefficient in column (2) for workers with less selective education. The coefficient based on the more precise alphabet position measure is significant at the 10% level using robust standard errors.[42] The parameter estimate implies a 5.2% wage increase associated with "moving" from 'Z' to 'A' among workers with lower education, which is an effect almost identical to the benefit of a year of education estimated on the whole

[41] Again, both alternative measures are normalized to be between 0 and 1.

[42] However, applying the more appropriate (conservative) clustering of residuals at the level of 26 last name initials, the p value on this coefficient increases from 0.092 to 0.120.

Table 6: Log Wage Regressions

	Whole Sample	Less Selective Schools	Highly Selective Schools
	(1)	(2)	(3)
Alphabet Position Measure Based on Letters' Numerical Order			
Last Initial	0.039	0.056	-0.049
	(0.036)	(0.037)	(0.120)
First Initial	0.012	-0.003	0.083
	(0.040)	(0.041)	(0.126)
Alphabet Position Measure Based on Population-Distribution Order			
Last Initial	0.042	**0.052**	-0.009
	(0.030)	(0.031)	(0.099)
First Initial	0.011	-0.001	0.060
	(0.029)	(0.030)	(0.089)
N	1,852	1,567	285

Note: Bolded coefficients are statistically significant at the 10 % level. Robust standard errors in parentheses. Additional controls are years of education, experience, its square, and a Prague dummy.

sample of 1,852 workers. Again, the first-name-initial coefficients never reach even marginal levels of statistical significance.[43] Overall, we find our wage-structure estimates to be mildly supportive of the ability-alphabet sorting prediction.

6 Conclusions

While economists have explored the labor-market effect of racial attributes of first names (Bertrand and Mullainathan, 2004; Fryer and Levitt, 2004) and have asked about the incidence on women

[43] As a further robustness check, we have estimated our specifications on the sample of employed married women. Marriage and name change should render the effect of one's last name smaller and possibly insignificant. Indeed, none of the (unreported) female 'alphabetical' coefficients we obtained reached even marginal levels of statistical significance. The sample of employed single women was too small (at 340) to allow for effective estimation.

changing their surname at marriage (Goldin and Shim, 2004), no attention has been paid so far to potential effects stemming from the widespread use of the alphabetical order.

In this paper, we are fortunate to access unique administrative data that report name initials. We find evidence highly suggestive of the use of the alphabet in admission policies of Czech secondary and tertiary schools. Among students admitted to the most selective secondary schools, those sorted low in the alphabet achieve higher test scores and presumably have higher ability. Among university applicants predicted to be close to the non-admission margin, those high in the alphabet enjoy higher chances of admission. These findings are robust to the use of different measures of one's position in the alphabet and also stand our natural test of asking about the effect of one's first-name-initial position in the alphabet, which we find to play no role. There is also some evidence that conditional on low education attainment, i.e. not being admitted to higher school levels, wages (and presumably ability) are higher for workers sorted low in the alphabet.

This set of findings can be explained by a simple model of school admission with students of three ability types (high, medium, and low) distributed independently of last name initial, where all high-ability and none of the low-ability students are admitted to selective schools, and where admission of medium-ability types is decided in a way affected by alphabetical sorting. We do not provide direct evidence of the various possible ways an alphabetical 'treatment' may be taking place in schools' admission policies. Yet, we believe that the combination of our findings and the absence of an alternative explanation lend our hypothesis substantial credibility.

Should our interpretation of the empirical findings be correct, there would be a non-negligible negative effect of apparently non-discriminatory practices for individuals with last names towards the bottom of the alphabet. Rationing of public services based on a lottery is optimal, but the use of a fixed "lottery ticket" (one's last name initial) throughout many lotteries (many schooling levels) is not fair. A simple remedy is to assign each application a numerical code at random and base sorting on this alternative lottery. We believe that our results motivate future research into the use of alphabetical listings in public decision making. For start, selective education programs are a feature of many European countries other than the Czech Republic.

Bibliography

Bertrand, M., and S. Mullainathan (2004) "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review* 94 (4), pp. 991.

The Economist (2001) "As easy as ZYX," Vol. 360, No. 8237: p.13.

Filer, R., Jurajda, Š., and J. Plánovský (1999) "Education and Wages in the Czech and Slovak Republics During Transition," *Labour Economics*, 6 (4), 581-593.

Filer, R., and D. Münich (2000) "Responses of Private and Public Schools to Voucher Funding: The Czech and Hungarian Experience," CERGE-EI Working Paper No. 160.

Fryer, R., Jr. and S. Levitt (2004) "The Causes and Consequences of Distinctively Black Names," *Quarterly Journal of Economics* 119 (3), pp. 767.

Goldin, C., and M. Shim (2004) "Making a Name: Women's Surnames at Marriage and Beyond," *Journal of Economic Perspectives* 18, 2(1), pp. 143-160.

Jacobs, N., and F. van der Ploeg (2005) "Guide to Reform of Higher Education: A European Perspective" CEPR Discussion Paper No. 5327.

Jurajda, Š. (2005) "Czech Returns to Schooling: Does the Short Supply of College Education Bite?" *Czech Journal of Economics and Finance* 55 (1-2), pp. 83-95.

Matějů, P., and J. Straková (2005) "The Role of the Family and the School in the Reproduction of Educational Inequalities in the Post-Communist Czech Republic," *British Journal of Sociology of Education* 26 (1), pp. 17-40.

McCarl, B.A. (1993) "Citations and Individuals: First Authorship Across the Alphabet," *Review of Agricultural Economics* 15: 307-312.

Moldanová, D. (2004) *Naše příjmení* [Our Surnames], Pankrac, Prague, 2nd edition.

Münich, D., Svejnar, J., and K. Terrell (2005) "Returns to Human Capital from the Communist Wage Grid to Transition: Retrospective Evidence from Czech Micro Data," *Review of Economics and Statistics* 87 (1), 100-123.

Münich, D. (2004) "School Quality and Student Choice," mimeo, CERGE-EI.

OECD (2004) *Education at a Glance*, OECD, Paris.

UIV - Ústav pro informace ve vzdělávání [Institute for Information on Education] (1999) *Školství na křižovatce* [Education System at a Crossroad], Prague.

van Ours, J., and V.A. Ginsburgh (2003) "Expert Opinion and Compensation: Evidence from a Musical Competition," *American Economic Review* 93 (1), 289-296.

Šimíková, I., Navrátil, P., and J. Winkler (2004) "Hodnocení programů zaměřených na snižování rizika sociálního vyloučení romské komunity," [Evaluatin of programs aiming at lowering the risk of social exclusion of the Roma minority], Research Institute for Labour and Social Affairs, Prague and Brno.