Substitution Effects in College Admissions

Mikkel Høst Gandil*

November 2021

Newest version here

Abstract

I show how local supply changes create ripple effects in a national educational market. Admitting an applicant to a program will free up a slot to be filled at her next-best alternative. To investigate such substitution effects I re-engineer the centralized admission system of the Danish tertiary education sector and simulate thousands of equilibria under counterfactual supply. I estimate potential earnings with a regression discontinuity design and quantify market-clearings. On average, a change of 10 slots leads to 15 applicants moving and substitution effects explain 40 percent of the variation in earnings. Substitution externalities are generally positive but vary in sign and magnitude. Using linear programming to compute unconstrained optima of the social planner I find that 8 percent of aggregate earnings are forgone by respecting the preferences of applicants and programs.

JEL Codes: C63,I21,I22,I24, I26, J24, H52
Keywords: Field of study, College admission, Program evaluation, RDD

*Department of Economics, University of Oslo. Postboks 1095 Blindern 0317, Oslo, Norway. Telephone: +45 28736769. Mail: mikkel.gandil@econ.uio.no. The Norwegian Research Council funded this research under project no. 275906. I am grateful to Edwin Leuven, Bjørn Meyer, Anders Humlum, Sturla Løkken, Monique de Haan, Manudeep Bhuller, Paolo Giovanni Piacquadio, Lasse Jensen, Morten Nielsen, Simon Damkjær as well as seminar participants for helpful comments and suggestions.

1 Introduction

When educational programs are oversubscribed, admitting an applicant to one program will free up a slot at her next-best alternative which can, in turn, be filled by an applicant from yet another program. Through these flows of applicants, supply changes in an educational program impose externalities on other programs. While central for policy evaluation, these substitution effects are empirically almost completely unexplored.

This paper intends to bridge the gap between estimated returns to education and supply-side policy evaluation in a market with applicant substitution effects. With a centralized and strategy-proof allocation mechanism, it is possible to clear educational markets for counterfactual supply while holding demand (ie. applications) constant. In other words, it is possible to investigate a full counterfactual market equilibrium in terms of individual allocations and outcomes. In this paper, I reverse-engineer the allocation mechanism for the entire Danish tertiary education sector and simulate thousands of supply changes while tracking all resulting applicant flows between programs. I quantify the substitution effects in terms of earnings exploiting the regression discontinuity design (RDD) embedded within the allocation mechanism. I combine the discontinuity design with a parametric selection model which allows me to investigate the effect of the supply changes on the full distribution of predicted earnings among all applicants in the market. To my knowledge, this is the first attempt to quantify full counterfactual market equilibria in terms of individual allocations and earnings in a national market for higher education.

I show that, on average, a change in supply of 10 slots induces 15 applicants to move – an externality of 50 percent in terms of applicant flows. However, the size of the externality varies widely between programs even within the same field. Using predicted earnings I show that by ignoring substitution effects, traditional program evaluations would miss 40 percent of the variation in earnings across experiments. While the sign and magnitude of the externalities vary between programs, I find a positive externality for 70 percent of programs. In other words, if a first-order local average treatment effect (LATE) is positive, ignoring substitution effects most likely leads to an underestimation of the aggregate gain from increased supply. However, I find that similar programs in terms of educational content and estimated LATEs vary in total gains from expanding supply once substitution is taken into account. Higher cutoffs in terms of grade-point averages (GPA) are generally predictive of larger absolute substitution externalities. Lastly, I find that respecting the preferences of applicants and programs incurs an efficiency loss of around 8 percent in terms of aggregate earnings.

This paper shows that when a policymaker implements targeted supply-side policies, applicant sorting and rationing have important effects on aggregate outcomes. This implies that policy-makers must take a broad perspective when deciding on how to ration education. The results are developed by exploiting the centralized nature of college admission in Denmark. However, the qualitative insight that substitution effects matter is not confined to centralized markets nor to markets for higher education. Rather, it is a specific example of how chains of substitution may counteract or enhance the effects of targeted policies. Thus one can expect similar effects in other markets both across countries and educational sectors and in markets for other goods such as housing or labor.

I begin by introducing a framework of centralized allocation for large economies inspired by Azevedo and Leshno (2016). In this framework, the educational market is cleared by program cutoffs rather than prices, and substitution effects following a program expansion can be represented by movement in cutoffs akin to a pecuniary externality. I show how relevant policy effects can be formulated in terms of multiple Local Average Treatment Effects (LATE, Imbens and Angrist (1994)) retrieved from a Regression Discontinuity Design (RDD) and changes in cutoffs. As a LATE is only informative of marginal supply changes, I propose an alternative approach where parametric outcome equations combined with simulations allow for a broader analysis going beyond marginal changes in supply while keeping track of all applicant flows. With the framework in place, I estimate selection adjusted program-level earnings equations using comprehensive Danish register data on applicants spanning three decades. I show that the structural selection models can replicate the corresponding Wald estimates of LATEs lending credibility to the estimated parameters of the potential outcome equations. To improve the out-of-sample prediction of earnings I shrink estimates using Empirical Bayes.

On a separate sample of all applicants to Danish higher education in 2016, I predict all counterfactual outcomes for all applications (applicant $\times$ program combinations). Using this sample, I conduct over four thousand simulations where, in each simulation, I modify the number of slots in a single program and let the model clear the market. I describe the results in terms of applicant flows and in terms of earnings where I perform a variance decomposition and compute estimates of marginal effects of programs expansions in terms of first-order gains (ie. on the margin of the manipulated program) and higher-order gains (ie. on margins between other programs). I find that the first-order gain per added slot is approximately linear in the magnitude of the supply change. This implies that compositional changes of complier groups for non-marginal supply changes are not first-order concerns in this context and that LATEs are externally valid for larger supply changes. However, gains from higher-order movements are less linear and to a limited extent covary with the gains from first-order moves. I then characterize the program-level joint distribution of marginal returns on first-order and higher-order margins and the association between externalities and publicly-observed cutoffs.

Lastly, I quantify the efficiency loss incurred by respecting the preferences of applicants and programs and the assignment algorithm. In a simplified assignment mechanism, I construct four supply-side scenarios where all programs may be changed simultaneously. To quantify the efficiency loss I compare allocations from the assignment mechanism to a counterfactual case where a social planner can freely allocate applicants while respecting supply constraints. Using linear programming I calculate the optima of the social planner and find that around 8 percent of aggregate earnings are foregone by respecting the preferences and the assignment algorithm across the four scenarios. Additionally, I find tentative evidence that eliminating rationing of education entirely would lead to an increase in earnings between 2 and 3.8 percent compared to the baseline scenario. This increase comes with no cost to equality. While this computation is done in a simplified framework with strong restrictions on general equilibrium effects, the resulting efficiency gain indicates that rationing of education might be sub-optimal from the perspective of a social planner.

This paper is related to several different branches of economic research. Firstly, there is a large body of research focusing on the returns to tertiary education. Due to the nature of the educational systems, studies on American data typically focus on returns to institutions (Dale and Krueger, 2002; Hastings, Neilson, and Zimmerman, 2013; Zimmerman, 2014; Arcidiacono, Aucejo, and Hotz, 2016; Andrews, Li, and Lovenheim, 2016; Mountjoy and Hickman, 2020), Several European and South-American papers investigate returns to specific fields, see Altonji, Arcidiacono, and Maurel (2016) for a review. Among these, a number of studies employ regression discontinuity designs (RDD) to estimate returns to fields of study, either in terms of admission (Hastings, Neilson, and Zimmerman, 2013) or completion Kirkeboen, Leuven, and Mogstad (2016) of the marginal applicant of a given program. Several papers have used such approaches to estimate returns to admittance or completion on Danish admission data (Heinesen and Hvid, 2019; Humlum and Meyer, 2020; Daly, Jensen, and Le Maire, 2020; Andersen, Hørlück, and Sørensen, 2020).

A second related body of research finds its origin in the work on school choice by Gale and Shapley (1962) and Abdulkadiroğlu and Sönmez (2003) among others. Azevedo and Leshno (2016) develop large market asymptotics of stable matching mechanisms and show that they can be represented by in a framework where cutoff structure plays the role of prices in clearing demand and supply. Abdulkadiroğlu, Angrist, Narita, and Pathak (2017) use the same large market framework to construct a research design for the estimation of causal effects exploiting that random tie-breaking moves applicants between programs. Abdulkadiroglu, Angrist, Narita, and Pathak (2022) expand the framework to encompass non-random tie-breaking in a regression discontinuity framework. These articles exploit applicant switching in centralized mechanisms to construct research designs but do not investigate the role of applicant switching for policy evaluation.

Several papers exploit the large market properties to estimate welfare in mechanisms where truth-telling is not a dominant strategy. Abdulkadiroğlu, Agarwal, and Pathak (2017); Agarwal and Somaini (2018); Fack, Grenet, and He (2019); Kapor, Neilson, and Zimmerman (2020). Contrary to these papers I take preferences as truthful and policy-invariant, thereby holding demand constant while changing the supply. Rather than assuming a specific structure on applicant utility, my analysis of outcomes will be in terms of earnings, which is uncontroversial in the social welfare literature (Sandmo, 2015).

A few papers investigate applicant flows explicitly. Agarwal (2015, 2017) investigates the effects of changing capacities on allocations in the centralized medical residency matching market in the US and Bucarey (2018) investigates the crowding out of applicants in Chile when introducing financial aid. My paper differs from theirs in three significant ways. Firstly, the Danish system is fully centralized and I have access to all inputs as well as the algorithm. This means that simulations reflect realistic counterfactual scenarios of an entire educational market. Secondly, while Agarwal and Bucarey estimate preferences, I avoid this structural modeling by observing rank-ordered lists from a robust mechanism. Finally, as I estimate earnings equations, I link substitution effects directly to traditional treatment effects type estimates of returns to fields and thereby price the externalities which are not commonly captured in program evaluations.[1]

The importance of substitution effects for policy interventions has been investigated by Manning and Petrongolo (2017) in the context of local labor markets where markets overlap geographically. The authors document that geographic ripple effects dilute the effect of local stimulus policies. In Danish higher education, I show that treatment effects are on average not attenuated but rather increased by the presence of substitution effects. In their working paper Kirkeboen, Leuven, and Mogstad (2016) briefly investigate the indirect effects of increasing capacity in science programs using first-stage estimates of complier compositions from their fuzzy regression discontinuity framework and IV estimates of returns contingent on the next-best alternative. While the Scandinavian centralized setting is similar to this paper, my paper adds to their analysis in significant ways. Firstly, they consider one round of indirect effects at the field level and ignore longer chains of substitution. In contrast, I simulate the entire mechanism and follow each applicant potentially through their entire rank-ordered list of programs including the possibility of non-assignment, which I show is an important margin. Additionally, I allow for heterogeneity within fields across applicants by parametrizing outcomes. I show that the program level focus is important as even within fields there are large differences in substitution effects that are missed by aggregation. Kline and Walters (2016) explicitly investigate the role of substitution for cost-benefit analysis in the context of the Head Start program in the US and show that accounting for the cost of the non-admitted compliers in other public programs can significantly alter the cost-benefit ratio of an expansion.[2] They briefly investigate the importance of rationed substitutes and conjecture that not accounting for rationing provides a lower bound on the rate of return to program expansion. In the context of Danish higher education, I show that this does not always hold. In estimating and predicting counterfactual outcomes, I draw inspiration from Abdulkadiroğlu, Pathak, Schellenberg, and Walters (2020) where the authors investigate the role of value-added for high school choice in New York City.

2 Substitution effects in centralized mechanisms

To clarify which estimates are important for policy evaluation I employ a stylized model of a centralized matching market. Similarly to Azevedo and Leshno (2016) I assume a continuum economy where a finite set of programs, P, indexed by p, is matched to a continuum of applicants. The capacity of a program is given as S_p. Applicants are defined by their type, $\theta = (\succ^{\theta}, e^{\theta}, Y_{\theta}, X_{\theta})$. where $\succ^{\theta}$ is a strict preference ordering of programs and $e^{\theta} \in (0,1)^P$ is a vector of eligibility score, where programs prefer applicants with higher eligibility scores. $Y_{\theta} = \{Y_p^{\theta}\}$ is a vector of length P of potential outcomes.Finally, X_{θ} is a set of observable characteristics. Let η be the probability measure over the set of all types.

[1]Additionally, Tanaka, Narita, and Moriguchi (2020) provide reduced-form evidence that crowding out occurred when introducing centralized admission to education in Japan.

[2]Feller, Grindal, Miratrix, Page, et al. (2016) in the same context find differential effects depending on the counterfactual allocation.

Azevedo and Leshno (2016) show that a stable matching in a continuum economy is unique and can be characterized by a vector of program cutoffs, $C \in (0,1)^P$ which equate demand and supply. The demand of applicant type θ, $D^\theta(C)$ is defined as her favorite program among those where the eligibility score of e_p^θ exceeds the cutoff. The cutoffs can in turn be thought of as prices clearing supply and demand.In the Deferred Acceptance mechanism (DA) it is the dominant strategy for applicants to report true preferences regardless of other applicants or supply (Abdulkadiroğlu and Sönmez, 2003). Assuming no application costs, no restriction on the rank-ordered lists and rational applicants, the reported preferences of applicants are, therefore, not a function of supply, and the market can be fully cleared for any given supply holding submitted applications constant.[3]

To quantify the effects of supply changes on outcomes, assume that program p performs a marginal expansion of capacity of dS_p. Given that p is oversubscribed, the change in the number of admitted applicants to p equals the change in capacity. The vector C' contains the cutoffs in the stable matching using the new vector of supplies . Equating supply and demand it follows that:

$$\begin{aligned} dS_p = dD(C_p) =& \eta(\{\theta : D^\theta(C') = p, D^\theta(C) \neq p\}) \\ =& \int_\theta \mathbf{1}\{\theta : D^\theta(C') = p, D^\theta(C) \neq p\} d\eta(\theta) \\ =& \sum_{p' \neq p} \int_\theta \mathbf{1}\{\theta : D^\theta(C') = p, D^\theta(C) = p'\} d\eta(\theta). \end{aligned} \tag{1}$$

This formalization shows that marginal applicants come from other programs and that these applicants are heterogeneous in type. By moving into program p the applicants will realize their potential outcome associated with p, Y_p^θ. The aggregate change in outcomes for this set of applicants, therefore, is given by:

$$\Delta_p = \sum_{p' \neq p} \int_\theta \left[Y_p^\theta - Y_{p'}^\theta \right] \mathbf{1}\{\theta : D^\theta(C') = p, D^\theta(C) = p'\} d\eta(\theta) \tag{2}$$

Dividing Equation (2) with (1) gives the average gain in outcomes of the marginal applicants associated with a marginal increase in program p, which is equivalent to the Local Average Treatment Effect (LATE, Imbens and Angrist (1994)):

$$LATE_p = \frac{\sum_{p' \neq p} \int_\theta \left[Y_p^\theta - Y_{p'}^\theta \right] \mathbf{1}\{\theta : D^\theta(C') = p, D^\theta(C) = p'\} d\eta(\theta)}{\sum_{p' \neq p} \int_\theta \mathbf{1}\{\theta : D^\theta(C') = p, D^\theta(C) = p'\} d\eta(\theta)} \tag{3}$$

Equation (3) shows that the IV estimator will readily identify the treatment effect on compliers without observing types of applicants.[4] The irrelevance of knowledge of the margin of choice

[3] I maintain these assumptions throughout the paper and discuss the implications in Section 9.

[4] In other words, even without full rank-ordered lists and knowledge of alternatives, the IV estimate identifies the effect of admission on the margin. To see this, note that the complier share, as defined in the LATE framework in this context, can be obtained by summing over alternatives and integrating over types: $Pr\left(D_i(C') = p, D_i(C) \neq p\right) = \sum_{p' \neq p} Pr(D_i(C') = p, D_i(C) = p') = \sum_{p' \neq p} \int_\theta \mathbf{1}\{\theta : D^\theta(C') = p, D^\theta(C) = p'\} d\eta(\theta)$.

is a specific formulation for the centralized DA mechanism of the general identification results presented by Heckman, Urzua, and Vytlacil (2008). This also formalizes the reasoning of Kirkeboen, Leuven, and Mogstad (2016) that the LATE of admission to program p is identified in a fuzzy regression discontinuity design using threshold crossing as an instrument and controlling for the position in the waiting list in a DA-mechanism.[5]

With a single-instrument IV, the LATE can be estimated under three assumptions; Independence of the instrument, an exclusion restriction, and a monotonicity condition. Infinitely close to the program cutoff, an offer is random, ensuring that independence of potential outcomes is satisfied. The exclusion restriction is plausible as the instrument reflects a decision rule in the algorithm – a program offer is the only plausible way that it might affect outcomes. Following Imbens and Angrist (1994), applicants can be separated into four sets; never-takers, always-takers, defiers, and compliers. Never-takers are those who would not be admitted regardless of crossing the cutoff or not and consists of two subsets: Those who do not apply to the program and those admitted to higher prioritized programs. The set of always-takers is empty. To see this, recall the definition of the cutoff as the minimum eligibility score of those admitted. Due to the lack of justified envy, a necessary condition for admission is crossing this cutoff. This rules out the existence of always takers.By the same reasoning, the set of defiers is empty; crossing the cutoff can only make applicants (weakly) more likely to accept be accepted. The standard monotonicity condition for a just-identified IV is therefore plausible.

Using Equations (1) and (3) the aggregate effect on the newly admitted to program p can then be formulated in terms of a LATE and a change in supply.

$$\Delta_p = LATE_p \times dS_p = LATE_p \times dD_p(C), \tag{4}$$

which shows that the LATE is the policy-relevant treatment effect (Heckman and Vytlacil, 2001). However, using only (4) to evaluate a supply change policy overlooks that not only does the demand of program p change, but the demands of other programs change as well. The link between changed supply in one program and changes in the entire vector of cutoffs reflects the substitution of applicants and is equivalent to a pecuniary externality in a standard supply and demand framework. To evaluate a policy of program expansion an earnings-maximizing social planner computes the total gain, Δ^T, which is the sum of program-specific gains from a change in program p:

$$\Delta^T = \underbrace{LATE_p \times dD_p(C)}_{\text{First-order}} + \underbrace{\sum_{p' \neq p} LATE_{p'} \times dD_{p'}(C)}_{\text{Higher-order}}. \tag{5}$$

Equation (5) shows that the informational requirement for policy evaluation using the LATE framework is substantial. The social planner needs knowledge of all changes in cutoffs and

[5]Kirkeboen, Leuven, and Mogstad (2016) argue that one needs to condition on the next-best program to obtain interpretable LATEs. This is only the case if one wants to uncover heterogeneity in returns. A standard fuzzy RDD is sufficient to uncover the LATE of interest in the context of marginal supply changes, as it implicitly weights compliers on the margin.

estimated LATEs for all programs where $dC_{p'}/dS_p \neq 0$ making. The large informational requirement is not a specific feature for centralized mechanisms as substitutes are often rationed (see Kline and Walters (2016) for an example using Head Start in the US). However, contrary to common applications of the LATE framework, in the context of centralized admissions, the individual compliers are directly identifiable in the production data through simulation. Therefore, centralized admission systems allow researchers to explicitly investigate the importance of such substitution patterns without restrictive assumptions on applicant behavior or market-clearing.

Alternative to LATE for policy evaluation when compliers are known As outlined above, a proper policy evaluation using IV methods requires multiple LATE estimates. This has several drawbacks; Firstly, while IV is consistent, estimates can be extremely noisy (Young, 2019). Secondly, using cutoff-crossing as an instrument is only possible in oversubscribed programs. Thirdly, while LATEs might be accurately estimated they are not readily informative on issues of social welfare such as economic inequality. Finally, policy evaluations using LATEs are inherently local and may not be robust to large policy changes where the composition of the complier group changes substantially.

To handle these issues I assume a parametric structure on the potential outcomes and utilize application and capacity data and knowledge of the assignment algorithm to identify complier groups of policy experiments directly. I assume that the outcomes can take only positive values and assume the following parametric structure:

$$Y_p^\theta = \exp\left(X_\theta \beta_p\right), \tag{6}$$

where X_θ is a vector of observable characteristics of type θ and β_p are program-specific parameters that map the characteristics to outcomes. The aggregated effects can then be written as

$$\Delta^T = \sum_p \sum_{p' \neq p} \int_\theta \left[\exp\left(X_\theta \beta_p\right) - \exp\left(X_\theta \beta_{p'}\right)\right] \mathbf{1}\{\theta : D^\theta(C') = p, D^\theta(C') = p'\} d\eta(\theta) \tag{7}$$

The parametric structure is restrictive in that it selects the features that are relevant for predicting outcomes. Further, the program-specific parameters must be estimated. The gain, however, is that given estimates of β_p, I can implement non-marginal supply changes and calculate expected changes in outcomes as well as outcome levels across the whole distribution of applicants.[6] The complier compositions are computed directly by simulating the assignment mechanism and given estimates of β_p I can calculate all the empirical equivalents of the elements of Equation (7).

[6]Note, however, that I assume β_p to be constant under counterfactual supply. I thereby assume away general equilibrium effects that affect payoffs. I discuss this assumption in the conclusion.

3 Context and simulation of the assignment mechanism

In Denmark, tertiary programs are generally divided between short-cycle 2-year programs, medium-cycle professional bachelors (such as teaching), and long-cycle academic bachelors at universities (where most graduates will proceed to complete a master's degree). Education is free and students receive generous grants and loan opportunities to cover living expenses. In line with most continental European systems, applicants apply to specific program-institution combinations and enrollments mainly occur in the fall semester.In general, programs set their own capacities though some programs with high unemployment rates of graduates are prohibited from expanding capacity. Programs receive funds from the central government based on completed coursework and graduation rates.

Allocation to tertiary education is administered by the Ministry of Higher Education and Science in a centralized allocation system (in Danish: *Den Koordinerede Tilmelding*). Each year the allocation round matches between 70 and 90 thousand applicants to around 800 programs. When applicants outnumbers slots in a program, applicants are generally admitted in two quotas. Quota 1 (abbreviated Q1) admits applicants according to the grade-point average (GPA) from secondary school which is a combination of nationally centralized exams and grades given during the term. An applicant will enter with the same GPA in the rankings in multiple programs and only applicants with a GPA or foreign equivalent can be admitted in Q1. The GPA is calculated with one decimal and a lottery number is used as a tie-breaker. An alternative is Quota 2 (abbreviated Q2), in which the ranking criteria are chosen by the educational institution under constraints set by the ministry. The most popular approaches in Q2 are combinations of specific course grades and CV requirements, though there is a lot of variation in these criteria. The ranking process is performed by the program admission offices and the ministry only observes the final ranking in Q2.

Each applicant can apply to eight programs. Under each program, the applicant can signal whether they want to be evaluated in Q2. If so, the applicant will have to supply further information to the program in order to be ranked. Additionally, the applicant can signal that they want to be evaluated for a "standby" slot both in Q1 and Q2. This system with quotas and standby means that while the applicant can only rank eight programs, the system can observe an applicant with more priorities.[7] In the mechanism, along with this modified rank-ordered list, an eligibility score is observed for each applicant in each relevant quota.

Program admission offices observe the program-specific application and not the remaining programs on the rank-ordered list of the applicant. This makes it very difficult for the admission offices to act strategically. Prior to allocation, the institutions report a capacity for each quota. After the ministry runs the mechanism, programs are allowed to increase but not decrease their capacity and the mechanism is run again. My data is from the final run of the mechanism. The allocation mechanism is a modified Student Proposing Deferred Acceptance (abbreviation: DA,

[7]For example, a single application from the perspective of the applicant can entail 4 applications from the perspective of the mechanism: Q1, Q2, Q1-standby, and Q2-standby. If admitted to a standby-slot, and an admitted applicant rejects an offer, the offer is given to an applicant in the standby-quota according to the ranking. If not admitted, a standby slot guarantees an offer in the next academic year. Some programs also provide the option of enrolling in the Spring semester, which will sometimes be represented as an additional program in the mechanism.

Gale and Shapley (1962); Abdulkadiroğlu and Sönmez (2003)). In principle the DA is strategy-proof. However, in practice, specific features create strategic incentives. For instance, the limit on the number of programs on the rank-ordered list may force applicants to truncate their lists. However, only 3 percent of the applicants submit a full list, implying a low level of truncation at the bottom of the list.[8] Each year the results of the allocation mechanism are made available on the Ministry website. This information includes the number of students allocated to each quota and remaining slots. Admission cutoffs in Q1 in terms of GPA are published and treated as front-page news in the media. Cutoffs in Q2 are not informative as programs vary in their ranking function which is not generally known – nor in most cases formalized.

Simulating the assignment mechanism With knowledge of the algorithm and production data, I construct a version of the entire mechanism for the allocation of tertiary education in Denmark. I manage to allocate over 98 percent of applicants correctly in numerous years. I can, therefore, keep track of all substitution patterns when changing features of the market.

To illustrate the simulation model, Figure 1 presents the result of changing the capacity of the Medicine program at the University of Copenhagen (KU) in the 2016 admission round. The program is heavily oversubscribed and has one of the highest Quota 1 cutoffs. Figure 1a shows the flows resulting from a decrease in the number of slots in Quota 1 of 10 percent. The y-axis shows which combination of field and program length applicants come from, while the x-axis shows where they end up. The dashed line indicates the outflows from the long-cycle Medicine group, including the program at University of Copenhagen. The darker the square, the larger the flow.

The largest flow is *between* Medicine programs. In other words, pushing applicants out of Medicine at KU pushes them into other Medicine programs. The second-largest flow is on the extensive margin, pushing applicants out of admission entirely. This partly reflects that some applicants only apply for a single program (and may apply again the following year.). Figure 1a also shows a large flow into Dentistry, indicating that these fields are substitutes. In turn, applicants are pushed out of Dentistry as evidenced by the flows in the lowest row. The same pattern can be seen for the medium-cycle health programs, which mostly consist of Nursing programs.

Figure 1b shows the result of an *increase* of capacity at the University of Copenhagen. While many flows occur along the same margins, this is not universally true.In other words, substitution patterns are complex and complier shares from a contraction do not equal the complier shares of an expansion.[9] Fortunately, my simulation model keeps track of these flows regardless of the size of supply changes and without imposing structure on applicant preferences. The

[8]As noted by Calsamiglia, Haeringer, and Klijn (2010); Fack, Grenet, and He (2019), applicants might leave out unrealistic programs at the top of the rank-ordered list. In the conclusion, I discuss the issues this kind of truncation might pose for the validity of my findings. Further, the decision to apply for standby is non-trivial, as it does not enter as a separate application. In order to not waste slots, the algorithm is nested within a loop where non-filled slots are transferred between quotas between each iteration of the algorithm. This means that the algorithm does not necessarily terminate, though the resulting matching is stable due to the properties of DA.

[9]For marginal experiments, one could in principle use the first stages from IV regressions where the instrument is interacted with the next best field as approximating the flow shares as done in the working-paper version of Kirkeboen, Leuven, and Mogstad (2016). However, with discretionary slot allocation and varying program sizes, the definition of a marginal expansion is unclear in practice and first-stages are likely to be noisy.

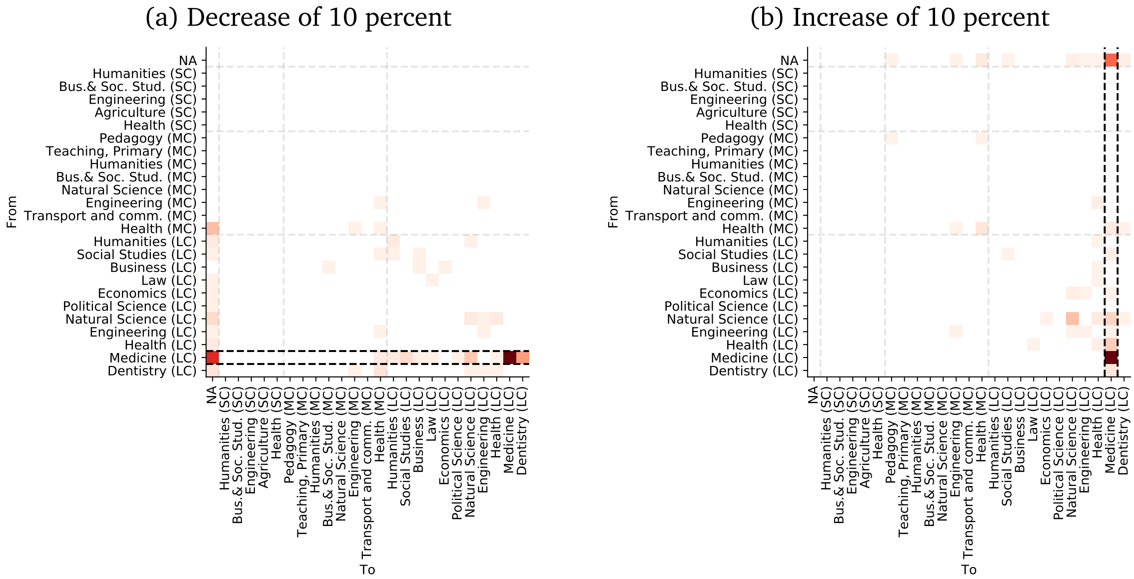


Figure 1: Applicant flows following a change in capacity for Medicine at KU

The simulations are based on a re-engineered version of the Danish assignment mechanism using 2016-data. In Figure 1a the number of slots in Quota 1 for Medicine at the University of Copenhagen is decreased by 10 percent. In Figure 1a the capacity is increased by 10 percent. Programs are grouped into field-length combinations, where the length can be short-cycle (SC, 1 to 3 years), medium-cycle (MC,3 to 4 years), and long cycle academic programs (LC, 3 years or more, academic bachelor programs). The dashed lines indicate the first order margin by field×length.

flows are completely determined by the applicant data and accurately reflect what would have occurred had the supply been changed while holding demand constant.[10]

4 Data and descriptive statistics

Population data on income, grade-point averages (GPA), initiated educational spells and socio-economic variables are obtained from Statistics Denmark. Raw production data from the admission system is provided by the Danish Ministry of Higher Education. This data contains waiting lists for each program-quota combination, priorities of applicants and capacity, and the admittance outcome. Data is available from 1992, but due to lack of data simulation with a satisfactory level of precision is possible from 2016.[11]

Estimation sample For estimation of outcomes, I take all applications from applicants in the admission data for whom I observe GPA, a Quota 1 waiting-list number, and positive income in the registers. The income concept is the log of average personal pre-tax income excluding transfers 7 to 9 years after application. With the exception of Medicine, no program is longer than five years, Thus the outcome measure should be interpreted as early career earnings. The

[10]These simulations are performed for the 2016 admission round, which is the same round that I will perform simulations on in Section 7.

[11]Each program-year combination has a numeric identifier. Though large programs maintain the same identifier throughout the sampling period, many smaller programs change due to institutional merges and changes in program content. To maintain consistency over time I map the identifier to fields and institutions through the register data. If over 95 percent are enrolled in a given program, I link the two. I match the remaining program-year combinations manually. In the case of institutional mergers, I map the programs to the latest institution observed in the data.

combination of register availability and the log specification entails that foreign applicants who do not migrate to Denmark are not observed in the data. Additionally, individuals with no income in all three years are excluded from the sample. As a measure of skill, I calculate the GPA rank within graduating cohort. For applicants where I do not observe the GPA (mainly foreign applicants), I impute it using a nearest-neighbor regression on the rank in Quota 1. In the control function specification, I exclude programs that do not have support in the running variable on either side of the cutoff. Further, I exclude programs where the admittance rate is less than 2 percent or higher than 98 percent. I do not impose this restriction in models where I do not control for selection.

Simulation sample The sample used for the simulations of capacity changes contains the population of applications in 2016. For some applicants, I lack data on GPA, which is necessary for computing expected outcomes. As all applicants enter this sample, regardless of whether they apply in Quota 1, I cannot impute GPA rank by the same procedure as in the estimation sample. For most applicants, an alternative GPA is reported in the production data. This GPA is including potential bonuses given due to a short time span between graduating high school or due to raising subjects to a higher level.[12] For foreigners, the GPA conversion follows the guidelines of the Ministry of Higher Education, though as this is done by the program administrators these GPAs are subject to error.[13] The corresponding GPA-ranks are imputed by a Random Forrest regression trained on those applicants where both GPA-rank and the alternative GPA are reported. For the subset of applicants where no variant of GPA is observed, I set the value of the GPA rank to 0.5.

4.1 Descriptive statistics

Table 1 present descriptive statistics for the samples. The unit of observation is the applicant, and the discrepancy between the number of applicants and applicants reflects that on average applicants file 1.8 applications (850,290/478,803). The first column shows that the applicant GPA rank in own cohort is slightly above the average in the applicant cohort with a mean value of 0.55. In the estimation sample, 32 percent of applicants apply to more than one program within the same field (and program length), and 50 percent file at least one Quota 2 application. Income 8 years after the application is on average 300 thousand DKK in 2015-prices, roughly equivalent to 45 thousand USD.[14]

The following four columns display the corresponding statistics for the samples used for sampling correction. As I describe in section 5 I estimate four different sets of control functions where I use cutoffs in Quota 1 or Quota2 and exploit cutoffs in own program and higher prioritized programs. The sample size in terms of applications diminishes reflecting that the research design requires binding cutoffs and I restrict attention to applicants close to these cutoffs. The average GPA rank is slightly higher in the Quota 1 samples with values of 0.59 for correction using own program cutoff and 0.56 using higher prioritized cutoffs. The simulation sample

[12]Applicants applying within two years of high school graduation can multiply their GPA with 1.08.
[13]For instance, the same applicant can be observed with multiple values of the alternative GPA across applications.
[14]The average exchange rate in 2015 was 6.72 DKK/USD.

Table 1: Descriptive statistics on samples

		Estimation sample	Program cutoff Q1	Q2	Higher priority cutoff Q1	Q2	Simulation sample
GPA rank	Mean	0.55	0.59	0.49	0.56	0.49	0.51
		(0.28)	(0.26)	(0.25)	(0.26)	(0.25)	(0.27)
Applications in same field	Mean	0.32	0.41	0.46	0.74	0.79	0.43
		-	-	-	-	-	-
Applies in Q2	Mean	0.50	0.51	1.00	0.56	1.00	-
		-	-	-	-	-	-
Income	Mean	300.36	298.46	296.38	297.22	297.28	-
		(145.58)	(147.19)	(140.21)	(146.60)	(139.83)	-
log(Income)	Mean	5.46	5.45	5.46	5.44	5.46	-
		(0.99)	(1.00)	(0.95)	(1.00)	(0.95)	-
N applicants		478,803	302,149	151,244	157,705	75,290	82,794
N applications		850,290	480,752	210,525	255,002	104,924	204,900

The table reports summary statistics. The first five columns report statistics on the samples used for estimating outcome equations while the last column reports the statistics for the simulation sample. Standard deviations are reported in parentheses. Standard deviations are only shown for continuous variables.

contains *all* 83 thousand applicants in 2016. Of these applicants, 43 percent apply to more than one program within the same fields, and on average applicants file 2.5 applications. This ratio is higher than in the estimation sample in the first column. This is due to the lack of sampling restrictions in the simulation sample.

5 Estimating and predicting potential outcomes

I now present the methods used to obtain estimates of the parameters governing the expected potential outcomes. As in Equation (6) I assume that the potential outcomes are linked to programs via the following log-linear model:

$$\log(Y_{ip}) = y_{ip} = X_i\beta_p + \varepsilon_{ip}, \tag{8}$$

where y_{ip} is the log of a positive scalar outcome of interest of individual i admitted in program p and X_i is a vector of observable and predetermined covariates while ε_{ip} is unobserved. The vector of observables, X_i, contains a constant, a second-degree polynomial in the GPA-rank in the applicant's high-school cohort and a linear year trend. If applicants sort on potential gains, conditioning on preferences may illicit heterogeneous returns as shown by Kirkeboen, Leuven, and Mogstad (2016) and Abdulkadiroğlu, Pathak, Schellenberg, and Walters (2020). I therefore also include dummies for whether the lower-ranked alternative is within the same field and whether the application is the last priority on the rank-ordered list of the applicants. The outcome y_{ip} is only observed for applicants admitted to p. Taking the expectation of equation (8) for those admitted yields:

$$E[y_{ip}|X_i, i \in p] = X_i\beta_p + E[\varepsilon_{ip}|X_i, i \in p]. \tag{9}$$

I estimate β under different assumptions on the unobserved component, $E[\varepsilon_{ip}|X_i, i \in p]$.

A set of estimates are obtained under the assumption that the expected value of the unobserved component conditional on covariates is zero, $E[\varepsilon_{ip}|X_i, i \in p] = 0$:

$$E[y_{ip}|X_i, i \in p] = X_i\beta_p \tag{10}$$

Under this conditional independence assumption, an unbiased estimate of β_p can be obtained by Ordinary Least Squares (OLS). For each program, I estimate the potential outcomes of those admitted. For a subset of programs, it is not feasible to estimate program-specific parameters in the estimation sample as these programs are too new for applicants to realize outcomes. For predicting these programs I estimate field-specific versions of equation (10). All models based on conditional independence are estimated with heteroskedasticity-robust standard errors.

If applicants sort into programs based on the unobservable part of the potential outcomes such as in the Roy model (Roy, 1951), OLS estimates of β_p will be biased. To control for selection I exploit the cutoff representation in a Fuzzy Regression Discontinuity Design. As previously described, the Danish admission system works with program-quotas instead of programs. This implies that the cutoff representation is in terms of quotas indexed by q instead of programs, with a many-to-one mapping between quotas and programs.[15] In the data, I observe the ranking of each applicant in each quota. By flexibly controlling for the position in the waiting list of a given quota, admission can be instrumented with a dummy for crossing a cutoff. In this context, such a strategy gives rise to two instruments, using the cutoff in Quota 1 and Quota 2. However, one can also exploit variation in higher ranked alternatives, see Humlum and Meyer (2020). Marginally *not* crossing a cutoff in a higher ranked alternative can serve as an instrument for admission in a program. In this context, this approach adds two additional instruments, one for Quota 1 and one for Quota 2.

As an instrument for admission in program p I define a binary instrument for crossing the cutoff in a quota q in program p, $Z_i^{pq} = \mathbf{1}\{e_{iq} \geq 0\}$, where I center the running variable, e_{iq} around the cutoff in quota q in the year. Conditional on the running variable, e_{iq}, Z_i^{pq} is a valid instrument for D_i^p. I specify the following selection equation:

$$D_i^p = \mathbf{1}\left\{\gamma Z_i^{pq} + f\left(e_{iq}\right) + \epsilon_{iq} > 0\right\}, \tag{11}$$

where $f\left(e_{iq}\right)$ is a linear spline with a knot at zero.

Assuming that $(\varepsilon_{ip}, \epsilon_{iq})$ are jointly normal-distributed, the model in Equation (9) becomes a standard probit selection model:

$$E[y_{ip}|X_i, i \in p] = X_i\beta_p + \psi_p\lambda\left(\hat{\gamma} Z_i^{pq} + \hat{f}\left(e_{iq}\right)\right), \tag{12}$$

where $\lambda(\cdot)$ is the inverse Mills-ratio and $\psi_p = corr(\varepsilon_{ip}, \epsilon_{iq})\sigma_\epsilon$.

This model embeds the identifying variation from the regression discontinuity design in a switching regression framework and imposes structure through equation (9). To include the instrument, Z_i^{pq} in equation (12) and avoid identifying off functional form and the running variable alone, there need to be applicants on both sides of the cutoff. This implies that the con-

[15] Mechanically, each quota is given a numeric identifier and submitted rank-ordered lists are expanded lexicographically by priority and quota identifier.

trol function method is only feasible for over-subscribed programs or in cases where applicants are restricted from admission into higher-prioritized programs.[16]

Due to the multi-quota nature of the admission system and the specification of the instrument, there is two-sided non-compliance; Firstly, applicants may have $Z_i^p = 0$ but be over the threshold in another quota within p. This implies that applicants below the threshold gain admission.[17] Secondly, applicants above the threshold might also be eligible to a higher-ranked program and thus get admitted there. The two-sided non-compliance imply a correspondance between the selection model and the LATE estimate from a fuzzy regression discontinuity design using IV. To validate the model, following Kline and Walters (2019), I calculate the LATE of being admitted to p in this model and compare it to the corresponding IV specification of the fuzzy regression discontinuity design using two-stage least squares. The methods for calculating the LATE in the control function approach and the standard IV approach is outlined in Online Appendix A.1.1. Standard errors are obtained from the two-step variance estimator derived by Heckman (1979).

5.1 Joint distribution of parameters, shrinkage, and prediction

Under appropriate assumptions, the two set of estimates of β_p are unbiased but noisy measures of the true parameters. This noise is especially an issue with the control function method on small programs with correspondingly small samples. As the estimates are needed for predictions, I follow Abdulkadiroğlu, Pathak, Schellenberg, and Walters (2020) and apply an empirical Bayes shrinkage estimator to the estimates, thereby reducing the mean squared error of the predicted outcomes (Robbins, 1992; Morris, 1983).I assume the following hierarchical model:

$$\hat{\beta}_p | \beta_p \sim \mathcal{N}\left(\beta_p, \Omega_p\right) \tag{13}$$

$$\beta_p \sim \mathcal{N}\left(\mu_\beta, \Sigma_\beta\right) \tag{14}$$

Estimates of μ_β and Σ_β are obtained through maximum likelihood estimation of (13) and (14), where Ω_p is approximated by the estimated covariance matrices of the estimation methods described above. For the control function estimates, I include ψ_p in the vector of parameters. The resulting estimates of the hyperparameters, μ_β, and $\hat{\Sigma}_\beta$, are in turn used to obtain empirical Bayes posterior means for β_p:

$$\beta_p^* = \left(\hat{\Omega}_p^{-1} + \hat{\Sigma}_\beta^{-1}\right)^{-1} \left(\hat{\Omega}_p^{-1}\hat{\beta}_p + \hat{\Sigma}_\beta^{-1}\hat{\mu}_\beta\right),$$

where ($\hat{\Omega}_p$ is approximated by the covariance matrix of $\hat{\beta}_p$. In essence, this procedure shrinks imprecise estimates towards the mean of the estimated coefficients. The procedure is described in detail in Appendix A.1.2.

[16]As the assignment function is known, there is no need to include X_i in the selection equation. As the running variable is just a rescaling of GPA (In Quota 1) and a function of the mass of applicants, I exclude the running variable from the structural equation. For each program, I stack application years where the program is oversubscribed.

[17]This contrasts with the theoretical model outlined in Section 2. The theoretical outline did not allow for always-takers because each program had one quota. In the empirical section I observe always-takers because I only model one type of quotas at a time.

The predicted outcomes in the simulation sample are calculated as $\hat{Y}_{ip} = \exp\left(X_i\beta_p^*\right)$. For each program, I also obtain a predicted potential outcome if not admitted. For the last program on the rank-ordered list of an applicant, this predicted outcome is interpreted as the outcome in the case of not being admitted to any program. In addition to improving prediction, the hyperparameters are also informative on the joint distribution of parameters in the educational production function and by comparing the hyperparameters across the estimation methods I can investigate the importance of correcting for selection.

6 Estimation of outcomes

Balancing and non-parametric RD I begin by showing balancing and threshold-crossing effects in the full sample where programs are stacked. Figures 2a and 2b show balancing of predetermined variables and I find no evidence of effects of sorting around the cutoff.[18] The first stage, displayed in Figure 2c, is large at 0.4. In other words, crossing the threshold increases the probability of admittance by 40 percentage points. The level of admittance conditional on crossing the cutoff is around 60 percent. This reflects that applicants will only be admitted to a program if they have not been admitted to a higher prioritized program. The reduced form estimate is 0.053, which means that crossing the threshold is associated with an increase in income of approximately 5.3 percent, though the discontinuity is not visually convincing. A fuzzy regression discontinuity estimate of the application-weighted LATE of admittance is 0.09 (se=1.7).Though this is weak evidence of positive returns to the preferred field, this estimate masks considerable heterogeneity, as I will return to below. Additionally, the gains reflect early-career outcomes and do not take into account that lifetime income profiles may differ substantially across programs.

A necessary condition for IV to provide consistent estimates of the program-specific LATEs is a sufficiently strong first stage. As the control function is estimated on program level, Table 2 presents the distribution of first stage F-statistics for the four combinations of quotas and the definition of threshold crossing. Across identification methods, threshold crossing appears to be a strong instrument for admittance with median F-statistics between 16.8 and 40.4. However, the F-statistics are in general much stronger for program cutoffs compared to using cutoff in higher prioritized programs. This motivates the imputation order presented in Section 5.1. Some F-statistics are small which will bias the IV estimates toward the OLS counterpart. This bias will be mitigated by shrinking the estimates towards the mean of the selection adjusted estimates.

The distribution of estimated LATEs As argued by Vytlacil (2002), Brinch, Mogstad, and Wiswall (2017) and Kline and Walters (2019), similarity between LATEs of the structural model and the simple 2SLS IV bolsters the credibility of estimates of β obtained through structural modeling. Figure 3a plots the estimated program-specific LATEs using the control function approach against the corresponding IV estimates using the variation from crossing own-program

[18]Because I use the position in the waiting list as the running variable, the densities are by definition uniform. Statistics on possible density manipulation would therefore solely reflect composition effects by stacking programs. I, therefore, do not report these statistics.

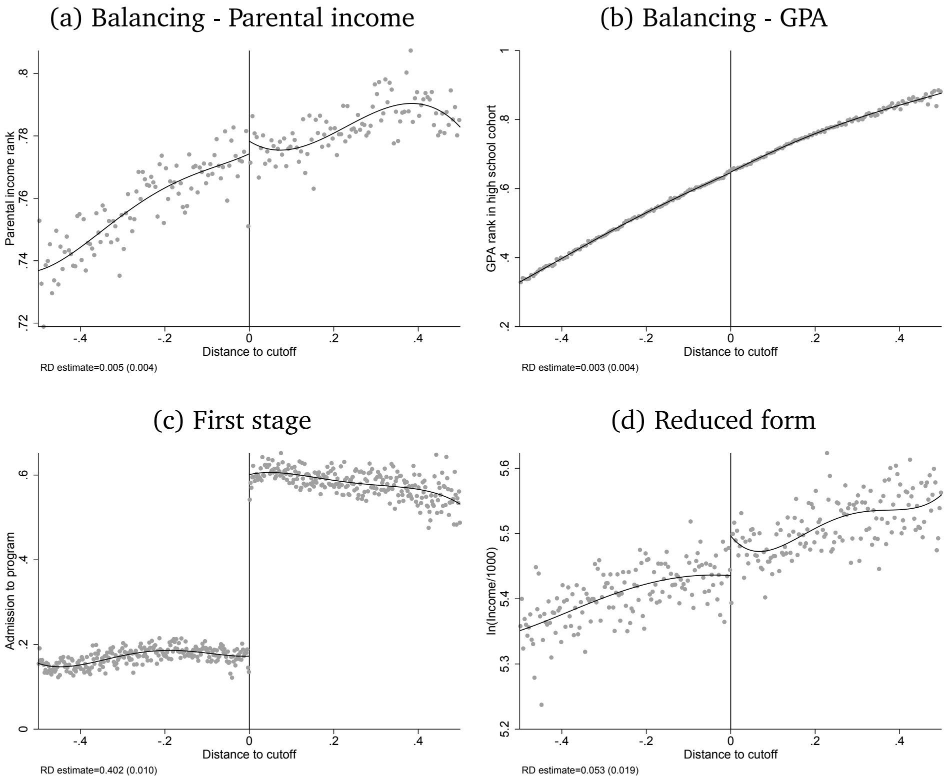


Figure 2: Regression discontinuity plots

Effects of crossing cutoffs are estimated using the rd-robust package in Stata with standard bandwidths. For the graphs, I subtract the program mean by regressing the dependent variable on a program fixed effect. Binned scatter plots are plotted with 250 bins on each side of the cutoff. The effects are estimated in the full sample across all program-year combinations and due to composition effects slopes should be interpreted with caution. For the IV estimation, observations exactly at cutoff are removed from the sample. Standard errors are clustered at the individual level as applicants can appear multiple times both across programs and years.

Table 2: F-statistics for program specific IV

Cutoff	Quota	Count	Mean	Std	Min	25%	50%	75%	Max
Program	1	268.0	122.5	252.9	0.0	12.1	40.4	125.7	2316.8
	2	202.0	156.4	274.2	0.1	16.1	58.7	167.6	2037.3
Higher priority	1	225.0	38.5	58.0	0.4	7.6	16.8	43.7	425.3
	2	111.0	28.2	32.6	0.0	5.9	18.5	35.0	170.8

The table provides descriptive statistics on the first stage F-statistic from program-specific 2SLS regressions with different running variables and associated instruments and robust standard errors. Controlling for running variables is implemented by a linear spline.

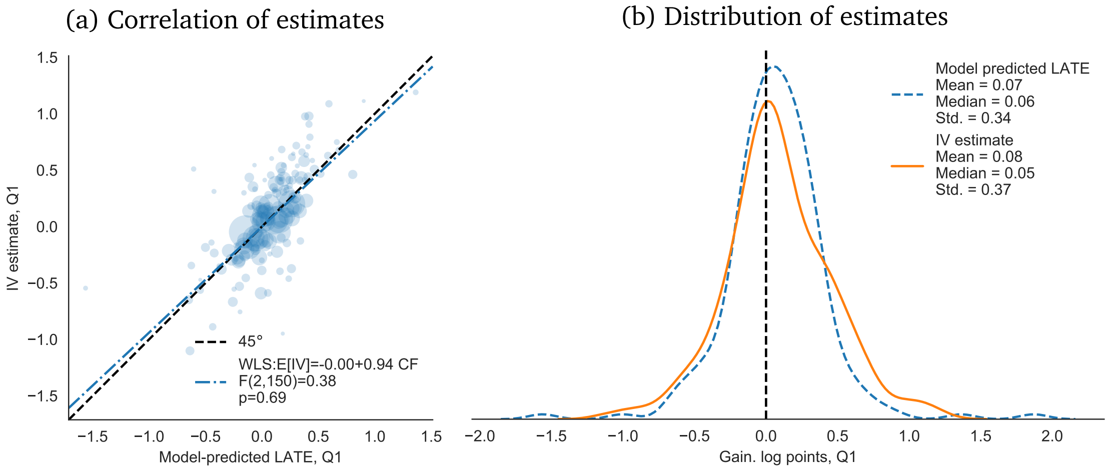


Figure 3: Model-based and IV estimates of LATE

The figure shows LATE estimates calculated from the selection correction model in Equation (12) and IV estimates of a standard fuzzy regression discontinuity design implemented with two stage least squares. The source of variation is crossing own program cutoff in quota 1. Figure 3a shows the model-based estimates on the x-axis and the corresponding IV estimate on the y-axis. The data points are weighted by sample size. The legend shows the result of weighted least squares and the corresponding F-test for whether the points are on the 45-degree line up to sampling noise. 2SLS estimates with an F-statistic lower than 10 or a p-value larger than 0.7 have been excluded. Figure 3b shows the marginal distribution of LATEs using IV and control functions.

cutoffs. To avoid comparing imprecisely estimated zeros, the programs included in the figure all have strong first stages and a p-value of the IV estimate smaller than 0.7. While the control function models are structurally specified and include controls, the figure shows a tight link between the resulting LATEs and the corresponding non-parametric IV estimates. A Wald-test of the hypothesis that the estimates are on the 45-degree line fails to reject the hypotheses with a p-value of 0.69. This provides suggestive evidence that the parametric control function approach recovers the relevant LATEs. The marginal distributions are also similar as evidenced in Figure 3b. Using the control function approach, the program-weighted mean gain of admittance for compliers in a program is 7 percent with a standard variation of 0.34. This matches the IV estimate of 8 percent.

The joint distribution of parameters and importance of correcting for selection To investigate the role of selection correction, Table 3 displays the estimated priors of the empirical Bayes model presented in Section 5.1. These estimates should be interpreted as weighted means of parameters across programs.Without controlling for selection, the average marginal return to GPA rank *whitin a program* is 0.12, implying that moving up a decile in the GPA distribution and conditional on admittance to a specific program is associated with an earnings increase of 1.2

percent as evidenced by the last row in column (1). When models are estimated across fields this parameter increases slightly to 0.16. Focusing on the sample for which sample correction using own-program cutoffs is feasible, the parameter drops by 50 percent to 0.08 in column (3), implying that rationed programs are not a random sample of programs. Once controlling for selection, the average marginal return drops marginally to 0.07. The parameter on the Mills-ratio in column (4) is positive, which reflects that the error term in selection equation (11) correlates positively with outcomes. This can be interpreted as a positive effect of motivation on outcomes. As the returns to GPA diminishes once controlling for selection, a possible interpretation is that, on average within programs, GPA tends to correlate positively with motivation. Using the higher-prioritized program cutoff as an instrument for admission, we observe higher returns to GPA as seen in columns (5) and (6). The coefficient on the inverse Mill's ratio changes sign, which is in line with expectations, as the selection equation reverses the instrument compared to baseline specification.

Columns (7) and (8) contain the corresponding results for Quota 2. Controlling for selection similarly decreases in the average marginal return to GPA from 0.9 to 0.5. The coefficient on the Mills ratio is however negative, implying that motivation and income correlate negatively. However, the applicants in Quota 2 have all incurred application costs to qualify and are therefore more selected than in Quota 1. Thus the variation in motivation picked up by the error term in the selection model is conditional on already having shown a certain level of motivation to enter the quota in the first place.[19]

Generally, having the same field as the next-best field on the rank-ordered list lowers the expected return to admission. This indicates that individuals reveal part of their type through their rank-ordered lists of applications. Further, if program expansion simply shuffles applicants around within the same fields, gains are likely to be smaller.

Predicting returns in simulation sample In the simulation sample, containing all applications in 2016, I predict the expected annual income based on the shrunken parameters, β_p^* where I set the year to 2010. The models used for predicting incomes are hierarchical; the selection correction models are preferred to models based on conditional independence assumptions. Table 4 presents the hierarchy and the share predicted by each method. The models with selection correction for crossing own-program cutoff are used to predict 64 percent of application outcomes. Estimates from crossing cutoffs in higher prioritized programs account for 8 percent. For some programs, I am not able to estimate selection correction models and therefore rely on the conditional independence assumption. Program level CIA-models predict 13 percent, the largest share next to the own-program cutoff correction model. In the sample, I am unable to predict 8 percent of the applications due to a lack of support in the estimation sample. This is largely due to new programs opening up which imply that income is not yet realized. These outcomes are imputed from the estimated prior, μ_β, using the own program cutoff in quota 1, which are the estimates presented in column 3 in Table 3.

Income is measured in 1,000 Danish Kroner (DKK) in 2015 prices. The distribution of

[19]Further, as Quota 2 is mechanically assigned a lower priority in the applicant rank-ordered list than Quota 1, there are strictly fewer options to enroll if under the cutoff and the nature of non-compliance therefore differ.

Table 3: Joint distribution of estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GPA	0.321	0.437	0.242	0.206	0.294	0.677	0.265	0.197
	(0.041)	(0.030)	(0.060)	(0.061)	(0.106)	(0.027)	(0.030)	(0.141)
GPA^2	-0.125	-0.174	-0.124	-0.086	-0.125	-0.527	-0.170	-0.128
	(0.197)	(0.195)	(0.176)	(0.182)	(0.170)	(0.371)	(0.162)	(0.119)
Same field, lower	0.001	-0.008	-0.004	-0.007	-0.014	-	-0.006	-0.011
	(0.089)	(0.115)	(0.055)	(0.051)	(0.054)	-	(0.037)	(0.032)
Same field, higher	-	-	-	-	-	-0.023	-	-
	-	-	-	-	-	(0.124)	-	-
Last appl. on ROL	-0.002	-0.004	-0.006	0.000	-0.004	0.037	-0.001	-0.011
	(0.051)	(0.047)	(0.024)	(0.028)	(0.029)	(0.067)	(0.023)	(0.040)
Year	0.003	0.000	0.001	0.002	0.000	0.000	0.000	0.002
	(0.011)	(0.012)	(0.010)	(0.010)	(0.008)	(0.009)	(0.009)	(0.009)
Constant	5.326	5.319	5.344	5.254	5.312	5.218	5.344	5.391
	(0.266)	(0.310)	(0.269)	(0.266)	(0.235)	(0.083)	(0.260)	(0.241)
λ	-	-	-	0.055	-	-0.006	-	-0.032
	-	-	-	(0.036)	-	(0.123)	-	(0.020)
Level	Program	Field	Program	Program	Program	Program	Program	Program
Correction	-	-	-	CF	-	CF	-	CF
Quota	-	-	1	1	1	1	2	2
Cutoff	-	-	Program	Program	Higher	Higher	Program	Program
$E\left[\frac{\partial y}{\partial GPA}\right]$	0.12	0.16	0.08	0.07	0.11	0.16	0.08	0.06

The table displays the estimated joint prior of parameters, μ_β in Equation (14) estimated from the program level estimates for the estimation methods outlined in Section 5. The dependent variable is the log of average income 7 to 9 years after admission. Standard errors, shown in parentheses below estimates, are the square root of the diagonal part of the estimate prior, Σ_β. Assuming uniform distribution of the ranked GPA, the average slope is calculated as $E\left[\frac{\partial y}{\partial GPA}\right] = \frac{1}{2}\beta_{GPA} + \frac{1}{3}\beta_{GPA^2}$.

Table 4: Imputation methods for $E[y|admission]$

Level	Selection correction	Quota	Cutoff	Share	Cumulative share
Program	RDD	1	Program	0.64	0.64
Program	RDD	1	Higher priority	0.09	0.73
Program	RDD	2	Program	0.02	0.75
Program	CIA	-	-	0.13	0.88
Field	CIA	-	-	0.05	0.92
-	-	-	-	0.08	1.00

The table show the distribution of imputation methods in the simulation sample. The order of the prediction methods correspond to the priority the the methods.

predicted incomes in the base level assignment in 2016 is displayed in Figure 4a. The mean predicted income is 229 thousand DKK (equivalent to 34 thousand USD in 2015 prices). For comparison, the average income for Danes aged 25-29 years in 2010 was 226 thousand DKK, implying that the magnitudes are plausible. Figure 4a also show the distribution of predicted earnings for first priority programs. The average predicted earnings would be 2 percent higher (233.8/229.2-1) if everybody would be given their first priority. This implies that rationing of education may be sub-optimal. I return to the the potential gain from eliminating rationing in section 8.

The distribution of expected gains of being admitted into Higher Education is shown in Figure 4b. These figures suggest that obtaining an admission is associated with an annual income gain of between 14 and 17 thousand DKK. The gain is slightly higher for applicants with more priorities. This may be due to differences between applicants who fill out more slots on their rank-ordered lists and those who only apply for one program. For example, younger people may be more willing to wait for admission next year than older applicants.

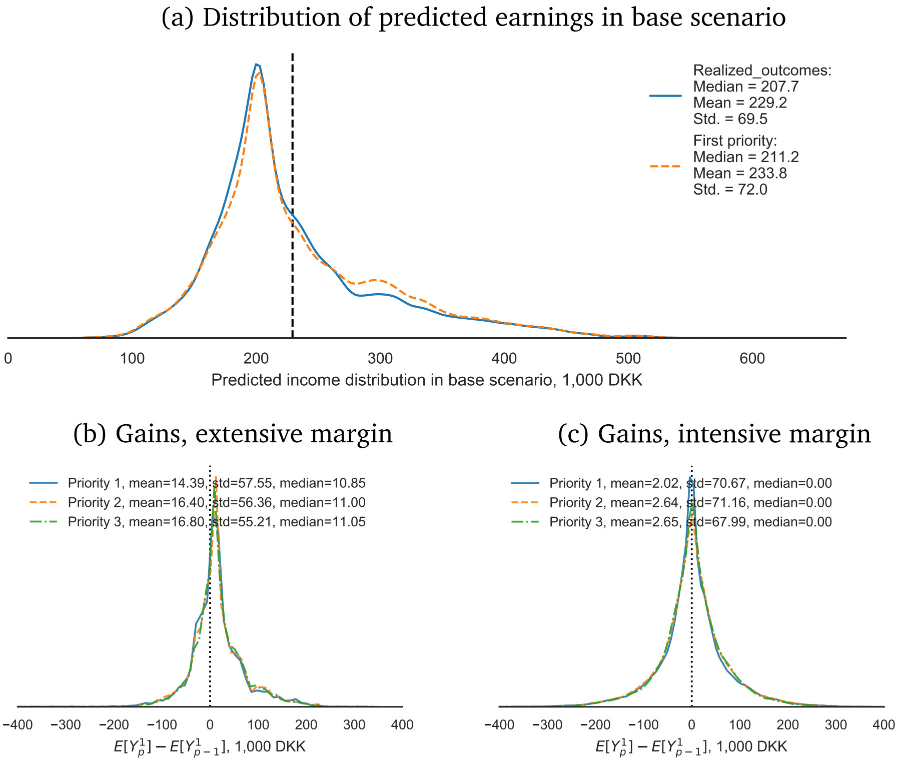


Figure 4: Predicted incomes

Figure 4a displays the distribution of predicted incomes in the simulation sample. The dashed lines show the points of truncation. Figure 4c shows the distribution of predicted gains from moving up on slot up a priority on a rank-order list, and Figure 4b shows the predicted gain from going from non-assigned to assigned. Incomes are in 1.000 DKK in 2010-prices.

Once admitted, the gains from moving up a priority on the rank-ordered list are essentially zero, though with considerable variation as evidenced in Figure 4c.

7 Clearing markets under counterfactual supply

To investigate the importance of applicant substitution I simulate a change in the supply of quota 1 slots for programs with more than 50 slots in Quota 1 in the application round of 2016. For each program, I manipulate slots to be between 70 percent smaller or larger than baseline. In order to avoid an overflow of applicants from Quota 2, I disable transfers of slots between quotas within the same program. To keep comparisons consistent, the baseline scenario is simulated in the same way. In total, I perform 4,186 simulations. For each experiment, I distinguish first-order moves and higher-order moves. First-order moves are applicants moving into and out of the program where supply is manipulated.[20] Higher-order moves are all moves that do not occur a the margin of the manipulated program.

Comparing first-order and higher-order moves Figure 5a shows that, on average, when a counterfactual supply induces 10 applicants to change program, this is associated with 15 applicants

[20] As a consequence of the deferred acceptance, an expansion of a program can only cause applicants to move into the program. Correspondingly, lowering the number of slots can only cause applicants to move out of the program.

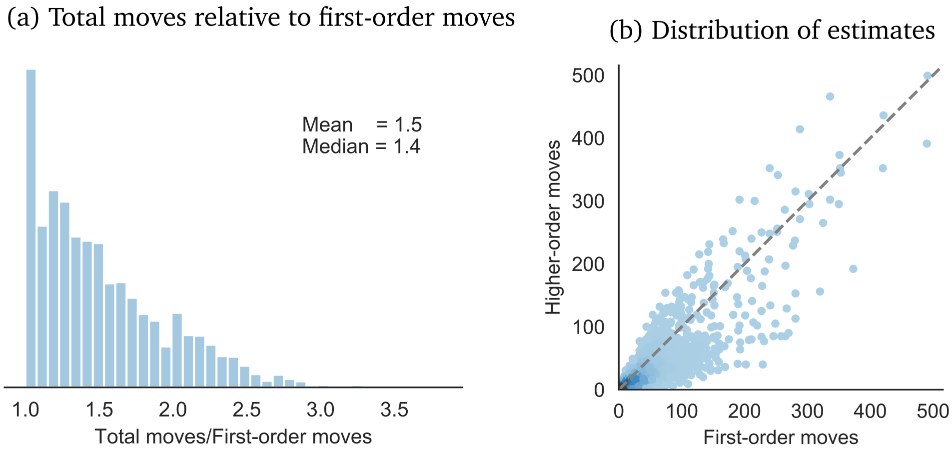


Figure 5: First-order and higher-order moves

Figure 5a plots the ratio of the total number of movers relative to the first order movers. The unit of analysis is a simulation. Figure 5b plots the number of higher-order moves against the number of first-order moves. Points are shaded according to the density of points.

changing allocations. Thus the share of applicants *indirectly* affected by a supply change is fifty percent of the share directly affected. The dispersion is, however, large, with some experiments inducing twice as many indirect shifts than direct shifts. In other words, depending on the program, a change in supply may induce a sizable externality on other programs through the structure of the applicant rank-ordered lists.

Decomposition of gains The difference in total earnings between an experiment and the baseline scenario, Δ^T, can be decomposed into four distinct sources according to whether it is from first-order (F) or higher-order (H) flows and whether it occurs on the margin between programs (intensive, I) or on the margin between receiving an offer or not (the extensive margin, E):

$$\Delta^T = \underbrace{\Delta^{F,I} + \Delta^{F,E}}_{\text{First-order}} + \underbrace{\Delta^{H,I} + \Delta^{H,E}}_{\text{Higher-order}}.$$

To investigate the share of the variation in aggregate earnings across experiments explained by the different channels I perform a variance decomposition of the stacked experiments and present the results in Table 5. The first column shows the share that each variance component attributes to the total variance in the experiments. The additional columns present different combinations of the components. The first-order components explain 60.5 percent of the variation in earnings. The corresponding share for higher-order effects is 9.3 percent and the remaining 30 percent is due to the interaction of first- and higher-order effects. This implies that 39.5 percent of the actual variation is missed by not taking higher-order effects into account. All covariance terms are positive. In other words, a higher first-order return to a program is associated with additional higher-order returns, and the correlation coefficient between first- and higher-order effects is 0.64. The last two columns of Table 5 show a decomposition into intensive and extensive movers. In isolation, the intensive margin accounts for 24.6 percent whereas the extensive margin accounts for 28.7 percent.

Table 5: Variance decomposition of earning gains in simulations

	Share	First-order	Higher-order	Intensive	Extensive
$Var(\Delta^{F,I})$	16.2	X		X	
$Var(\Delta^{F,E})$	16.9	X			X
$Var(\Delta^{H,I})$	2.6		X	X	
$Var(\Delta^{H,E})$	2.3		X		X
$2Cov(\Delta^{F,I}, \Delta^{F,E})$	27.3	X			
$2Cov(\Delta^{H,I}, \Delta^{H,E})$	4.4		X		
$2Cov(\Delta^{F,I}, \Delta^{H,I})$	5.9			X	
$2Cov(\Delta^{F,E}, \Delta^{H,E})$	9.4				X
$2Cov(\Delta^{F,I}, \Delta^{H,E})$	5.8				
$2Cov(\Delta^{F,E}, \Delta^{H,I})$	9.1				
Sum	100.0	60.5	9.3	24.6	28.7

The table displays the contribution to the total variance in gains from moves across experiments. $\Delta^{\cdot,I}$ and $\Delta^{\cdot,E}$ denote the gains on the intensive margin (program-to-program) and extensive margin (no-offer-to-program) respectively. The four right-most columns contain different summations of the variance components.

Program-level returns to adding slots The variance decomposition shows that higher-order effects are large and positively correlated with first-order effects. To investigate program-level heterogeneity further, I compute implied program level marginal returns to increasing supply. For each program, I have performed 14 experiments ranging from subtracting 70 percent to adding 70 percent of slots which depending on the size of the program correspond to a different number of slots. To quantify the marginal effect of supply changes I regress the first-order (F), higher-order (H), and the aggregate total (T) gains of the experiments on the number of slots changed:

$$\Delta^g_{pe} = \gamma^g_p(S_{pe} - S_{p0}) + \varepsilon_{pe}, \quad g \in \{F, H, T\}, \tag{15}$$

where Δ^g_{pe} is the gain from a change e to program p, S_{pe} is the number of slots in program p in simulation e and S_{p0} is the number of slots in the baseline scenario. Gains are measured in 1,000 DKK. The set of parameters, γ^g_p, can be interpreted as the marginal gain from adding a slot in a given program.

The usefulness of these measures depends on the underlying linearity between changes in slots and gains. Linearity could be violated for a number of reasons: Firstly, the applicant pool might not be large enough to fill up a program after an expansion, meaning that at some point the marginal gain from expansion drops to zero. Secondly, if the composition of the complier groups depends on the magnitude of supply change, the marginal effect of supply changes might change as a function of changing composition. To gauge the validity of the slope as a metric, Figure 6a plots the R^2 from the regressions in Equation (15). Total and first-order gains exhibit a very large degree of linearity with a median value of R^2 of 0.91 and 0.89 respectively. This implies that the first-order complier population does not change dramatically with the magnitude of the supply change and lends credence to the external validity of LATEs for first-order gains. The higher-order gains, however, exhibit less, though still substantial, linearity with a median R^2 of 0.64. The high degree of linearity implies that the set of γ^g_p provides valuable information on the effects of changes in supply. In combination with the positive correlation between first- and higher-order gains further suggests that a first-order LATE is a

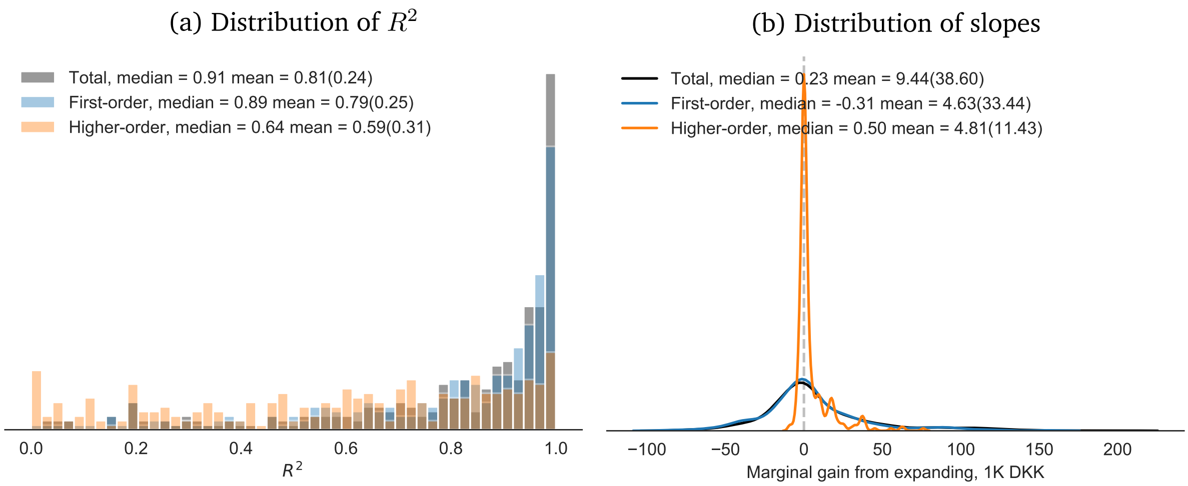


Figure 6: Marginal gain of changing supply

Figure 6a shows the distribution of R^2 from regressing aggregate gains on the number of slots as in Equation (15). Figure 6b shows the distribution of corresponding slopes, γ_p^g.

good approximation of the sign of the higher order effects.

The distribution of γ_p^g is shown in Figure 6b. The median slope for total gains is 0.23, meaning that the median marginal gain of adding 1 slot to a program is a gain of 230 DKK. However, there are large differences between programs, as some programs yield as much as 200 thousand DKK per added slot. The magnitude of first-order gains is generally larger than higher-order gains as evidenced by the standard deviations (33.44 and 11.43 respectively). Table 6 breaks down the distribution of marginal gains by program length and field. On average, long-cycle health and social science programs yield large expected marginal returns. An expansion of a long cycle health program is on average associated with a first-order gain of 25 thousand DKK. But factoring in the average higher-order effect of 16.5 thousand DKK the actual gain is 41.9 thousand. In other words, failing to factor in higher-order moves results in only identifying 60 percent of the gain from a policy expansion. For long-cycle social science programs, only about half of the gains are realized by first-order moves. Notice, however, that the standard deviations are large implying heterogeneity *within* fields. First-order effects may be counteracted by higher-order effects. For long-cycle humanities programs, the first-order loss of 500 DKK is more than compensated for by higher-order gains, such that the total marginal gain is 1.1 thousand DKK. Thus the policy implications of changing slots in humanities reverse once higher-order gains are taken into accounts. Additionally, medium-cycle STEM programs on average yield a marginal first-order loss of 4 hundred DKK. But higher-order effects mediate this effect and the total effect is, therefore, positive with a value of 2.7 thousand. Though Table 6 shows heterogeneity in returns to program expansion, it is uninformative on the interdependence between first-order and second-order gains within fields. To investigate how higher-order effects might change policy recommendations, Figure 7 plots the higher-order gains against first-order gains. The dashed black line represents break-even points where a marginal first-order gain is neutralized by the higher-order loss (or vice-versa). The programs are colored according to total gain and all programs to the right of the dashed line have positive total marginal returns.

Table 6: Marginal gains of program expansion

Length	Field	First-order	Higher-order	Total	N
Long-cycle	Health	25.4 (54.8)	16.5 (15.9)	41.9 (59.2)	12
	Humanities and communication	-0.5 (29.4)	1.6 (7.0)	1.1(31.9)	57
	STEM	3.5 (33.3)	4.2 (6.3)	7.7 (35.1)	57
	Social science and business	15.3 (48.7)	15.4 (19.5)	30.7 (58.6)	53
Medium-cycle	Health	-1.1 (14.5)	-1.1 (2.4)	-2.3 (14.9)	24
	STEM	-0.4 (42.9)	3.0 (4.5)	2.7 (41.8)	15
	Social science and business	5.1 (22.3)	2.9 (4.2)	8.0 (20.3)	17
	Teaching	-3.7 (9.7)	-0.7 (1.4)	-4.3 (10.4)	24
Short-cycle	STEM	5.8 (20.0)	1.5 (2.9)	7.3 (21.8)	10
	Social science and business	2.9 (16.5)	1.2 (2.4)	4.1 (17.9)	23

The table provides means (and standard deviations in parentheses) of the estimated slopes within fields using the model in equation (15). Only non-zero slopes are included.

The figure provides a number of takeaways. Firstly, for 69 percent of the programs, the marginal gains taking higher-order moves into account are larger than the first-order gains. This implies that the externality of investing in the added supply of slots is in general positive. Secondly, a number of programs, such as Business at Aarhus University (AU) and bachelor of engineering in design at the Technical University of Denmark (DTU), yield large marginal first-order returns to program expansion while having small second-order effects. These programs typically have low entry barriers and therefore do not create long chains of applicants. Thirdly, programs with similar first-order marginal gains can exhibit very different second-order gains, exemplified with the Business program at Copenhagen Business school which despite its similarity in content with the Aarhus program yield much larger higher order gains. Finally, even though a program yields negative first-order marginal returns this may be redeemed by higher-order gains. These programs are displayed in the upper shaded triangle. Examples of these are certain specialized business programs at CBS. However, the majority of programs are outside this triangle, which implies that negative first-order gains are not, in general, fully mitigated by higher-order gains.

Cutoffs as a measure of externality The substitution effects are due to the rationing of programs. As such, one would expect that the size of the externality is increasing in the degree of rationing and thereby the value of the GPA cutoff. Figure 8 plots ratio of the absolute value of the higher-order slope over the absolute value of the first-order slope against the GPA cutoff in the year of the simulation-sample.[21] The figure shows that the magnitude of the externality is approximately constant with a value below 0.5 for cutoffs below 10 while it increases dramatically for higher cutoffs, where the magnitude approaches 2. In other words, assuming that both slopes are positive, the higher-order earnings gain is twice as large as the first-order gain for programs with high cutoffs. This continues to hold when controlling for field and institution dummies, though controlling for institutions diminishes the ratio in the top slightly. This implies that the cutoffs are a good predictor of the magnitudes of substitution effects.

[21]Passed exams are graded on a scale between 2 and 12. In combination with bonus for multiple A-levels and an early application bonus cutoffs can therefore take values between 2 and slightly over 12.

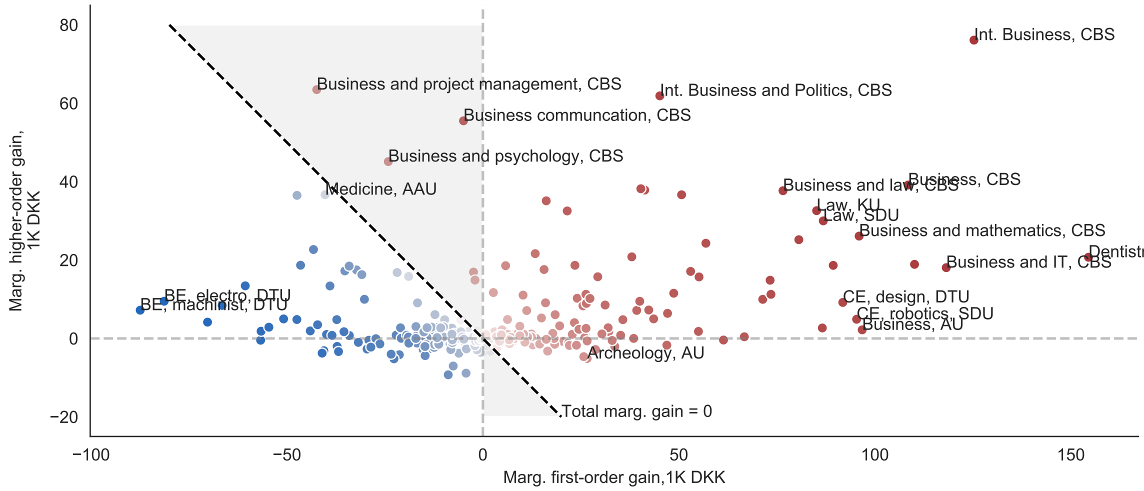


Figure 7: First-order and higher-order gains

The figure plots estimated higher-order marginal gains against first-order marginal gains. The programs are colored according to their aggregate gains, which can be read off the distance to the dashed line. Selected programs are annotated, CE and BE are abbreviations for civil and bachelor engineering respectively.

8 Efficiency loss from respecting preferences

As shown above, applicants switching between programs constitute sizable externalities that constrain policy-makers. A natural next step is to quantify how much applicants' preferences and programs' choice of ranking function (i.e. program preferences) constrain the policymaker when deciding on capacities of multiple programs simultaneously. I consider four scenarios that a policy-maker might consider when manipulating capacities. These four scenarios are defined by capacity vectors resulting from the following thought experiments: i) The baseline capacity in 2016. ii) Give everybody first priority. iii) Give everybody program with highest predicted earnings. iv) Give everybody program with highest predicted earnings including non-assignments.

For each scenario, I compute the capacities as the number of applicants who would get a seat in each program and simulate the corresponding matching using the reported preferences and eligibility scores. To quantify the importance of the constraints posed by applicants' preferences and program rankings in these scenarios I construct a counterfactual scenario where a social planner may allocate applicants to programs freely while disregarding rank-ordered lists and eligibility scores but while respecting the capacity of each program. I assume that the social planner maximizes predicted earnings and the solution to this maximization problem constitutes the upper bound of total earnings for a given set of capacities. The deviation of the matching in the admission system from this unconstrained maximum is a measure of the constraint imposed by applicant preferences and program priorities.[22]

This maximization problem of the social planner is equivalent to a discrete Monge-Kantorovich transportation problem (Monge, 1781; Kantorovich, 1942) where programs are suppliers and

[22] I only allow the social planner to allocate an applicant to a given program if they have applied to the program but ignore the applicant's ranking of programs.

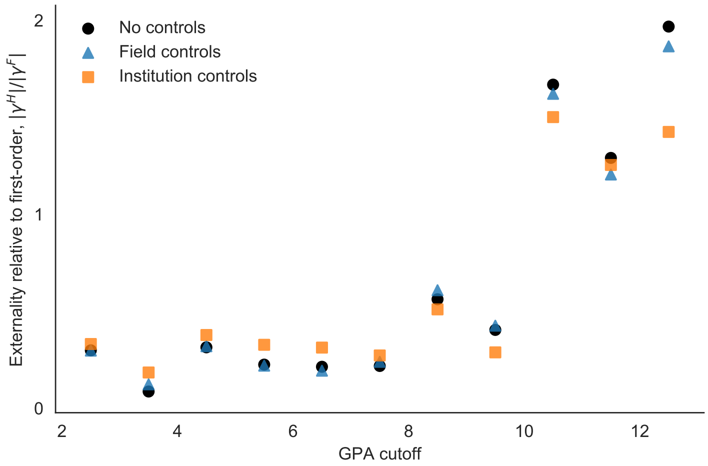


Figure 8: Substitution externality as a function of GPA cutoff

The figure plots the ratio of the absolute value of the higher-order slope over the absolute value of the first-order slope, $|\gamma^H|/|\gamma^F|$, as the function of the realized GPA-cutoff in 2016. The triangles and squares represent residuals from OLS where the outcomes are regressed on field and institution dummies respectively.

applicants are consumers with unit demand.[23] Due to the multi-quota nature of the admission system where applicants might apply in multiple quotas, there is no obvious way to set capacities across quotas within programs in the scenarios. I, therefore, restrict my attention to Quota 1 where applicants are ranked according to GPA, and remove all other quotas and applications. To keep scenarios comparable I recompute the baseline scenario 1 with only Quota 1 applications. For all scenarios, I allow the optimization routine to not admit a given student. This exercise rests on the credibility of the predicted earnings and the restricted version of the admission system, and the results should therefore be interpreted with appropriate caveats. The results for the four scenarios are displayed in Table 7.

Scenario 1: Baseline scenario The baseline matching achieves aggregate predicted earnings of 15.6 billion DKK. For the baseline capacities, however, the maximum predicted earnings amount to 16.9 billion DKK, implying a loss in predicted outcomes of eight percent. This loss of efficiency is due to the structure of preferences, eligibility scores, and the algorithm. Moving to an optimal allocation would increase inequality by 24 percent as measured by mean log deviations (MLD). The increase in inequality is partly a result of applicants shifting up and down in earnings as evidenced by the rank correlation of 0.68 between predicted earnings in the baseline scenario and the optimal scenario given the same capacities. A switch to an optimal allocation would affect around 55 percent of the applicants. 16 percent would receive a new program and half of those would be switching fields. This indicates the (estimated) program-specific match components are important over and above a field-specific component. The majority of flows, however, occur along the extensive margin. 12 percent of applicants

[23] I implement this problem as a minimum-cost-flow problem on a bi-partite directed graph where the nodes are a disjoint set of programs and applicants and an edge between a program and an applicant is present whenever an applicant has applied for the program. The problem is solved by balancing the problem by adding a sink to allocate empty slots and using the algorithms provided by Hagberg, Schult, and Swart (2008). The unconstrained optimum of scenario 4 can be easily computed by picking the maximum predicted earnings for each applicant. The solutions to the transportation problem in scenario 4, therefore, constitute a check on the calculations.

Table 7: Counterfactual allocations and quantification of constraints

Scenario		Naive	Matching	Optimal	Match/Opt-1
1. Baseline	Aggregate pred. earnings (bill. DKK)	-	**15.57**	16.90	-0.08
	Inequality (MLD)	-	**0.04**	0.03	0.24
	Rank corr. with baseline pred. earn.	-	**1.00**	0.68	0.46
	Share changing status	-	**0.00**	0.55	-
	Share changing program	-	**0.00**	0.16	-
	Share changing field	-	**0.00**	0.08	-
	Share entering on extensive margin	-	**0.00**	0.12	-
	Share exiting on extensive margin	-	**0.00**	0.27	-
2. First priority	Aggregate pred. earnings (bill. DKK)	16.16	16.16	17.64	-0.08
	Inequality (MLD)	0.04	0.04	0.03	0.23
	Rank corr. with baseline pred. earn.	0.78	0.78	0.68	0.14
	Share changing status	0.51	0.51	0.63	-0.19
	Share changing program	0.17	0.17	0.22	-0.22
	Share changing field	0.09	0.09	0.11	-0.26
	Share entering on extensive margin	0.34	0.34	0.22	0.53
	Share exiting on extensive margin	0.00	0.00	0.19	-
3. Highest outcome	Aggregate pred. earnings (bill. DKK)	17.73	16.57	18.16	-0.09
	Inequality (MLD)	0.04	0.04	0.04	0.10
	Rank corr. with baseline pred. earn.	0.70	0.79	0.69	0.14
	Share changing status	0.63	0.43	0.65	-0.35
	Share changing program	0.29	0.16	0.25	-0.38
	Share changing field	0.14	0.08	0.13	-0.41
	Share entering on extensive margin	0.34	0.25	0.27	-0.08
	Share exiting on extensive margin	0.00	0.02	0.13	-0.82
4. Highest outcome (incl. NA)	Aggregate pred. earnings (bill. DKK)	18.16	16.81	18.16	-0.07
	Inequality (MLD)	0.04	0.04	0.04	0.09
	Rank corr. with baseline pred. earn.	0.69	0.77	0.69	0.11
	Share changing status	0.65	0.42	0.65	-0.36
	Share changing program	0.25	0.16	0.25	-0.36
	Share changing field	0.13	0.08	0.13	-0.37
	Share entering on extensive margin	0.27	0.18	0.27	-0.31
	Share exiting on extensive margin	0.13	0.07	0.13	-0.46

The "Naive" column shows result of selecting the outcomes freely according to the strategy. The “Matching” column shows the implied matching using the observed rank-ordered lists and eligibility scores. The "Optimal" column shows the optimal solution given the implied capacities resulting from allocating capacity corresponding to the allocations in the “Naive” column while optimizing aggregated predicted earnings. All shares are compared to the baseline matching scenario.

would be admitted whereas 27 percent would be forced out of education – a net outflow of 15 percent. Thus, optimality would imply that fewer applicants would be admitted.[24]

Scenario 2: All applicants get their highest priority In the second scenario, I let the capacity be determined by the first priorities of applicants. Online Appendix Figure 10a shows that in this scenario, the largest increases in field sizes would be medium-cycle health programs (e.g. nursing and midwives), and long-cycle humanities, social studies, and business programs. Granting every applicant their first priority would yield an aggregate outcome of 16.16 billion DKK, a gain of (16.16 /15.57-1)=3.8 percent. This is larger than the two percent increase found in section 6 due to the exclusion of Quota 2 applicants from this exercise. Though this effect is estimated for a restricted market of Quota 1 applicants this result indicate that it might be desirable to do away with rationing of programs entirely. Unsurprisingly, the admission system allocates everybody to their first priority and yields the same effect as the naive assignment. Notably, moving away from rationing has no cost on inequality which is maintained at 0.04 in terms of mean log deviations. While aggregate earnings increase, the earnings under optimal assignment increase as well to 17.6 billion DKK. The result is again that 8 percent of aggregate earnings are lost due to preferences. A move to optimal allocation from the baseline scenario implies large movements on the extensive margin, where 22 percent of applicants would enter education, while 19 percent would exit, implying a net inflow of 3 percent. Figure 9 shows the resulting field sizes from allocating applicants optimally compared to the field sizes if everybody would have gotten their first priority. The largest absolute decrease is in long-cycle humanities programs, implying low gains to these programs. These applicants are mostly pushed out of education entirely, implying high outside options of entering education in a given year. As the elimination of rationing increases aggregate expected earnings and applicants' reported preferences are observable to the policy-maker they seem to be a reasonable guide for setting capacities. Further, if applicants rank according to expected utility including other non-pecuniary aspects, the gain in expected earnings is most likely underestimating the utility gain from giving everybody access to their preferred education.[25]

Scenario 3 and 4: All applicants get the program with highest predicted earnings The third and fourth scenarios assume knowledge of all potential outcomes of all applicants and are therefore unrealistic scenarios. Nonetheless, the two scenarios represent the boundary of achievable earnings. Were everybody to be assigned the program with their highest predicted earnings (including the possibility of being assigned the outside option) the aggregate earnings would amount to 18.16 billion DKK, an increase compared to the baseline of (18.16/15.57-1)=17 percent as seen in Table 7. In scenario 3 the social planner sets capacity to reflect the naive optimal assignment *within the ranked ordered list* ignoring the outside option. This lowers the naive gain to 17.7 billion DKK as some applicants have better outside options. The resulting assignment results in aggregate earnings of 16.6 billion, reflecting that percent of earnings are foregone by not being at the optimal allocation.

[24]An important caveat is that the outside opting is estimated in a static framework and partly reflects future admittance into education. A high outside option will therefore both reflect different educational choices and possibly maturing of the applicant.

[25]This holds with the caveat that reported preferences are not top-censored. Figures corresponding to 9 for the two remaining scenarios are in Appendix Figure 11.

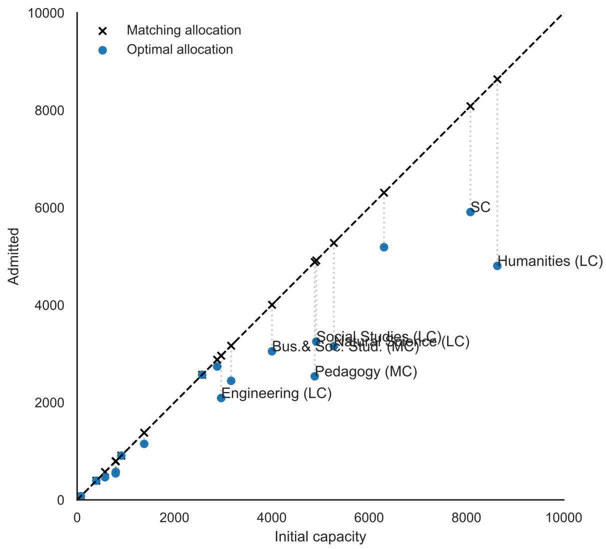


Figure 9: Field sizes with no rationing

The figure shows field sizes in the baseline scenario 2 and the resulting optimal assignment in terms of aggregate predicted earnings given the implied capacities. Changes of less than ten observations are excluded. Corresponding figures for scenarios 2 and 4 are shown in Online Appendix Figure 11.

In the fourth scenario, the social planner takes outside options into account, and thus the naive allocation equals the optimal assignment with predicted earnings of 18.16 billion DKK. Unsurprisingly, the naive (and optimal) allocation pushes applicants out of education but also pulls in applicants resulting in a net inflow of 14 percent (calculated as 0.27-0.13). When running the assignment mechanism, the resulting aggregate earnings amount to 16.81 billion, an increase compared to the baseline of 8 percent, but a loss relative to optimal allocation of 7 percent.

This quantification exercise rests on a number of important assumptions which are stricter than needed for providing evidence of spill-overs which has been the objective of the previous sections of the paper. I assume that the predicted earnings are correctly estimated and outside options are held constant. Thus the thought experiment is a one-year change where after allocation returns to normal. Additionally, I take applicant preferences as given and ignore that applicants might have left out programs on their application lists. In the presence of application costs, applicants may leave out programs both at the top and bottom of their lists, and the lists would therefore presumably change if program capacities undergo large changes. The computations are done ignoring the multi-quota nature of the Danish admission system and realized policy will for this reason alone differ from the results presented here. Interpretation of results in this section should, therefore, reflect the constrained nature of this exercise in optimal allocation.

9 Conclusion

Market clearing in the educational market can enhance or counteract the effects of local supply changes and without knowing the externality imposed on other programs through market-clearing, policy recommendations based on local effects alone may be severely off. While this finding is discouraging I find that the externality on average is positive, implying that if the first-order estimate is positive, the value of investing in increasing the supply of oversubscribed programs is, on average, underestimated.

It is important to note that the quantitative results hinge on an assumption of unchanged demand. While the mechanism is in principle strategy-proof, the number of programs that applicants can rank is limited. Applicants may therefore leave out unrealistic options entirely as investigated by Calsamiglia, Haeringer, and Klijn (2010). Thus, assuming that a supply change is public knowledge, non-marginal changes in supply might induce applicants to change behavior either by adding or removing programs from their rank-ordered lists. In other words, demand may adjust to a changed supply. As rank-ordered lists are generally not filled out, this may especially be an issue of truncation at the top of the list. A method to counter this would be to estimate preferences under assumptions outlined by Fack, Grenet, and He (2019) and fill out rank-ordered lists. However, one would need to impose a rigid structure on preferences, and the qualitative conclusion that substitution chains matter would remain unchanged. If anything I conjecture that if one had access to full rank-ordered lists of all programs, the substitution chains would likely become longer. The present results are therefore most likely a lower bound on the magnitude of externalities.

The results are obtained in the Danish context of centralized assignment but the qualitative finding that substitution effects matter does not hinge on this specific context nor on the centralized nature of the market. An educational slot is by definition indivisible and applicants will (in most cases) have unit demand. This means that whenever supply is limited, substitution may substantially change policy implications derived from local effect estimates. However, the results also generalize to other markets. Interpreting the cutoffs as prices, the results suggest that pecuniary externalities are important even without distributional concerns. In other words, if the purpose is to formulate policy, well-specified, policy-relevant treatment effects should be complemented with considerations of market clearings.

References

ABDULKADIROĞLU, A., N. AGARWAL, AND P. A. PATHAK (2017): "The welfare effects of coordinated assignment: Evidence from the New York City high school match," *American Economic Review*, 107(12), 3635–89.

ABDULKADIROĞLU, A., J. D. ANGRIST, Y. NARITA, AND P. A. PATHAK (2017): "Research design meets market design: Using centralized assignment for impact evaluation," *Econometrica*, 85(5), 1373–1432.

ABDULKADIROGLU, A., J. D. ANGRIST, Y. NARITA, AND P. A. PATHAK (2022): "Breaking ties: Regression discontinuity design meets market design," *Econometrica*.

ABDULKADIROĞLU, A., AND T. SÖNMEZ (2003): "School choice: A mechanism design approach," *American economic review*, 93(3), 729–747.

ABDULKADIROĞLU, A., P. A. PATHAK, J. SCHELLENBERG, AND C. R. WALTERS (2020): "Do Parents Value School Effectiveness?," *American Economic Review*, 110(5), 1502–39.

AGARWAL, N. (2015): "An empirical model of the medical match," *American Economic Review*, 105(7), 1939–78.

——— (2017): "Policy analysis in matching markets," *American Economic Review*, 107(5), 246–50.

AGARWAL, N., AND P. SOMAINI (2018): "Demand analysis using strategic reports: An application to a school choice mechanism," *Econometrica*, 86(2), 391–444.

ALTONJI, J. G., P. ARCIDIACONO, AND A. MAUREL (2016): "The analysis of field choice in college and graduate school: Determinants and wage effects," in *Handbook of the Economics of Education*, vol. 5, pp. 305–396. Elsevier.

ANDERSEN, M. B., S. HØRLÜCK, AND C. L. SØRENSEN (2020): "Effekt af optag på videregående uddannelser for marginale studerende," Discussion paper, Working Paper.

ANDREWS, R. J., J. LI, AND M. F. LOVENHEIM (2016): "Quantile treatment effects of college quality on earnings," *Journal of Human Resources*, 51(1), 200–238.

ARCIDIACONO, P., E. M. AUCEJO, AND V. J. HOTZ (2016): "University Differences in the Graduation of Minorities in STEM Fields: Evidence from California," *American Economic Review*, 106(3), 525–62.

AZEVEDO, E. M., AND J. D. LESHNO (2016): "A supply and demand framework for two-sided matching markets," *Journal of Political Economy*, 124(5), 1235–1268.

BRINCH, C. N., M. MOGSTAD, AND M. WISWALL (2017): "Beyond LATE with a discrete instrument," *Journal of Political Economy*, 125(4), 985–1039.

BUCAREY, A. (2018): "Who pays for free college? Crowding out on campus," *Job market paper*.

CALSAMIGLIA, C., G. HAERINGER, AND F. KLIJN (2010): "Constrained school choice: An experimental study," *American Economic Review*, 100(4), 1860–74.

DALE, S. B., AND A. B. KRUEGER (2002): "Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables," *The Quarterly Journal of Economics*, 117(4), 1491–1527.

DALY, M., M. F. JENSEN, AND D. LE MAIRE (2020): "University admission and preferred field of study," Discussion paper, Working Paper.

FACK, G., J. GRENET, AND Y. HE (2019): "Beyond truth-telling: Preference estimation with centralized school choice and college admissions," *American Economic Review*, 109(4), 1486–1529.

FELLER, A., T. GRINDAL, L. MIRATRIX, L. C. PAGE, ET AL. (2016): "Compared to what? Variation in the impacts of early childhood education by alternative care type," *Annals of Applied Statistics*, 10(3), 1245–1285.

GALE, D., AND L. S. SHAPLEY (1962): "College admissions and the stability of marriage," *The American Mathematical Monthly*, 69(1), 9–15.

HAGBERG, A. A., D. A. SCHULT, AND P. J. SWART (2008): "Exploring Network Structure, Dynamics, and Function using NetworkX," in *Proceedings of the 7th Python in Science Conference*, ed. by G. Varoquaux, T. Vaught, and J. Millman, pp. 11 – 15, Pasadena, CA USA.

HASTINGS, J. S., C. A. NEILSON, AND S. D. ZIMMERMAN (2013): "Are some degrees worth more than others? Evidence from college admission cutoffs in Chile," Discussion paper, National Bureau of Economic Research.

HECKMAN, J. J. (1979): "Sample selection bias as a specification error," *Econometrica: Journal of the econometric society*, pp. 153–161.

HECKMAN, J. J., S. URZUA, AND E. VYTLACIL (2008): "Instrumental variables in models with multiple outcomes: The general unordered case," *Annales d'Economie et de Statistique*, pp. 151–174.

HECKMAN, J. J., AND E. VYTLACIL (2001): "Policy-relevant treatment effects," *American Economic Review*, 91(2), 107–111.

HEINESEN, E., AND C. HVID (2019): "Returns To Field Of Study In The Medium Term – Instrumental Variables Estimates Based On Admission Thresholds," Discussion paper, Working Paper.

HUMLUM, A., AND B. B. MEYER (2020): "Artificial Intelligence and College Majors," Discussion paper, Working Paper.

IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

KANTOROVICH, L. V. (1942): "On the translocation of masses.(Dokl.) Acad," *Sci. URSS*, 37(3), 199–201.

KAPOR, A. J., C. A. NEILSON, AND S. D. ZIMMERMAN (2020): "Heterogeneous beliefs and school choice mechanisms," *American Economic Review*, 110(5), 1274–1315.

KIRKEBOEN, L. J., E. LEUVEN, AND M. MOGSTAD (2016): "Field of study, earnings, and self-selection," *The Quarterly Journal of Economics*, 131(3), 1057–1111.

KLINE, P., AND C. R. WALTERS (2016): "Evaluating public programs with close substitutes: The case of Head Start," *The Quarterly Journal of Economics*, 131(4), 1795–1848.

——— (2019): "On heckits, late, and numerical equivalence," *Econometrica*, 87(2), 677–696.

MANNING, A., AND B. PETRONGOLO (2017): "How local are labor markets? Evidence from a spatial job search model," *American Economic Review*, 107(10), 2877–2907.

MONGE, G. (1781): "Mémoire sur la théorie des déblais et des remblais," *Histoire de l'Académie Royale des Sciences de Paris*.

MORRIS, C. N. (1983): "Parametric empirical Bayes inference: theory and applications," *Journal of the American statistical Association*, 78(381), 47–55.

MOUNTJOY, J., AND B. HICKMAN (2020): "The Returns to College (s): Estimating Value-Added and Match Effects in Higher Education," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, 1(2020-08).

ROBBINS, H. E. (1992): "An empirical Bayes approach to statistics," in *Breakthroughs in statistics*, pp. 388–394. Springer.

ROY, A. D. (1951): "Some thoughts on the distribution of earnings," *Oxford economic papers*, 3(2), 135–146.

SANDMO, A. (2015): "Chapter 1 - The Principal Problem in Political Economy: Income Distribution in the History of Economic Thought," in *Handbook of Income Distribution*, ed. by A. B. Atkinson, and F. Bourguignon, vol. 2 of *Handbook of Income Distribution*, pp. 3–65. Elsevier.

TANAKA, M., Y. NARITA, AND C. MORIGUCHI (2020): "Meritocracy and Its Discontent: Long-run Effects of Repeated School Admission Reforms," Discussion paper, Research Institute of Economy, Trade and Industry (RIETI).

VYTLACIL, E. (2002): "Independence, monotonicity, and latent index models: An equivalence result," *Econometrica*, 70(1), 331–341.

YOUNG, A. (2019): "Consistency without inference: Instrumental variables in practical application," Discussion paper, Working Paper.

ZIMMERMAN, S. D. (2014): "The returns to college admission for academically marginal students," *Journal of Labor Economics*, 32(4), 711–754.

A Appendix

A.1 Methods

A.1.1 Calculation of the Local Average Treatment Effect

This appendix briefly describe how the Local Average Treatment Effect (LATE) is calculated in the control function approach and how it is calculated in a standard fuzzy discontinuity design.

Control function estimation of LATE The control function estimates are estimated on a program level using a standard two-step switching regression framework. To simplify notation I suppress the program and quota indices used in the main text. Let D_i be a binary indicator for receiving offer. Again, e_i is the eligibility score centered around the cutoff and $Z_i = \mathbf{1}(e_i > 0)$. For each program with associated sample of applicants I estimate by probit the selection equation with a linear spline in the running variable:

$$D_i = \mathbf{1}\left\{\gamma_1 Z_i + \gamma_2 e_i + \gamma_3 Z_i \times e_i + \epsilon_i > 0\right\}.$$

Let Φ be the standard normal distribution and ϕ the associated density. Using the estimated parameters in the selection equation, I construct the inverse Mill's ratio and estimate the following regressions for treated and untreated respectively by OLS:

$$\begin{aligned} Y_i &= X_i\beta_1 + \psi_1 \frac{\phi(\hat{\gamma}_1 Z_i + \hat{\gamma}_2 e_i + \hat{\gamma}_3 Z_i \times e_i)}{\Phi(\hat{\gamma}_1 Z_i + \hat{\gamma}_2 e_i + \hat{\gamma}_3 Z_i \times e_i)} + u_i, \text{ if } D_i = 1 \\ Y_i &= X_i\beta_0 + \psi_0 \frac{-\phi(\hat{\gamma}_1 Z_i + \hat{\gamma}_2 e_i + \hat{\gamma}_3 Z_i \times e_i)}{1 - \Phi(\hat{\gamma}_1 Z_i + \hat{\gamma}_2 e_i + \hat{\gamma}_3 Z_i \times e_i)} + v_i, \text{ if } D_i = 0 \end{aligned}$$

Again, using the estimated values of the parameters from the selection model I calculate the following individual correction term:

$$\Gamma_i(e_i) = \frac{\phi\left(\hat{\gamma}_1 + (\hat{\gamma}_2 + \hat{\gamma}_3)e_i\right) - \phi\left(\hat{\gamma}_2 e_i\right)}{\Phi\left(\hat{\gamma}_1 + (\hat{\gamma}_2 + \hat{\gamma}_3)e_i\right) - \Phi\left(\hat{\gamma}_2 e_i\right)}.$$

I then proceed to construct expected outcomes for $d \in (0, 1)$ using the estimates of β_d. The LATE in the control function approach, τ_{CF}, is calculated as the sample analog to following expectation over the full sample:

$$\tau_{CF} = E[X_i(\hat{\beta}_1 - \hat{\beta}_0) + (\hat{\psi}_1 - \hat{\psi}_0)\Gamma_i(e_i)],$$

Instrumental variable estimation of LATE The IV estimate of the LATE in the fuzzy regression discontinuity framework is obtained using the following first stage:

$$D = \alpha_1 Z_i + \alpha_2 e_i + \alpha_3 Z_i \times e_i + \epsilon_i$$

The predicted treatment value is then used in the following second stage:

$$Y_i = \tau_{IV}\hat{D}_i + \eta_2 e_i + \eta_3 Z_i \times e_i + u_i,$$

where the estimate of τ_{IV} is the LATE in the fuzzy regression discontinuity design.

A.1.2 Empirical Bayes shrinkage

In this section I provide details on the Empirical Bayes shrinkage procedure outlined in Section 5.1. As mentioned in the main text, this procedure follows Abdulkadiroğlu, Pathak, Schellenberg, and Walters (2020) closely. The estimates based on conditional independence and the control function approach each return a set of program specific estimates, $\left\{\hat{\beta}_p\right\}_{p=1}^{P}$, where for the control function I include the parameter on the correction term, $\hat{\psi}_p \in \hat{\beta}_p$.

Let K be the length of the vector of parameters. Under the hierarchical model outlined in Equations (13) and (14), the likelihood of the estimates for program p conditional on the unobserved parameters, β_p, and the associated covariance matrix, Ω_p, is

$$\mathcal{L}\left(\hat{\beta}_p|\beta_p,\Omega_p\right) = (2\pi)^{-K/2}\left|\Omega_p\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\left(\hat{\beta}_p-\beta^p\right)'\Omega_p^{-1}\left(\hat{\beta}_p-\beta^p\right)\right)$$

Assuming that my estimates of Ω_p are accurately approximated, the integrated likelihood function conditioning only on hyperparameters is then

$$\begin{aligned}\mathcal{L}^I\left(\hat{\beta}_p|\mu_\beta,\Sigma_\beta,\Omega_p\right) &= \int \mathcal{L}\left(\hat{\beta}_p|\beta_p,\Omega_p\right)dF\left(\beta_p|\mu_\beta\Sigma_\beta\right)\\ &= (2\pi)^{-K/2}\left|\Omega_p+\Sigma_\beta\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\left(\hat{\beta}_p-\mu_p\right)'\left(\Omega_p+\Sigma_p\right)^{-1}\left(\hat{\beta}_p-\mu_p\right)\right).\end{aligned}$$

Empirical Bayes estimates of the hyperparameters are obtained by maximizing the integrated log likelihood function where I plug in estimates, $\hat{\Omega}_p$, for Ω_p:

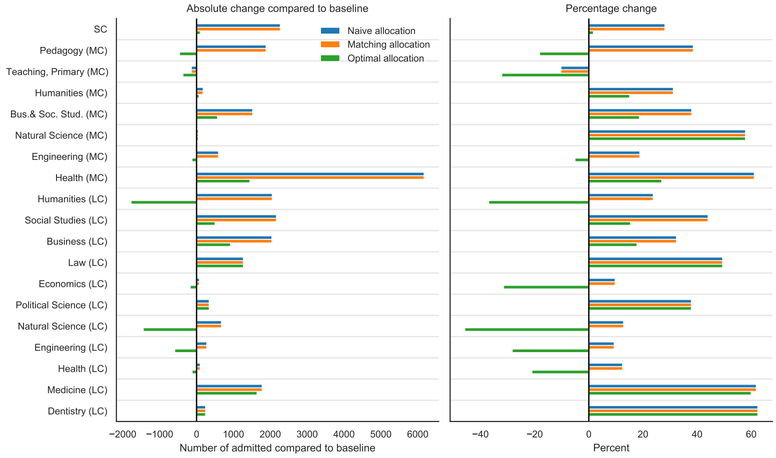
$$(\mu_p,\Sigma_p) = \arg\max\sum_p \log\mathcal{L}^I\left(\hat{\beta}_p|\mu_\beta\Sigma_\beta,\hat{\Omega}_p\right).$$

In Table 3, I report the square root of the diagonal elements of $\hat{\Sigma}_p$ under the parameters in $\hat{\mu}_p$. Using the estimates, $\left(\hat{\mu}_p,\hat{\Sigma}_p,\hat{\Omega}_p\right)$, posteriors of β_p and Ω_p are obtained as follows:

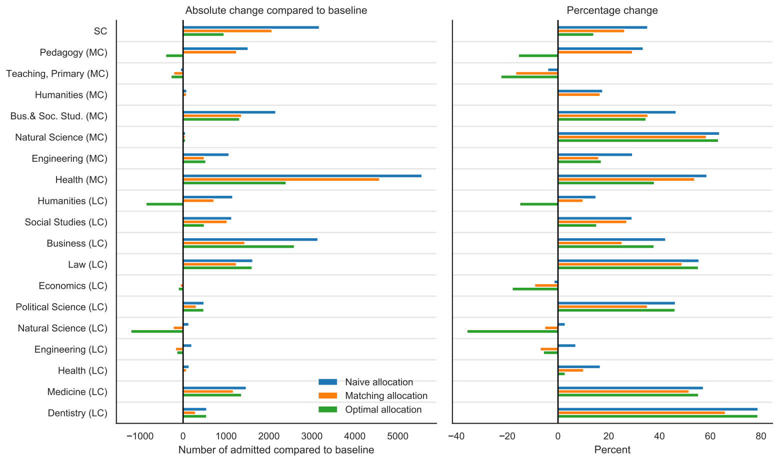
$$\begin{aligned}\beta_p^* &= \left(\hat{\Omega}_p^{-1}+\hat{\Sigma}_\beta^{-1}\right)^{-1}\left(\hat{\Omega}_p^{-1}\hat{\beta}_p+\hat{\Sigma}_\beta^{-1}\hat{\mu}_\beta\right)\\ \Omega_p^* &= \left(\hat{\Omega}_p^{-1}+\hat{\Sigma}_\beta^{-1}\right)^{-1}.\end{aligned}$$

The procedure is implemented in Python and code is provided in the data appendix.

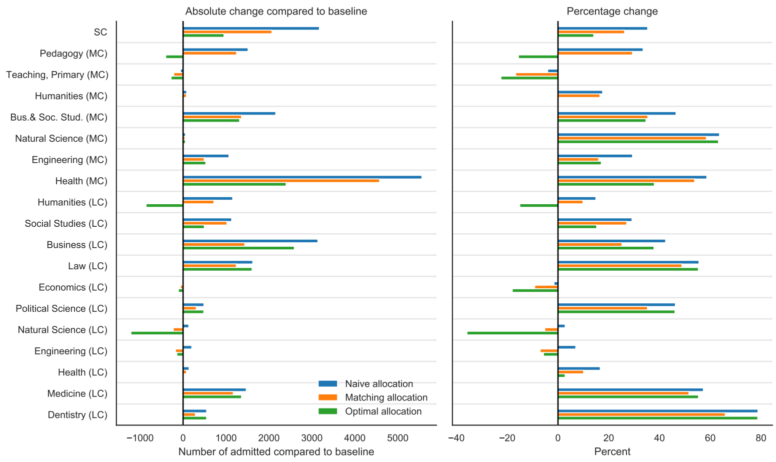
B Additional Figures



(a) Scenario 2: Give everybody first priority



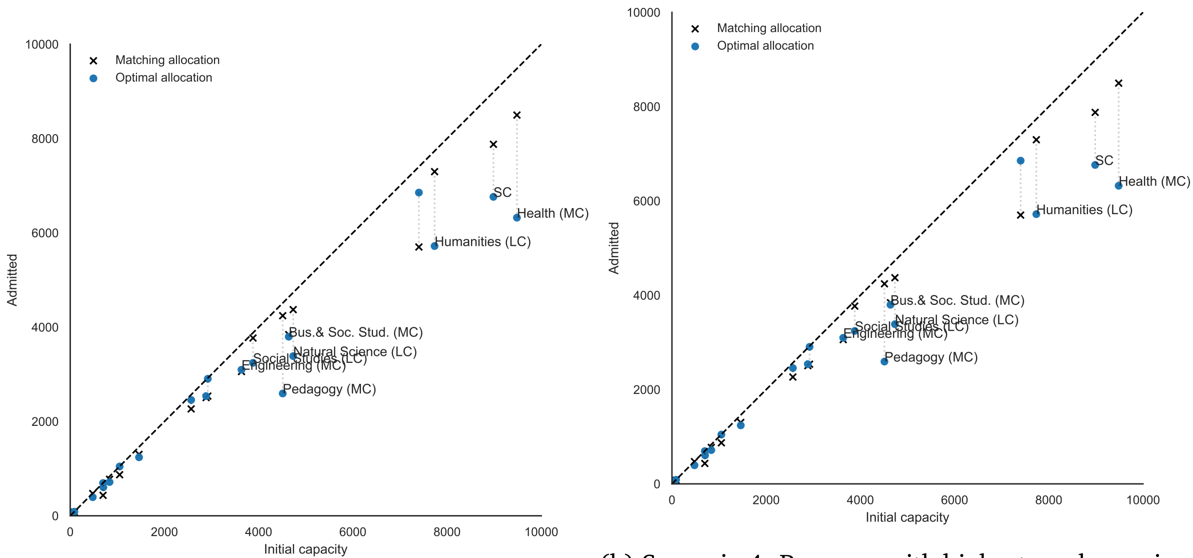
(b) Scenario 3: Give everybody program with highest pred. earnings



(c) Scenario 4: Give everybody program with highest pred. earnings (incl. non-assignments)

Figure 10: Changes field sizes and counterfactual scenarios

The figure shows changes in field sizes allocation under scenarios 2, 3, and 4 compared to the baseline scenario. The "Naive assignment" shows the result of selecting the outcomes freely according to the strategy. The “Matching assignment” shows the implied matching using the observed rank-ordered-lists and eligibility scores. The "Optimal assignment" shows the optimal solution given the implied capacities resulting from allocating capacity corresponding to the allocations in the “Naive” scenario while optimizing aggregated predicted earnings. All shares are compared to the baseline matching scenario. Flows and stocks with less than ten observations are excluded.



(a) Scenario 3: Program with highest pred. earnings

(b) Scenario 4: Program with highest pred. earnings (incl. non-assignments)

Figure 11: Changes field sizes and counterfactual scenarios

The figure shows field sizes allocation under scenarios 2, 3, and 4, the number of admitted using the actual assignment mechanism with given preferences of applicants and programs and the resulting optimal assignment in terms of aggregate predicted earnings given the implied capacities. Changes of less than ten observations are excluded.