Admission to Selective Schools, Alphabetically

Štěpán Jurajda and Daniel Münich
CERGE-EI

September 1, 2008

Abstract

One's position in an alphabetically sorted list may be important in determining access to oversubscribed public services. Motivated by anecdotal evidence, we investigate the importance of the position in the alphabet of Czech students for their admission chances into oversubscribed schools. Empirical evidence based on the population of students graduating from secondary schools and applying to universities is consistent with the use of alphabet in admission procedures at both secondary and tertiary level. A simple student-school matching model suggests that the repeated use of such admissions implies potential efficiency losses.

JEL Codes: H49, J78, I29

Keywords: Admissions, Alphabetical order, Serial position, Order effects

Acknowledgements 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. 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 Julie Cullen, Peter Katuščák and Peter Orazem. We also thank Ondřej Šteffl and Petr Vagenknecht of Scio for access to the National Comparative Exam data. 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 authors and have not been endorsed by CERGE-EI or the GDN.

Address CERGE-EI, Charles University Prague and Academy of Sciences of the Czech Republic, Politickych veznu 7, Prague 11121, Czech Republic. E-mail: `stepan.jurajda@cerge-ei.cz`

1 Introduction

Sorting based on 'alphabetical order' is a fact of everyday life. Whether this systematic sorting provides an advantage to those positioned high in the alphabet is often the object of popular discussions.[1] Customers may choose their service provider from the top of an alphabetically sorted directory, employers using the apparently non-discriminatory alphabetical order may be more attentive to job applicants who are interviewed first, etc. Yet, so far there is little evidence on the issue, thanks in large part to lack of data with individual initials.[2]

The question of non-discriminatory sorting is particularly important when allocating a prize or distributing a rationed good or oversubscribed public service. Consider, for example, musical competitions, which have been shown to determine life-time career success of professional musicians. Even though the goal of such (blind) competitions is to reflect the quality of each player, van Ours and Ginsburgh (2003) show that the randomly assigned order in which musicians play in a competition has a strong effect on their success.

In this paper, we study sorting effects in another competition that is also meant to reflect only the quality of contestants and where sorting may also play an important role, but which can affect entire population cohorts: We analyze the student selection process in oversubscribed secondary schools and universities. Specifically, we ask whether students with last names sorted high in the alphabet enjoy higher chances of being admitted to selective schools. We find supportive evidence at both secondary and tertiary level.

Even if unequal access to quality education can have serious consequences for life-time labor market outcomes, alphabetical sorting can be defended as a way of randomizing access among

[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.

[2] The major exception are studies of citation bias against coauthors of scientific studies whose last names are sorted low in the alphabet; see, e.g., McCarl (1993), Einav and Yariv (2006) or Praag and Praag (in press).

equally talented applicants in face of capacity constraints. However, using a stylized model of student-school matching we suggest that the repeated use of alphabetical sorting at entry to both secondary and tertiary education can lead to efficiency losses.

Our analysis is based on the experience of students in the Czech Republic, which provides a useful case to study for three reasons. First, studying alphabetical sorting effects in admission procedures in the Czech Republic is motivated by several pieces of anecdotal evidence. We know of cases where lists of applicants with multiple student characteristics (including test scores) prepared for admission committees are sorted according to the alphabet. When applications are evaluated based on multiple criteria in absence of a clear summarizing measure, marginal cases at the top of such 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. A similar effect could be present in those universities, which use an oral exam and call applicants to these exams in alphabetical order.[3] Finally, in some cases Czech universities openly use the alphabetical order to break ties among applicants with identical admission test scores. We quote from the official specification of the admission procedure at one department of Charles University Prague, a prestigious Czech 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)."[4]

Second, the reliance on alphabetical sorting in admission procedures can only be of importance in a highly selective schooling system, where student rationing is extensive. The Czech Republic is a case in point as it features a highly selective admission process at both secondary and tertiary

[3] Such practice is currently used, e.g., at the Philosophical Faculty of Charles University Prague.

[4] The announcement, in Czech, was originally posted at `http://prijimacky.ff.cuni.cz` and is now available at `http://home.cerge.cuni.cz/munich/alpha1.html`.

schooling level, and thus provides a good example of other European selective education systems.[5] At 13%, the country has one of the lowest tertiary attainment rates in the OECD (OECD, 2007)[6] and students entering the university system typically come from selective academic secondary programs serving less than 15% of each cohort of secondary-school students. Tuition-free public universities provide the bulk of tertiary education and they tend to reject about a half of applicants each year. In 1999—the year our data come from—55% of all applicants to Czech universities were not able to enroll in any program.

Third, there is unique administrative data available on study achievement and university admission experience of the whole population of Czech secondary-school graduates in 1999. Specifically, we observe national school-leaving-exam test scores from mathematics and the native Czech for all graduates from secondary programs, together with their initials. For all of these high-school graduates, we also see which universities they apply to, together with the admission decision.[7]

To provide a framework for our analysis, we build a simple model of student-school matching, where alphabetical sorting is allowed to play a role for applicants on the margin of admission. There is a *group* of such marginal applicants thanks to either noisy admission exams or discrete support of the admission 'score' measure. If only a part of the marginal group can be admitted due to capacity constraints, alphabetical sorting is invoked, either overtly or covertly. The model implies that in presence of alphabet-based admission practices at selective schools, students admitted to such schools with last names in the bottom part of the alphabet should on average have higher ability and that this sorting should be stronger in more selective schools. Next, we note that the presence of such alphabet-ability sorting in selective high-school programs, which prepare students

[5] School-specific selection of applicants at the undergraduate level is applied in Germany and the United Kingdom; at the master level, it is applied in most EU countries (Jacobs and van der Ploeg, 2006; Aghion et al., 2008).

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

[7] We do not observe students of apprenticeship programs, which do not lead to a school-leaving exam and do not send students to universities.

for university education, has consequences for college admissions based on noisy entrance exams. Among marginal university applicants, a 'Z' applicant from a selective high school is likely to be of higher ability compared to an 'A' applicant from the same program and this information should be used to improve the efficiency of student-school matching.

Our empirical analysis starts with the national study-achievement tests administered to the population of Czech students graduating from secondary schools in 1999. We find evidence consistent with the student sorting predicted by our model: students with 'Z' last-name initials perform better in tests compared to 'A' students and this performance gap is larger in more selective schools, consistent with the use of alphabetical sorting in admissions at selective secondary programs. Next, we study the success of student applications to universities. We find a significant negative effect of being sorted low in the alphabet according to last-name initial on admission chances of marginal applicants. Throughout the empirical 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.

One could argue that if schools select among applicants of similar ability based on the alphabetical order, it is only one of many alternative random justifiable ways of rationing. However, our model implies that the use of alphabet-based sorting at both secondary and tertiary level is likely to lead not only to distributional but also efficiency consequences. We conclude the paper with a calibration exercise aimed at quantifying the extent of inefficient matching of students with universities and find that the efficiency loss is likely to be small.

The paper is organized as follows. In the next section, we describe our data and use them to offer stylized facts about the Czech education system. The theoretical framework and testing strategy are outlined in Section 3. Sections 4 and 5 present the test-score and college-admission analysis, respectively. The calibration exercise is presented in Section 6 and some tantalizing wage analysis based on a 1996 household survey is provided in Section 7. The last section concludes.

2 The Czech Education System and our Data

The structure of the Czech educational system parallels those of other European countries. All universities are public and tuition-free.[8] Despite significant growth in total enrollment during the 1990s, about a half of applicants were not able to enroll in any university program as of the end of last century and the tertiary attainment rate of the Czech population aged 25-34 remains starkly low at 14% even as of 2005 (OECD, 2007).

The low tertiary attainment rate is perhaps not surprising given that a major group of secondary-level students attends apprenticeship programs, which offer dismal prospects of continuing on to higher education. Most of the apprenticeship schools offer three-year programs, which do not lead to a school-leaving comprehensive examination, 'Maturita' in Czech. These exams, administered at the end of 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 or the German 'Abitur' exam and form a pre-requisite for tertiary education.[9]

Students taking the 'Maturita' exams come from three types of Czech four-year secondary programs: apprenticeship, specialized and academic. Four-year apprenticeship programs leading to the 'Maturita' exam typically focus on craft skills. Examples of specialized secondary programs include construction or nursing schools. The academic programs are typically strong in both humanities and mathematics and offer the best chances of continuing on to university education.[10]

---

[8] Enrollment in private colleges emerged only after 1999. Even today, private tuition-based tertiary education remains miniscule in the Czech Republic.

[9] 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.

[10] Filer and Münich (2000) offer a detailed description of 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 1,642 schools. Exams were held simultaneously and the results were processed centrally.

Our data provide standardized test scores (on a 0 to 100 scale) corresponding to students' mathematics skills and to their command of the native Czech language. 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. Out of the total of 105,979 tested students, we observe name initials for over 97 thousand students and among these, 90,597 have valid test scores from both mathematics and Czech language tests available. 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.[11]

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. Students graduating from academic programs also have by far the highest chance of being admitted to universities. It is therefore not surprising that academic high schools are in excess demand, as illustrated by the 1998 admission rates for each type of secondary program presented in the last row of Table 1.[12]

[11] In the Czech Republic, there are no types of last names related to a history of family wealth such as "van" or "von" (Moldanová, 2004). The country is also highly ethnically homogenous, with only one sizeable minority, the Roma. However, with minor exceptions, Roma students do not enter selective secondary schools or universities (Šimíková et al., 2004).

[12] The admission-rate differences are smaller than expected due to binding limits on the number of secondary-school applications a student is allowed to submit. These limits lead to strategic misrepresentation of preferences; they lower

The reported standard deviations of admission rates within school type reflect regional variation in cohort size and location and quality of individual programs.[13] The admission process is governed independently by individual schools, which base their admission decision on entrance exams and grades from elementary education.

The school-leaving test score differences presented in Table 1 are consistent with more selective secondary schools having higher admission standards. We confirm this intuition using a survey of the academic programs administered in 1996, which reports elementary-school grades of admitted students (UIV, 1996); specifically, we find that elementary-school grades of the weakest three students admitted to an academic secondary program are significantly better the more selective the program is. Further evidence along these lines is shown in Figure 2, which is based on the OECD's international PISA survey from 2003. The figure presents the distribution of PISA test scores among 2,180 15-year-old elementary-school graduates applying to Czech secondary programs leading to the 'Maturita' exam.[14] Applicants indicate the school type of their first choice. The vertical line in each graph corresponds to the share of all applicants admitted. Two facts stand out. First, PISA test scores at the margin of admission are indeed higher in the more selective academic programs. Second, the density of applicants' ability at the margin of admission, as reflected by the PISA tests, is higher in the more selective type of programs. It may therefore be harder for more selective schools to distinguish applicants' quality within the group of marginal applicants.

---

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.

[13] Although we follow all students in each secondary school, such that we know the school-specific success rate in college admissions, we do not observe the admission rates of individual secondary schools themselves. We have available admission rates for each secondary program type for each of 76 Czech districts (NUTS-4 territorial units). We pool the districts falling within the capital city of Prague into one district. The average district population size (excluding Prague) is approximately 100 thousand.

[14] We do not cover applicants to the apprenticeship programs that apply almost no selection.

2.2 College Admissions

Secondary-school graduates can submit an unlimited number of university applications. An application process typically consists of a written exam and in a subset of faculties includes also oral exams. We work with the 1999 administrative register of individual university applications covering all 116 distinct faculties of Czech public universities.

Competition for university education is fierce as only 29% of all applications were admitted. This is not surprising given that universities are tuition-free and given the strict quotas on total enrollment set by the Ministry of Education. In fact, universities are penalized for each additional student enrolled beyond the quota limit. Looking across the 116 faculties, 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.

We have merged the register of university applications with the secondary-school test data described above. The college-application register reports the success or failure of each individual application (whether a given student was admitted to a particular school), but falls short of providing faculty-specific admission test scores and does not give name initials. We therefore focus our analysis of college admissions on those 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 such "fresh" secondary-school graduates constitute 55% of all applications and 61% of university admissions in 1999.

The resulting data provide information on a total of 116,479 applications submitted by 41,486 secondary-school graduates. We exclude from analysis the approximately 25 thousand applications where the student did not show up to take the exam; we therefore work with 89,443 applications.

3 The Use of Alphabet and Testing Strategy

What would be the consequences of alphabetical sorting in school admissions? Clearly, under the assumption that ability and last-name initials are independent and that students do not adjust their application strategy based on their position in the alphabet,[15] the presence of an alphabet-affected admission process would lead to a negative correlation between being admitted to selective schools and one's position in the alphabet, conditional on applying. An interesting consequence of such admission processes is that there would also be a positive correlation between ability and one's numerical position in the alphabet among students admitted to highly selective schools as well as among students enrolled in easily accessible schools. To see this point in a simple setting, 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 the 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 within both less selective schools (thanks to medium-ability Zs) and more selective schools (thanks to medium-ability As).

To present this argument more formally and to gain further insight into the consequences of alphabet-based school admission, we build a simple model of school-student matching, in which admissions aim to select the most able applicants using noisy entrance exams. We assume that schools admit students based on an admission test score S, which reflects students' ability a (distributed as standard normal[16]), and that the test score S has only discrete support. Selective schools are

[15] We are not aware of any public discussion of the issue of alphabet sorting in admission procedures in the Czech Republic. It appears that neither the students nor the schools consider this issue important.

[16] This assumption is in line with the distribution of PISA test scores of secondary-school applicants presented in

limited in the number of students they can admit such that they directly admit all applicants with a test score strictly above an admission threshold score S^T and admit only a fraction of marginal applicants who scored exactly S^T on their admission exam.[17] The selection among marginal applicants is based on alphabetical sorting. In other words, admissions are decided using a lexicographic order on S and N, where $N = 1,\ 2, ..., 26$ denotes one's position in the alphabet.

This formulation of admissions captures the essence of the alphabet-based admission mechanisms discussed in the Introduction. It corresponds exactly to the practice at those schools that openly use the alphabetical order to break ties among applicants with identical admission test scores, but it can be thought of as providing a more general description of alphabet-based admission procedures. Schools using a continuous test score may consider as marginal all those applicants falling into a confidence interval implied by the presences of measurement error in S around the threshold value of the score. Such marginal applicants can be thought of as having the discrete test threshold value S^T. Similarly for applicants who appear marginal based on multiple evaluation criteria in alphabetically sorted lists or those called to oral exams in the alphabetical order.

Given that last name initials are assumed orthogonal to ability, the expected ability of a directly admitted applicant with initial N, denoted $\overline{a_{DA}}(N)$, corresponds to the formula for the expectation of a truncated standard normal distribution:

$$\overline{a_{DA}}(N) \equiv E\left[a|S > S^T, N\right] = E\left[a|S > S^T\right] = \frac{\phi(S^T)}{1 - \Phi(S^T)} \tag{1}$$

and does not depend on N. In equation (1), ϕ and Φ denote the probability density and the cumulative distribution function of the standard normal distribution, respectively. The expected

Figure 2 as well as with the distribution of 'Maturita' test scores of university applicants.

[17] We therefore assume that more selective schools have higher admission standards. This is consistent with the evidence presented in Section 2.1.

ability of marginal applicants, which also does not depend on N, can be expressed as[18]

$$\overline{a_M}(N) \equiv E\left[a|S = S^T, N\right] = S^T. \tag{2}$$

Next, consider the alphabetically sorted list of marginal applicants and denote by N^T the initial of the last marginal student who is admitted. The expected ability of admitted applicants with surname initial $N > N^T$, that is of students who are only admitted directly, equals $\overline{a_{DA}}$. For $N \leq N^T$, on the other hand, the school admits all marginal applicants with a given initial and the expected ability of all admitted applicants equals the average of the expected ability of direct and marginal admits, weighted by the population proportion of each group for each initial. Hence, the expected ability of all admitted students $\overline{a}$ can be expressed as follows:

$$\overline{a} \equiv E\left[a|S \geq S^T, N\right] = \begin{cases} \overline{a_{DA}} = \frac{\phi(S^T)}{1-\Phi(S^T)} \text{ for } N > N^T \\ \frac{\left(1-\Phi\left(S^T\right)\right)\overline{a_{DA}}+m\overline{a_M}}{(1-\Phi(S^T))+m} = \frac{\phi(S^T)+mS^T}{1-\Phi(S^T)+m} \text{ for } N \leq N^T, \end{cases} \tag{3}$$

where we denote by m the expected share of marginal applicants with a given initial on all applicants with that initial. We assume that the share m does not change with the degree of selectivity S^T and note that it does not depend on N, similar to $\overline{a_M}$. Equation (3) implies that the expected ability of admitted students is higher for those with $N > N^T$ compared to those with $N \leq N^T$. Clearly, different schools apply a different N^T threshold (there is a distribution of N^T across schools) such that the use of alphabetical sorting of marginal applicants implies a positive relationship between N and $\overline{a}$ in the population of admitted students.[19]

Interestingly, equation (3) also suggests that the ability difference across the alphabet, denoted

[18] In Equation (1), we assume that the discrete value of S corresponding to each interval of the continously distributed ability a equals the lowest value of a in the interval. Here, we make the simplifying assumption that the interval corresponding to value S^T is symmetrical around $a = S^T$.

[19] By the same argument, there ought to be a positive ability-alphabet sorting among those not admitted to selective schools.

Δ^S, which can be expressed as

$$\Delta^S \equiv E(a|S > S^T, N > N^T) - E(a|S > S^T, N \leq N^T) = m\frac{\phi(S^T) - S^T\left(1 - \Phi(S^T)\right)}{[1 - \Phi(S^T)]\left[(1 - \Phi\left(S^T\right)) + m\right]}, \quad (4)$$

grows with the degree of admission selectivity; i.e., it is a function increasing in S^T.[20] The expected ability gap between a 'Z' student and an 'A' student admitted to a selective school is higher the more selective a given school is in admitting students. This result rests on the assumption that the share of marginal applicants m does not depend on S^T. It holds more generally as long as m does not decrease relative to the share of directly admitted applicants $1 - \Phi(S^T)$ as S^T increases, that is as long as more oversubscribed schools are not disproportionately better at discriminating among applicants' ability compared to less selective programs.[21]

We are now ready to consider the consequences of repeated use of such admission procedures at the entry to both secondary and tertiary schools, where only students admitted to selective secondary schools can apply to universities. We further assume that skill production in secondary schools does not close the alphabet-ability gap described in equation (4). For the sake of simplicity, we assume that the ability of graduates of selective secondary schools, denoted $\widetilde{a}$, can be expressed as $\widetilde{a} = \delta\widetilde{N} + u$, where u follows the standard normal distribution, $\widetilde{N}$ is an appropriate linear transformation of N that guarantees that $E(\widetilde{a}) = 0$, and δ captures the positive dependence of the expected ability of students admitted to selective secondary schools on N. Following the logic provided above, and using $\widetilde{S}$ to denote test scores from a college entrance exam, the expected ability of marginal applicants to colleges is $E(\widetilde{a}|\widetilde{S} = \widetilde{S}^T, \widetilde{N}) = \widetilde{S}^T + \delta\widetilde{N}$. Clearly, colleges should select among the marginal applicants in *reverse* alphabetical order as such choice would result in higher average ability of admitted students compared to pure randomization or standard alphabetical

[20] This holds over the relevant range of S^T values, where the share of rejected applicants $\Phi(S^T)$ is above 0.3.

[21] The evidence presented in Figure 2 suggests that the density of ability at the margin of admission is higher in more selective schools. Therefore, unless the degree of noise (measurement error) in admission tests of more selective schools is dramatically smaller compared to less selective schools, the assumption is likely to be satisfied.

sorting.

How can we test the predictions of this simple model of alphabet-based school-student matching? First, our data on school-leaving exams of selective secondary school graduates allow us to ask whether 'Z' students display higher ability in 'Maturita' tests compared to 'A' students as predicted by equation (3). Furthermore, we can use the substantial variation in excess demand among the 1,642 Czech secondary schools with the 'Maturita' exam (see Section 4) to ask whether the ability gap between 'Z' and 'A' students is higher in more selective schools, as predicted by equation (4). Such evidence would be consistent with the use of alphabet, overt or covert, in high-school admissions.

Second, the information on the success of secondary school graduates in applying to universities allows us to test directly whether 'Z' applicants face higher chances of being admitted to colleges compared to 'A' applicants, conditional on having similar admission test scores. We expect only marginal applicants to be affected by their alphabetical position; hence, it is important that we can use the 'Maturita' test scores to predict who is a marginal applicant. Specifically, for each faculty, we observe the list of applicants with their 'Maturita' test scores and we know the total number of admits. We use this information, together with other predictors, to identify applicants who are likely to be close to the margin of acceptance—those at the percentile of the 'Maturita' exam distribution of applicants to a particular faculty that corresponds to the share of admitted applicants to that faculty.[22]

[22] We would ideally like to measure the alphabetical effect on admission only 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 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

4 Test Score Analysis

Our simple model suggests that alphabet-based admission procedures can lead to a positive correlation between ability and one's numerical position in the alphabet among students admitted to selective schools. Our test score data allow us to assess the presence of such ability-alphabet sorting among students of the most selective academic secondary programs and the less selective specialized and apprenticeship programs with a school leaving exam.[23] Specifically, we regress students' test scores on their position in the alphabet using the whole sample of test scores and also by school type.[24] Our main focus is on last-name initials, but we also include a measure of one's first-name alphabetical position as a natural check on our approach since we know of no reason why first-name initials should affect admission chances. Next, we offer a stronger test of our hypothesis: We ask whether the relationship between alphabetical position and test scores differs across schools that differ in how over-subscribed they are, as implied by equation (4), by interacting our excess demand measure with one's position in the alphabet.

We use two alternative measures of 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, given that each letter in the alphabet represents a population group of different size (see Figure 1), a more precise measure of one's position in an alphabetically ordered list consists of the fraction of

of schools that do and do not use alphabet-based admission procedures.

[23] We do not observe students of the short apprenticeship programs without 'Maturita' exams (see Section 2). Recall that these least-selective schools cover about 40% of the cohort. Hence, there could be some alphabet-ability sorting present across entire districts within the group of 'Maturita' students.

[24] In effect, we assume that 'Maturita' test scores reflect ability as of the time of admission. This is likely to hold within schools as students are not seated in classrooms in the alphabetical order in the Czech Republic. The assumption is perhaps 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.

population with last (first) name initial sorted higher in the alphabet. For the sake of comparability, both measures are scaled to give one's alphabetical percentile position between 0 and 1.

We find that more selective schools indeed do display higher test scores (and presumably ability) for those of their students who have last names sorted low in the alphabet. Tables 2, 3, and 4 bear out this claim. Table 2 presents regression coefficients of interest from the basic mathematics-test-score regressions, while Table 3 replicates this analysis for the Czech language test scores. The two panels of each table correspond to the two measures of one's position in the alphabet. In the first column of each table, we present the name-initials coefficients estimated off the entire sample of tested students. The parameter 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. The data imply a strong relationship between test scores and last-name-initial alphabetical position in the most selective schools—in the academic programs. The last-name-initial effects are not only statistically, but also economically significant: the gap between an 'A' and a 'Z' student in the predicted mathematics test score in the academic programs is 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.

When using the population-based position measures, we obtain a puzzling negative estimate of the first-name-initial position in the specialized schools. It turns out that this negative coefficient represents the sole violation of our natural specification test as all other first-name-initial coefficients in the subsequent analysis are not statistically significant.

Finally, in Table 4 we offer the stronger test of our sorting hypothesis by interacting the school excess demand measure with one's position in the alphabet. (We only present results based on the population-order measure.) Excess demand is captured by school admission rate, which likely

proxies for ability of admitted students and is therefore also separately controlled for in the estimated regressions. Indeed, in all estimated regressions, we find that the lower the admission rate (the higher the school's selectivity), the higher the test scores. In columns (1) and (2), we use both within-school-type and across-school-type variation in admission rates to identify the parameters of interest, i.e., the interaction of admission rate with alphabetical-position measures. Based on both math and Czech-language test-score data, we again find that students sorted low in the alphabet according to last name initial are achieving higher test scores, but this effect is lower the less selective the secondary school is (the higher its admission rate). The estimates imply that there is no alphabet-ability sorting at a school that does not select its students (has an admission rate of close to 1).

In columns (3) and (4), we estimate the interaction specifications relying only on differences in school admission rates within the group of students of the 320 academic secondary programs. The parameter estimates are all larger compared to those in columns (1) and (2), but are less precisely estimated, particularly in the Czech-language regression in column (4).

Overall, we find that students with surnames sorted low in the alphabet do achieve higher test scores on average and that this sorting "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. Of course, this interpretation of the evidence would be strengthened by a finding of no alphabet-ability sorting before the entry into secondary programs. The only existing Czech data with both name initials and ability measures for students graduating from primary-level programs (in 9th grade, before admission to secondary schools) consists of practice tests of the so-called 'National comparative exam'. These exams, conducted by Scio.cz, a private testing agency, have recently become a prototype of admission exams at many secondary schools. Unlike our main data, which covers the entire cohort of 'Maturita' students, these practice

tests are therefore taken by a selected group of highly motivated 9th graders. We regressed the mathematics and Czech language test scores of the 9,625 students who took these practice exams under certified conditions in 2005 on their last name initial position in the alphabet (first initial is not available) and found no statistically or economically significant coefficients.

5 College Admission Analysis

The administrative register of college applications and admission decisions allows us to test directly for the effect of one's position in the alphabet according to last name initial on one's chances of being admitted to over-subscribed colleges. Our hypothesis is that alphabetical sorting plays a role only for applicants on the margin of admission. Therefore, we would ideally like to identify marginal applications using scores from admission exams administered independently by each department or faculty. In the absence of this information, we predict admission probabilities for each application at each faculty using students' 'Maturita' test scores, which help us control for student ability, and using the average success rate in college admission of all students from a given secondary school, which helps us control for school-quality and reputation effects. In addition, we also use students' gender and age to predict admission chances. Note that we observe the complete pool of applications for each faculty such that the identification of marginal applications is school-specific. We can therefore test for the effect of the alphabet on admission decisions in different parts of the distribution of predicted admission chances, which is an important part of our overall testing strategy. Our hypothesis is that the 'alphabetical' effect is present only for those applications 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.

We proceed in two steps. First, we estimate admission probability equations separately for each of the 116 distinct faculties of Czech public universities. The success of individual applications is

captured using linear probability models controlling for the student and school quality measures described above.[25] 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 or are very likely or very unlikely to be admitted. (Note that the average predicted probability of admission is equal to the ratio of admitted students to all applications to a given faculty. The median predicted probability in fact closely corresponds to the margin of admission.)

In the second-step, we re-estimate the admission equation, this time on the pooled sample of applications to all faculties/universities. This pooled specification controls for one's position in the alphabet and is estimated separately for different parts of the percentile rankings. Using this second-step regression, we can ask about the predictive power of one's position in the alphabet on admission chances of applications that are likely to be in (above, below) the marginal-acceptance group. We do not include our applicant quality measures in the second-stage regression, but we additionally control for the overall level of excess demand at a given tertiary school.[26] As in Section 2.1, we use two simple measures of one's position in the alphabet. One is based on the numerical position of one's first- and last-name initial (1 to 26 divided by 26), the other reflects the fraction of the pool of applicants to a particular university program with last (first) name initial sorted higher in the alphabet.

Table 5 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 two horizontal panels distinguish between the three different types of position measures we use. Each column corresponds to a different part of the

[25] The two test scores and the school average success are 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.

[26] The results are not sensitive to including the quality measures in the second-stage regressions.

predicted admission probability 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 initial 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.[27] 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. For comparison, increasing one's 'Maturita' mathematical test score by one standard deviation leads to increasing the admission chances by 1 percent.[28]

---

[27] Statistical inference is only little affected by the clustering of unobservables at the level of last-name initial, motivated by the grouped level of variation in 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.

[28] The estimates presented in Section 4 imply that 'Z' secondary-level students are of higher ability, particularly those from the selective academic programs. Most of these students apply for colleges. In the college-admission regressions, we control for applicants' ability using 'Maturita' test scores, but these provide only an imperfect measure of applicants' university-specific ability. As long as observables are positively correlated with unobservables, we would therefore expect 'Z' students from highly selective secondary programs to be of higher university-specific ability compared to 'A' applicants even conditional on 'Maturita' test scores. Could it be that among marginal applicants from highly selective secondary schools, the negative admission-probability effect of being positioned low in the alphabet working through university admission sorting is outweighed by 'Z' applicants' higher unobservable ability? To find out, we

Our choice of the 40-60 percentile range is obviously arbitrary. We have therefore re-estimated the second-stage regression (with the faculty-specific alphabet position measure) for a set of double-decile (moving) windows in predicted percentile position. The estimated last-name-initial coefficients are displayed in Figure 3. 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. 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 -0.030 with a p value of 0.005 and a statistically insignificant small parameter estimate for the first-name initial position (-0.003).

---

interacted the last-name initial position with the selectivity (admission rate) of applicant' secondary schools. Based on the sample of marginal applicants from academic programs, we found, as expected, that the negative alphabetical effect is strongest for students from less selective secondary programs. The last-name initial coefficient was 0.12 while the interaction of the last-name position with secondary-school admission rate was -0.24; both coefficients were statistically significant at the 10% level. This implies no effect of last-name initial for applicants from academic programs with admission rates of below 50%, i.e., those one standard deviation or more below the average admission rate in these types of programs. The first-initial coefficients were both close to zero and statistically insignificant. This set of estimates reflects sorting on unobservable quality and is consistent with our overall interpretation of the evidence.

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 10,851 applications with predicted admission chances ranging within 0.05 of the actual admission rate (i.e., the average predicted probability) at each school. We have again obtained small and insignificant first-name coefficients while the last-name parameter was -0.031 with a corresponding p value of 0.02. In sum, it appears that the main finding is very robust to the way we identify the marginal group of applications.[29]

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 initial. In Figure 4 we present the step-function coefficients estimated off the 40-60 predicted-admission-probability percentile region and therefore corresponding to the linear coefficient of -0.03 in column (3) of Table 5. While there are strong 'spikes' for specific letters, the displayed pattern is consistent with the linear-effect assumption.

Overall, the evidence is consistent with the presence of a significant negative effect of being sorted low in the alphabet on admission chances of marginal applicants to colleges.

[29] We have also separately estimated our specifications for the Philosophical Faculty of Charles University—the faculty, which openly features alphabet-based tie-breaking practices. This is a highly selective school, which accepts only 12% of all applications. Using all 2,506 applications we obtain a large negative last-name initial coefficient of -0.045 with a corresponding p value of 0.02 (the first-name coefficient is small and statistically insignificant). The negative last-name effect comes from both marginal and very strong applicants, i.e., all those with predicted probability of admission above the 40th percentile. That ability sorting affects all strong applicants is perhaps plausible, given the extremely high excess demand at this school. All applicants above the admission margin are quite strong in the cohort-wide comparison; order effects working through, e.g., oral exams for 'Z' students taking place in late afternoon, can thus affect all candidates with above-marginal entrance test scores.

6 Efficiency Loss Calibration

We now return to our simple model of Section 3 to interpret the evidence suggesting that being sorted low in the alphabet leads to a disadvantage when trying to enter selective secondary as well as tertiary programs. The model suggests that in presence of noisy admission exams, tertiary schools aiming to admit the most able students should choose 'Z' marginal candidates over 'A' marginal applications, i.e., apply reverse alphabetical ordering, because of the alphabet-ability sorting present among graduates of selective secondary programs. In contrast to this efficiency prescription, our evidence suggests that both secondary and tertiary schools use the alphabetical order in the same standard fashion. We therefore ask whether, given the magnitude of our estimates, alphabet-based admission procedures could have sizeable efficiency consequences in a country, where students are subject to repeated selective screening into higher levels of education. To this effect, we calibrate and simulate the model of Section 3. Our measure of efficiency corresponds to the ability of colleges to select the most able applicants.

Our first task is to generate a simulated population of students of selective academic secondary schools that displays the same magnitude of alphabet-ability sorting that we recorded in Tables 2 and 3. We follow the descriptive statistics presented in Table 1 to generate a population of 26,000 applicants to these academic programs. The simulated applicant population displays the same distribution of surname initials as observed in our data (Figure 1). We assign each applicant i an ability index a_i drawn from the standard normal distribution. We set the admission threshold for direct admits at the 30th percentile of the ability distribution a and identify those between the 20th and 30th percentile as marginal applicants. This ensures that the simulated excess demand ratio matches the observed selectivity of these secondary programs (Figure 2) and that we generate the "right" number of admitted students. We therefore directly admit the top 70% of applicants and consider as marginal the following decile group. Marginal applicants are admitted if their

alphabetical position according to last name initial is above a randomly drawn position (integer) $r_i \in \{1, 2, ...26\}$, i.e., when $N_i \leq r_i$ using the alphabetical position notation of Section 3. The random draw of r corresponds to different schools using a different threshold initial for marginal admissions. Finally, we normalize the mean and the standard deviation of the ability distribution of admitted students to mimic the observed statistics reported in Table 1 and regress this measure of ability on the last name initial alphabetical position of admitted students. We obtain parameter estimates that are similar in magnitude to those reported for academic secondary programs in Tables 2 and 3. Our simulation therefore successfully mimics the empirical magnitude of the alphabet-ability sorting in secondary schools, despite its many simplifying assumptions; it provides a simple benchmark for assessing the size of the sorting effects estimated in Tables 2 and 3.

In the second step, we focus on college admissions. We assume that all of the students admitted to academic secondary programs in the first step of our simulation choose to apply to university after graduating. We also simulate another, similarly sized group of university applicants consisting of graduates of the specialized (technical) secondary schools. We abstract from the fact that students submit several applications and allow only one application per person. In line with the empirical admission rate of applicants, we allow half of applicants to be successful.

The ability distribution of the students of academic secondary programs at the time of graduation is generated as the sum of the ability a at admission (as of four years earlier) and a normally distributed noise with zero mean and standard deviation equal to that of the a distribution. This additional independent noise is meant to reflect the many additional determinants of students' skills that are at work during the 4-year secondary programs. We also generate the (normal) ability distribution of applicants from specialized technical schools without any alphabet sorting effects, in accordance with the parameters of Tables 2 and 3.

We are now ready to simulate college admissions as follows: the top 45% of applicants (from both academic and specialized secondary programs) are admitted directly while those in the 45-55

percentile range are considered marginal, consistent with the mechanism we applied for secondary-school admissions. Finally, we admit marginal applicants based on the alphabetical order in the same fashion that we used for simulating secondary-school admissions. To see whether the simulated 'alphabetical' effect on admission of marginal students is similar in size to that we observed in our data, we regress the college admission outcome from the simulation on the alphabetical position of applicants; we do so only for the marginal students, which now corresponds to those in the 40-60 percentile range, in line with the empirical evidence presented in Figure 2. We obtain a coefficient, which is an order of magnitude larger than that reported in Table 5. In order to reconcile our simulation with the empirical estimates, we therefore deduce that only 10% of Czech universities uses alphabetical sorting in their admissions. Here, the simulation is providing us with a formalized guess of the fraction of university faculties that use sorted lists of applicants or cut ties using the alphabetical order. We further assume that the other 90% of college admissions select marginal students at random.

We are now ready to quantify the extent of admission inefficiency for marginal college applicants from academic secondary schools. First, in the 10% of faculties that use the alphabetical order as the mechanism for admitting marginal applicants, we replace it with the reverse alphabetical order, following the prescription of the theoretical model. We find that the admission outcome under these two different selection rules differs in 70% of the cases. Second, we use the reverse alphabetical order in place of the random selection at the remaining 90% of faculties, which results in different admission results for 50% of marginal applicants. Summing up, we conclude that using the reverse alphabetical order would improve matches for 52% of marginal applicants from academic secondary programs. Given that marginal admits form about one fifth of all admitted college students in our simulation and that only half of marginal applicants come from the academic secondary programs, we conclude that the repeated use of the alphabetical order may lead to inefficient school-student matches for about 5% of students admitted to Czech universities. This is not a large effect, which is

not surprising given that only marginal university applicants from a subset of secondary-schools are affected. Nevertheless, our analysis illustrates the potential for efficiency losses from the repeated use of the same order for allocating rationed public services.

In the simulation, we abstract from the fact that secondary-school graduates can apply to several universities. If a given student is not admitted to one school on account of his last name being sorted low in the alphabet, she can perhaps get enrolled in another university program. This would make the size of the efficiency loss even smaller. On the other hand, the fact that students may not be able to enrol in their preferred programs implies another type of efficiency costs: These students will likely try to get admitted to their preferred program later on. Indeed, many Czech university graduates switch from one program to another after a year or two. Survival rates in tertiary education programs are traditionally quite low in the Czech Republic at 63%, compared to, e.g., 75% in Germany, or 76% in Turkey.[30]

7 Wage Analysis

An interesting question related to our analysis of school admissions is whether the consequences of the use of alphabetical order in admission procedures can also be detected in labor-market outcomes of the adult population. In particular, our sorting hypothesis implies 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).[31] This obviously depends on the extent to which wages reflect ability and also presumes the existence of alphabet-based admission procedures in history—affecting all

[30] Survival rates in tertiary education represent the proportion of those who enter a tertiary-type programme, who go on to eventually graduate (OECD, 2007).

[31] 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, one would not expect any average wage effect of the alphabet after controlling for educational attainment.

age groups in the labor force.

To provide tantalizing evidence on this question, we use retrospective survey data collected from over 3 thousands Czech households in December 1996. The data is unique in that it reports name initials of surveyed individuals.[32] 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 misreporting.[33] 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.[34] The 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. The mean log CZK (Czech Crown) monthly wage rate is 8.86.

Our approach 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 (normalized to be between 0 and 1), and (ii) the fraction of the population with first or last names sorted higher in the alphabet for each worker.

---

[32] The sample of 3,157 households is representative of the 1996 Czech population. 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.

[33] 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 household survey.

[34] 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.

The alphabet-ability sorting hypothesis presented formally in Section 3 implies a positive "effect" of the alphabet on ability (and hence wages) 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 academic secondary programs combined with 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 wage observations, while columns (2) and (3) divide the sample based on education type.

Although 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.[35] The parameter estimate implies a 5.2% wage increase associated with "moving" from 'A' to 'Z' among workers with lower education, which is an effect almost identical to the benefit of a year of education estimated on the whole sample of 1,852 workers.[36] Again, the first-name-initial coefficients never reach even marginal levels of statistical significance.[37] The wage sample corresponding to graduates from selective programs

[35] 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.

[36] The finding in Section 4 of no alphabet-ability sorting effects in less selective secondary programs with 'Maturita' exam does not contrast with the results of the wage analysis. The group of workers from less selective schools who report wages in the household survey consists primarily of apprentices who graduated from programs not offering the school-leaving 'Maturita' exam—that is graduates of the least selective secondary programs. The finding of the presence of alphabet-ability sorting among graduates of the most selective schools (academic high schools and colleges) combined with the finding of little such sorting in (technical) secondary programs with average selectivity levels would imply the presence of such sorting in the least selective programs, i.e., for apprentices without the 'Maturita' exam.

[37] 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

may be too small, at 285 observations, to detect alphabetical sorting effect. Overall, we find our wage-structure estimates to be somewhat supportive of the ability-alphabet sorting prediction.

8 Conclusions

While economists have explored the labor-market effect of racial attributes of first names (Bertrand and Mullainathan, 2004; Fryer and Levitt, 2004), studied the socio-economic impact of uncommon surnames (Collado et al., 2007; Güell et al., 2007), and asked about the incidence of women 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

significance. The sample of employed single women was too small (at 340) to allow for effective estimation.

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 can be optimal, but the use of a fixed "lottery ticket" (one's last name initial) throughout many lotteries (many schooling levels) is not fair and may even have efficiency consequences, as we illustrate using a calibrated simulation based on our school-student matching model. A simple remedy is to assign each application a numerical code at random and base sorting on this alternative lottery.

Our results motivate future research into the use of alphabetical listings in access to public schooling or housing. For start, selective education programs are a feature of many European countries other than the Czech Republic. The complete absence of public discussion of alphabetical sorting in school admissions in the Czech Republic may suggest that similar effects could be present and undetected in other competitions or countries.

Bibliography

Aghion, Philippe, Dewatripont, Mathias, Hoxby, Caroline, Mas-Colell, Andreu, and André Sapir (2008) "Higher aspirations: An agenda for reforming European universities," Bruegel Blueprint Series, Vol. V.

Bertrand, Marianne, and Sendhil 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.

Collado, M. Dolores, Ortín, Ignacio Ortuño, and Andrés Romeu (2007) "Surnames and social status in Spain," mimeo, Universidad de Alicante.

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

Einav, Liran, and Leeat Yariv (2006) "What's in a Surname? The Effects of Surname Initials on Academic Success" *Journal of Economic Perspectives* Vol. 20, No. 1: 175-188.

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

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

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

Güell, Maia, Rodríguez Mora, José V., and Chris Telmer (2007) "Intergenerational Mobility and the Informative Content of Surnames," mimeo, Universitat Pompeu Fabra.

Jacobs, Bas, and Frederick van der Ploeg (2006) "Guide to Reform of Higher Education: A European Perspective" *Economic Policy* 21 (47), 535 - 592.

Jurajda, Štěpán (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.

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

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

OECD (2007) *Education at a Glance*, OECD, Paris.

UIV - Ústav pro informace ve vzdělávání [Institute for Information on Education] (1996) *Jaké jsou naše střední školy* [Census of secondary schools], Prague.

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

van Praag, Mirjam, and B.M.S. van Praag (in press) "First Author Determinants and the Benefits of Being Professor A (and not Z): An Empirical Analysis of Non-Alphabetic Name Ordering among Economics Authors" *Economica*.

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

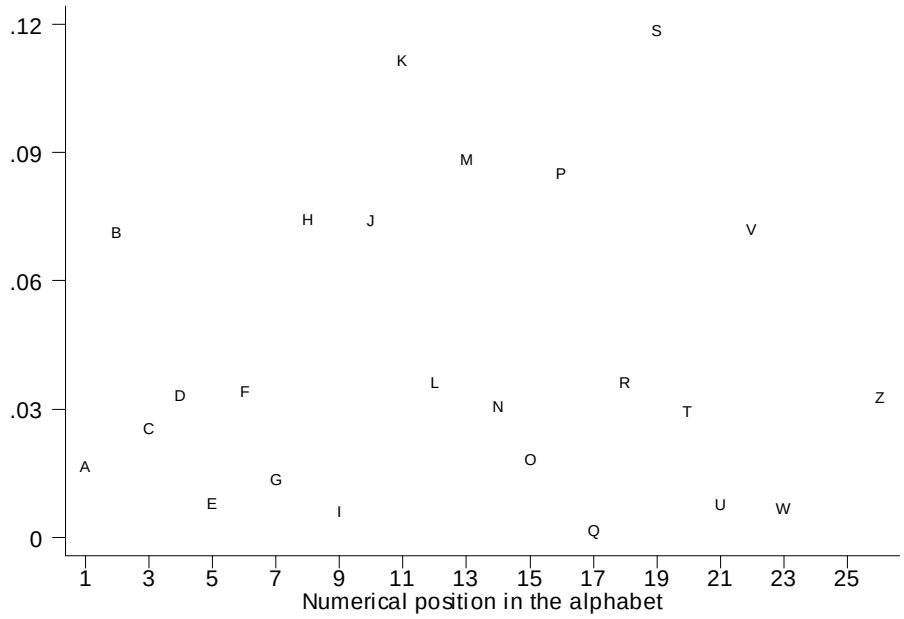

Figure 1: Distribution of Last Name Initials

Table 1: Mean Test Scores and Excess Demand by Secondary 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
Share of graduates applying to university	0.91	0.42	0.12
Share of graduates admitted to university	0.61	0.17	0.04
N	19,174	48,594	22,829
Admission rate of secondary program	0.63	0.71	0.83
	(0.10)	(0.10)	(0.11)

Note: Standard deviations in parentheses.

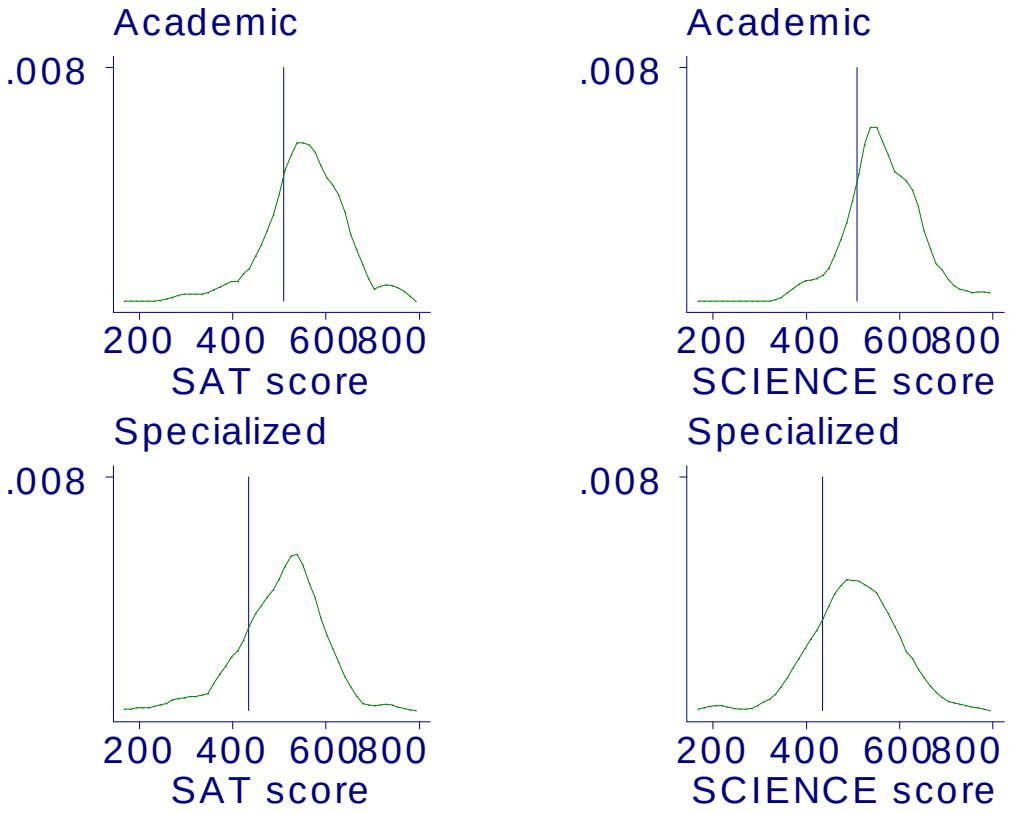

Figure 2: Distribution of 2004 PISA Test Scores for Students Applying to Secondary Schools with Country-Wide Margin of Admission as Vertical Line

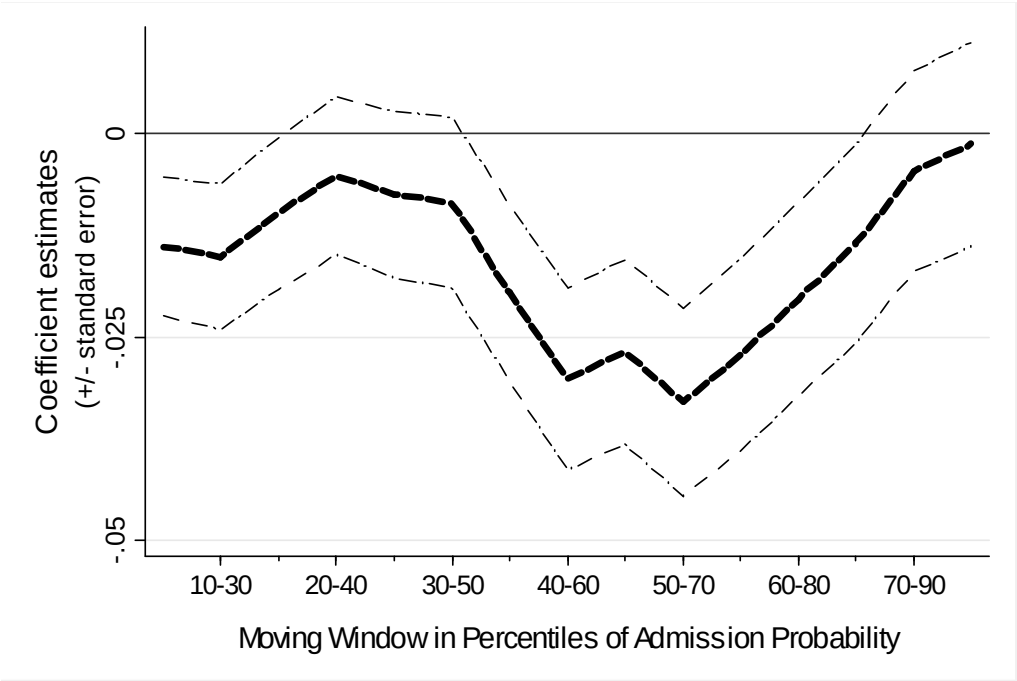

Figure 3: Last-Name Initial Coefficients Across Predicted Admission Distribution

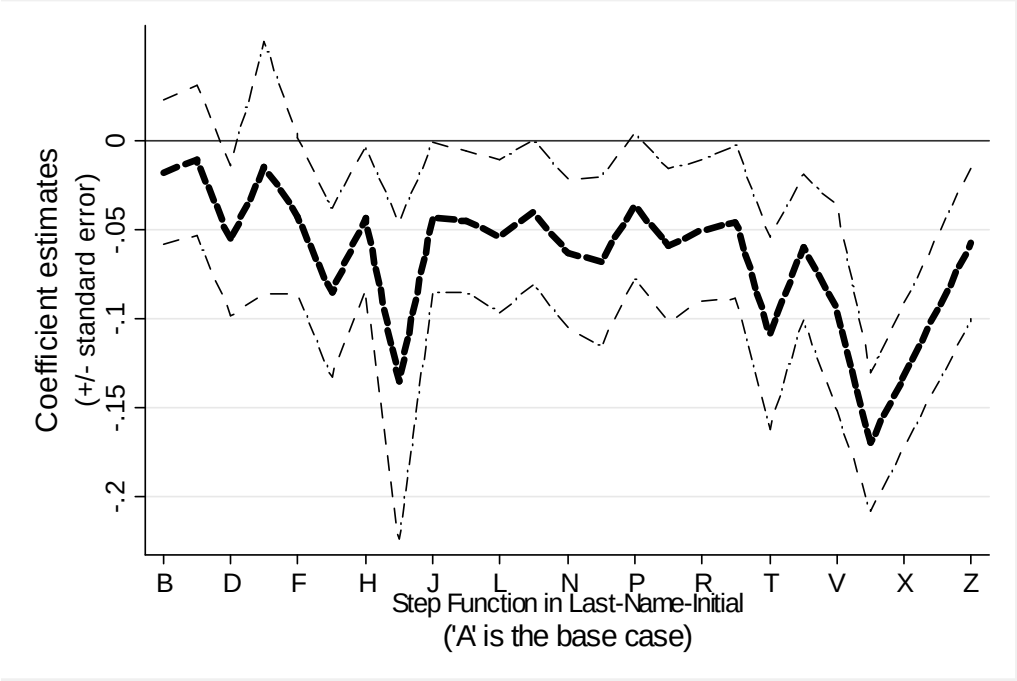

Figure 4: Step-Function Specification for the 40-60 Percentile Range

Table 2: 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.

Table 3: 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 2.

Table 4: 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 Population-Distribution Order				
Last Initial * Admission Rate	**-3.562**	**-2.044**	-4.352	-2.382
	(0.667)	(1.004)	(3.086)	(2.240)
Last Initial	**3.133**	**1.855**	**4.711**	2.339
	(0.618)	(0.089)	(2.035)	(1.610)
First Initial * Admission Rate	-1.268	1.035	-2.931	-1.801
	(1.481)	(1.221)	(2.911)	(2.751)
First Initial	1.032	-1.302	1.788	0.967
	(1.138)	(0.992)	(1.824)	(2.335)
Admission Rate	**-0.998**	**-5.092**	**-7.379**	**-5.644**
	(0.401)	(1.044)	(2.607)	(2.338)
N	91,599	91,599	19,174	19,174

Note: See notes to Table 2. The secondary-school admission rate measures school selectivity.

Table 5: 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	-0.008	-0.001	**-0.038**	-0.011
	(0.008)	(0.006)	(0.011)	(0.012)
First Initial	0.003	-0.011	-0.001	0.001
	(0.008)	(0.008)	(0.015)	(0.015)
Alphabet Position Measure Based on Faculty-Specific-Distribution Order				
Last Initial	-0.007	-0.002	**-0.030**	-0.009
	(0.007)	(0.007)	(0.008)	(0.011)
First Initial	-0.005	-0.011	-0.001	-0.010
	(0.012)	(0.007)	(0.014)	(0.012)
N	89,443	35,411	17,566	36,502

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 faculty-specific admission-probability distribution, which is based on student and school characteristics. Each regression additionally controls for a step function in faculty-specific admission rate (oversubscription).

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.