Substitution Effects in College Admissions

Mikkel Høst Gandil*

September 2021

Newest version here

Abstract

I show how local supply changes in programs create ripple effects in an educational market of an entire country. 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 equilibria under counterfactual supply. I estimate potential earnings with a regression discontinuity design and quantify market-clearings in terms of earnings. 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. I document a trade-off between earnings and inequality.

JEL Codes: C63,I21,I22,I24, I26, J24, H52

*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 as well as seminar participants at Statistics Norway for helpful comments and suggestions.

1 Introduction

The return to investing in education is a first-order issue for developed and developing countries alike. The cost of education makes out a large and growing share of GDP in most developed societies (OECD, 2020) while rising income differentials between educated and uneducated workers have spawned a large body of research on possible determinants such as trade and automation (Autor, 2019). At the same time, a large and growing literature has documented heterogeneity in returns to education both across individuals and across fields of study (see literature below). Presumably, a social planner would like to increase capacity in fields where returns of marginal applicants are high. However, with oversubscribed programs, admitting an applicant to one program will also 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 one program can impose externalities on other programs. While important for policy evaluation, these substitution effects are not reflected in regular estimates of returns to education and are empirically almost completely unexplored.

This paper intends to bridge the gap between estimated returns to education and supply-side policy in a market with applicant substitution effects. With a centralized and strategy-proof allocation mechanism, applicant behavior is not a function of supply. With access to the mechanism, it is, therefore, possible to clear markets for counterfactual supply while holding demand (ie. applications) constant. To perform such an analysis, 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. To quantify the substitution effects in terms of earnings I exploit the regression discontinuity design (RDD) implied by the allocation mechanism. I embed the discontinuity design in a structural selection model which allows me to investigate the effect of the supply changes on the full distribution of predicted earnings among all applicants. 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, implying 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 estimated earnings I show that flows that occur along the margin of the program where I change the supply account for 60 percent of the variation in aggregate earnings across experiments. In other words, by ignoring spillovers, 40 percent of the variation is missed. While the sign and magnitude of the externalities vary between programs, I find that for almost 70 percent of programs, flows along other margins create a positive externality. In other words, if a first-order 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 marginal gains vary in total gains from expanding supply once substitution is taken into account. Higher cutoffs in terms of grade-point averages (GPA) are generally associated with larger substitution externalities. With appropriate caveats, this suggests the following rules of thumb for policy-makers; If a program has a positive estimated return, the aggregate returns are most likely underestimated, and the more oversubscribed the program is, the larger this underestimation will be.

The findings of this paper imply that when a policymaker implements targeted supply-side policies, applicant sorting and rationing have important effects on aggregate outcomes. These results are developed by exploiting the centralized nature of college admission in Denmark. However, the qualitative insights are 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 markets.

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)) 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 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 model can replicate the corresponding Wald estimate of gains for marginal applicants 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 allocation mechanism clear the market. I describe the results in terms of applicant flows and earnings using variance decomposition and estimates of marginal effects of programs expansions in terms of first-order gains (ie. on the margins 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 size of supply changes. 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 approach the question of a trade-off between growth and inequality. I show that increases in aggregate earnings are associated with larger inequality. This trade-off is however not present within all fields, and policy-makers might therefore avoid the trade-off for certain supply-side policies. However, the estimated elasticity implies that supply policies that increase aggregate earnings will, in general, create gains among those with already high gains, which must be factored into the decision process if the policy-maker is inequality averse.

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). The largest quotas within Danish programs typically use high school GPA as a common tiebreaker. These papers exploit this feature and use cutoffs in terms of GPA to construct a fuzzy RDD. I add to these papers in a number of ways; I contextualize the RDD in a matching framework and show that the LATE is the policy relevant treatment effect for supply changes. Additionally, I exploit the full data and allocation rules of the Danish admittance system to obtain stronger first stages and estimate payoffs in programs with other tie-breakers than GPA.

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 matching mechanisms and show that the Student Proposing Deferred Acceptance mechanism can be represented by in a framework where cutoff structure plays the role of prices in clearing demand and supply that I utilize this insight in the theoretical part to clarify the policy effect of interest and phrase the substitution effects as pecuniary externalities where cutoffs play the role of prices.[1] Abdulkadiroğlu et al. (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. The authors show that this form of variation can be used to condition on the unobserved type and investigate the effect of attending charter schools. Abdulkadiroglu et al. (forthcomming) 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. By contrast, I parametrize outcome equations but let complier behavior be completely determined by the mechanism and stated preferences motivated by the fact that truth-telling is a dominant strategy in Deferred Acceptance. Several papers exploit the large market properties to estimate preferences in the absence of truth-telling properties, see Abdulkadiroğlu, Agarwal and

[1]The Danish allocation mechanism is a modified DA with the limitation that applicants can rank 8 programs rathers the submit a full ranking. I discuss this facet in Section 3.

Pathak (2017); Agarwal and Somaini (2018); Fack, Grenet and He (2019); Kapor, Neilson and Zimmerman (2020). These papers use the estimated preferences to investigate social welfare in mechanisms. Contrary to these papers I take preferences as truthful and policy-invariant, thereby holding demand constant while changing the supply.[2] Rather than assuming a specific structure on applicant utility, my social welfare analysis will be in terms of earnings, which is uncontroversial in the social welfare literature (Sandmo, 2015).

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. My paper differs from theirs in that I do not need a structural model of application behavior and counterfactual simulations are therefore less dependent on modeling assumptions. Altonji, Huang and Taber (2015) investigate consequences of compositional changes in the student body for those remaining in programs with outflows of applicants due to supply changes. I abstract from such compositional effects (such as peer effects) and instead focus on rationed programs with both inflow and outflow of applicants.

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 estimates preferences, I avoid this structural modeling by observing rank-ordered lists from a robust mechanism. I can therefore compute the full market equilibrium for tertiary education for all applicants in the market for Danish tertiary education without restrictive assumptions on preferences. Finally, as I estimate earnings equations, I can link substitution effects directly to traditional treatment effects type estimates of returns to fields and thereby price the externalities not commonly captured in program evaluations.[3] 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.[4] 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.

[2]For identification and interpretation, such utility models typically assume linear utility and normalizing the parameter on travel distance to -1. Piacquadio (2020) argues that this normalization implicitly favors applicants who prioritize substitutes to distance such as school quality. Interpreting aggregated utilities under such normalizations therefore imply very unusual social welfare functions. The mechanism that I simulate is in principle strategy-proof, but applicants are limited to filing 8 priorities which they generally do not. This suggests the presence of application costs. I discuss the implication of assuming truth-telling in the conclusion.

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

[4]Feller et al. (2016) in the same context find differential effects depending on the counterfactual allocation.

Except for Kline and Walters (2016) to my knowledge no other paper explicitly investigates the role of rationed substitutes for returns to supply-side policies in education. In estimating and predicting counterfactual outcomes, this paper draws inspiration from Abdulkadiroğlu et al. (2020) where the authors investigate the role of value-added for high school choice in New York City.

The paper proceeds as follows. Section 2 presents a framework for understanding the role of supply changes in a centralized assignment mechanism and how supply-side policies might be evaluated in the presence of substitution effects. Section 3 describes the institutional context of Danish tertiary education and a brief example of a simulation within the mechanism. Section 4 presents the econometric methods for estimating returns to program admission and Section 5 contains the sample definitions and report summary statistics. Section 6 presents the results of the estimations and in Section 7 I present the results of simulating market-clearing under counter-factual supply in terms of applicant flows and earnings. Section 8 concludes.

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 similar to the one presented by Azevedo and Leshno (2016). I assume a continuum economy with 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. Applicant types are defined by $\theta = (\succ^{\theta}, e^{\theta}, Y_{\theta}, X_{\theta})$, where $\succ^{\theta}$ is a strict preference ordering of programs for student type θ 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 if admitted to each program.[5] Finally, X_{θ} is a set of observable characteristics. Let η be the probability measure over the set of all types. Azevedo and Leshno (2016) show that a stable matching in a continuum economy is unique and can be characterized by a supply and demand equality in terms of a vector of program cutoffs, $C \in (0,1)^P$. 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. This formulation of demand closely resembles the multinomial choice-theoretic framework outlined by Heckman, Urzua and Vytlacil (2008) where the selection equation can be formulated in terms of the best alternative of those that the individual can afford. The cutoffs can in turn be thought of as prices clearing supply and demand. As the mechanism is Deferred Acceptance (DA), the dominant strategy of applicants is to report their true preferences regardless of other applicants or supply (Abdulkadiroğlu and Sönmez, 2003). From this follows that demand can be assumed constant and as the mechanism is unchanged, the market can be fully cleared for any given supply.[6]

[5] Contrary to the empirical context of Danish tertiary admission I assume away quotas. This implies that each program is represented by a single vector of eligibility scores and a scalar supply. In the simulations, i do not impose this restriction.

[6] Holding demand constant requires the possibility of submitting a full rank-ordered list and the absence of application costs (Fack, Grenet and He, 2019). I maintain these assumptions throughout this paper and discuss the implications in Section 8.

Assume that program p performs a marginal expansion of capacity of dS_p. Assuming that p is oversubscribed, the change in number of admitted to p equals the change in capacity. The stable matching using the new vector of supplies is given by C'. 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 the marginal applicants are taken from other programs and that these applicants are heterogeneous in type. The change in aggregate outcomes for the complier group to an expansion of supply in p is given by:

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

To evaluate the effect of a program expansion one needs an estimate of returns to program admission. The effect of interest can be framed in terms of the Local Average Treatment Effect (LATE) framework established by Imbens and Angrist (1994). Let crossing the cutoff be an instrument for admission. In this framework, a LATE can be estimated under the three conditions; Independence between the instrument and potential outcomes, an exclusion restriction that requires the effect to the instrument to only run through the program offer and a monotonicity condition which requires that all applicants act upon the instrument in (weakly) the same direction. These assumptions are plausibly fulfilled in a centralized DA-mechanism using a Fuzzy Regression Discontinuity Design (RDD) with waiting lists as running variables and a binary instrument for crossing the cutoff. Infinitely close to the program cutoff, an offer is random, ensuring that independence of potential outcomes is satisfied. As the instrument only reflects a decision rule of the algorithm there is no plausible way that it might affect outcomes in other ways than through program offer making the exclusion restriction a plausible assumption.

The last condition is monotonicity. Following Imbens and Angrist (1994), applicants can be separated into four sets depending on their response to the instrument; never-takers, always-takers, defiers, and compliers. In the DA mechanism, never-takers are those who would not be admitted regardless of crossing the cutoff or not. This set consists of two subsets: Those who do not apply to the program in question and those who are 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. Compliers are those who take up the offer if they cross the cutoff. Those are the applicants that do not get admitted to programs that they prefer over the program in question. This group is exactly equivalent to the set defined in Equation

(1).[7] By the same logic, note that Equation (2) constitute the aggregate gain in outcomes for the group of compliers. Therefore, dividing Equation (2) with (1) gives the Wald estimator of the LATE associated with a marginal increase in program p:

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

Equation (3) shows that the IV estimator will readily identify the treatment effect on compliers without observing types of applicants. In other words, even without full rank-ordered lists and knowledge of alternatives, the IV estimate identifies the effect of admission on the margin. The irrelevance of knowledge of the margin of choice 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.[8]

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.

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

Thus, in the context of supply changes, the LATE is what Heckman and Vytlacil (2001) refer to as the policy relevant treatment effect. H owever, 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.[9] The link between changed supply in one program and changes in the entire vector of cutoffs reflect the substitution of applicants and is equivalent to a pecuniary externality in a standard supply and demand framework. For the social planner, there is no a-priori reason to give preference to the outcome of those newly admitted to program p versus those newly admitted to program p'. 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

[7] 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:

$$\begin{aligned} Pr\left(D_i(C') = p, D_i(C) \neq p\right) &= \sum_{p' \neq p} Pr(D_i(C') = p, D_i(C) = p') \\ &= \sum_{p' \neq p} \int_\theta \mathbf{1}\{\theta : D^\theta(C') = p, D^\theta(C) = p'\} d\eta(\theta). \end{aligned}$$

[8] 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 the 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 the compliers on the margin.

[9] In a DA mechanism an expansion of p will weakly lower cutoffs in all programs, formally $\frac{dC_{p'}}{dS_p} \leq 0$. This is a reflection of the DA mechanism being strategy-proof. When supply is increased applicants can only do (weakly) better in terms of admittance.

gains:

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

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 cutoff and estimated LATEs for all programs where $\frac{dC_{p'}}{dS_p} \neq 0$ making. That informational requirements are large 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 complier groups are 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 of 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. In that case, one would need an estimate of the treatment effect for a different population.[10]

To handle these issues I instead utilize application and capacity data and knowledge of the assignment algorithm to identify complier groups of policy experiments directly and assume structure on the potential outcomes. While the outcomes have so far been left unspecified I now 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.[11] The aggregated effects can thus be written as

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

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,

[10] In the terminology of Heckman and Vytlacil (2001) the LATE might not be the policy-relevant treatment effect for large supply changes.

[11] This characterization is only appropriate for outcomes taking only non-negative values such as income.

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.[12] 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).

While Equation (7) shows one way of summarizing the effects of changing supply, as a social welfare measure it implicitly assumes indifference to economic inequality. To investigate a possible trade-off between inequality and economic growth I calculate the mean log deviations (MLD) inequality index on the baseline distribution of the predicted outcomes and the counterfactual distribution under changed supply. The MLD index is defined over a population vector of outcomes, $Y = \{Y_i\}$:

$$MLD(Y) = \frac{1}{N} \sum_{i=1}^{N} \ln \left(\frac{\overline{Y}}{Y_i} \right) \tag{8}$$

where $\overline{Y}$ is the average of Y. Correlating this measure against Δ^T will be informative on the existence of a trade-off between inequality and growth in higher education.

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 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 enrollment mostly occurs in the fall semester. Tertiary programs generally require a high school diploma and the programs can set additional requirements for admission. Usually, these requirements are based on having completed specific courses in high schools or having achieved a specific grade in a course. These requirements vary from program to program and across time. In general, programs set their own capacities though some programs with a high unemployment rate 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 70,000 and 90,000 thousand applicants to around 800 programs. When the number of applicants outnumbers the number of slots, applicants are generally admitted in two quotas. Quota 1 (henceforth Q1) admits applicants according to a GPA from secondary school. The same applicant will therefore enter with the same GPA in the rankings in multiple programs. Only applicants with a GPA or foreign equivalent can be

[12] 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 further in the conclusion.

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 (Q2), in which the criteria for choosing applicants are chosen by the educational institution under some 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.[13]

Each applicant can apply for eight programs. Under each program, the applicant can signal whether they want to be tried in Q2. If so, the applicant most often 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.[14] 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. 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.[15] In the mechanism, along with this modified rank-ordered list, each applicant is observed in each applicable quota with an eligibility score.

Program admission offices observe the program-specific application and not the remaining programs on the rank-ordered list of the applicant making 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 algorithm, programs are allowed to increase but not decrease their capacity and the mechanism is run again. The data I have access to is from the final allocation. The allocation mechanism is a modified Student Proposing Deferred Acceptance (abbreviation: DA, Gale and Shapley (1962); Abdulkadiroğlu and Sönmez (2003)).[16] In principle the DA is strategy-proof. However, in practice, the 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.[17] 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

[13] As this ranking is strict there is no need for a tie-breaker.

[14] 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.

[15] The order of these applications are determined by a numeric quota identifier, and the applications are ranked in ascending order of this number. Q1 will most often have numbers in the 10s, Q2 in the 20, Q1-standby in the 60s, and Q2-standby in the 70s. Some institutions have multiple groups within both quota 1 and quota 2.

[16] 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.

[17] 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. Thus, an applicant may receive a standby offer, resulting in non-assignment, while she would have been admitted to another program, had she not applied for a standby slot.

formalized.

3.1 Simulating the assignment mechanism

With knowledge of the algorithm and data on rank-ordered lists, eligibility scores, and quota capacities I construct a version of the entire mechanism for allocation of tertiary education in Denmark and I manage to allocate over 98 percent of applicants correctly in numerous years. The simulation model allows me to 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, which the program at the University of Copenhagen is a part of. 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 group of medium-cycle health programs, which mostly consist of Nursing programs. In other words, a change in the supply of one program creates substitution effects through a host of other 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; The flow matrix following an increase in capacity is not the transpose of the flow matrix following a decrease. In other words, substitution patterns are complex and complier shares from a contraction do not equal the complier shares of an expansion. 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. Fortunately, the simulation model keeps track of these flows regardless of the size of supply changes and without imposing structure on applicant preferences. The flows are completely determined by the applicant data and accurately reflect what would have occurred had the supply been changed while holding demand constant.[18]

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

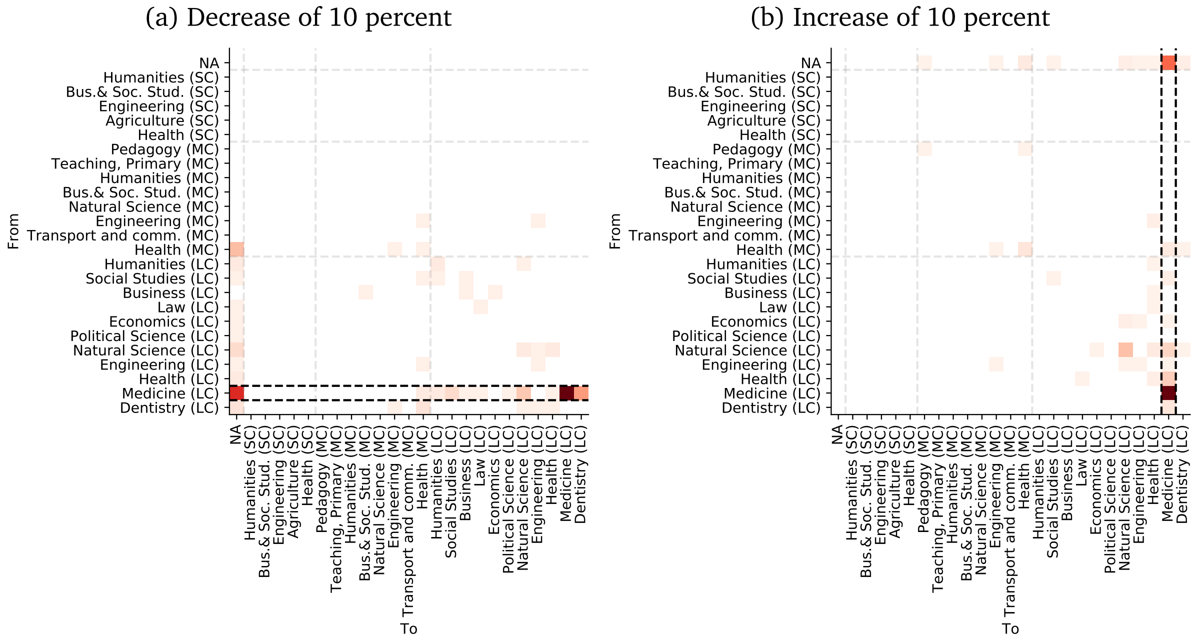


Figure 1: Applicant flows following a change in capacity for Medicine at KU

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 lenght 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.

4 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:

$$\log(Y_{ip}) = y_{ip} = X_i\beta_p + \varepsilon_{ip}, \tag{9}$$

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 et al. (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 those applicants admitted to program p. Taking the expectation of equation (9) for those admitted yields:

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

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 unobserved component is independent of observable characteristics $E[\varepsilon_{ip}|X_i, i \in p] = E[\varepsilon_{ip}] = 0$:

$$E[y_{ip}|X_i, i \in p] = X_i\beta_p \tag{11}$$

Under this assumption, which I refer to as selection on observables, 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 (11). 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 regression discontinuity design. As previously described, the Danish admission systems work with program-quotas instead of programs. Mechanically, each quota is given a numeric identifier and submitted rank-ordered lists are expanded lexicographically by priority and quota identifier. This implies that the cutoff representation is in terms of quotas indexed by q instead of programs, with a many-to-one mapping between quota and program.

I exploit discontinuities around the cutoffs in different quotas for each program. As described in Section 3, most programs have two quotas, Quota 1 where applicants are ordered according to high school GPA, and Quota 2 where programs select on other criteria. In the data, I observe the ranking of each applicant in each quota which facilitates a Fuzzy Regression Discontinuity Design. 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 an indicator variable 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. Conditional on the running variable, e_{iq}, Z_i^{pq} is a valid instrument for D_i^p. I specify the following selection equation:

$$D_i^p = \mathbf{1}\left\{\gamma Z_i^{pq} + f\left(e_{iq}\right) + \epsilon_{iq} > 0\right\}, \tag{12}$$

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 (10) becomes a standard probit selection model:

$$E[y_{ip}|X_i, i \in p] = X_i\beta_p + \psi_p\lambda\left(Z_i^p + f\left(e_{iq}\right)\right), \tag{13}$$

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 (10). To include the instrument, Z_i^p in equation (13) and avoid identifying off functional form alone, there need to be applicants on both sides of the cutoff. This implies that the control function method is only feasible for over-subscribed programs or in cases where applicants are restricted from admission into higher-prioritized programs.[19]

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.[20] Secondly, applicants above the threshold might also be eligible to a higher-ranked program and thus get admitted there. As the endogenous variable is whether the applicant gets an offer and we know the algorithm, the residual in the selection equation can be given a specific interpretation. In the presence of applications costs applying to multiple quotas in the same program will, all else equal, reflect that an applicant has a higher motivation for being admitted to program p. In the selection equation, the application to multiple quotas will be reflected in a high ϵ_{iq}. Given that the assignment mechanism is strategy-proof, it is natural to assume that an applicant has a higher motivation for a higher-ranked program than a lower-ranked program. Thus if an applicant crossed the cutoff but does not get an offer, this is due to her getting a better offer. In the selection equation, this will be reflected in a low ϵ_{iq}. The specific formulation of the selection thus allows for an interpretation of ϵ_{iq} as a measure of motivation.[21] A positive estimate of ψ_p is a sign that the higher motivation, as elicited in the selection model, is associated with higher outcomes. Additionally, a change in the estimates of β is informative on how motivation covary with observable attributes. Standard error are obtained from the two-step variance estimator derived by Heckman (1979).

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 Appendix A.1.[22]

4.1 Joint distribution of parameters, shrinkage, and prediction

Under the appropriate assumptions, the estimates of β_p are unbiased but noisy measures of the true parameters. This noise is especially an issue in the case of small programs with correspondingly small samples. As the estimates are needed for predictions I follow Abdulkadiroğlu et al. (2020) and apply an empirical Bayes shrinkage estimator to the estimates, thereby reducing the

[19]As the assignment function is known, there is no need to include X_i in the selection equation.

[20]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.

[21]If crossing the cutoff of a higher-ranked program is used as an instrument, the interpretation of ϵ_{iq} reverses.

[22]Notice that the equivalence result only holds under two-sided non-compliance. If a program does not admit anybody who does not cross the threshold in a quota there are no always-takers and the comparison breaks down.

mean squared error of the predicted outcomes (Robbins, 1992; Morris, 1983).[23] I assume the following hierarchical model:

$$\hat{\beta}_p|\beta_p \sim \mathcal{N}(\beta_p, \Omega_p) \tag{14}$$

$$\beta_p \sim \mathcal{N}(\mu_\beta, \Sigma_\beta) \tag{15}$$

Estimates of μ_β and Σ_β are obtained through maximum likelihood estimation of (14) and (15), 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 hyper parameters, 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.2.

The predicted outcomes in the simulation sample is calculated as:

$$\hat{Y}_{ip} = \exp\left(X_i\beta_p^*\right) \tag{16}$$

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.

Prioritizing prediction method The methods outlined above provide six sets of estimates for a given program; program and field-specific estimates under conditional independence and four RDD-methods, the combination of the two quotas and whether admission is instrumented by crossing the program cutoff or by crossing the cutoff of a higher prioritized program. As I only need one set of estimates for each program, the methods must be prioritized. The priority structure is as follows. Firstly I use control function estimates, preferring Quota 1 estimates to Quota 2 estimates and program cutoffs to cutoffs at higher prioritized programs. Missing estimates are then filled with program and field level OLS estimates in that order. The remaining missing outcomes are imputed using the μ_β from the control function using own-program cutoffs in Quota 1. Using this protocol I can impute all outcomes in the simulation sample and therefore calculate the full income distribution under counterfactual supply of slots.

[23]Empirical Bayes methods have become popular in educational contexts where prediction is important and data is noisy, see Angrist et al. (2017); Chetty, Friedman and Rockoff (2014) for examples of usage.

5 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. This data is sufficient to simulate the clearing of the mechanism. Data is available from 1992, but due to lack of data simulation with a satisfactory level of precision is possible from 2016. 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. For those admitted in the clearing in July, I observe enrollment in October. 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. Each program present in the simulation year, 2016, is mapped to previous programs and due to the opening of new programs and each program will differ in how many previous years are available for estimating the outcome equations.

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 gains. The 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 waiting list 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.

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 raising subjects to a higher level.[24] For foreigners, the GPA conversion follows the guidelines of the Ministry of Higher Education, though as this is done by the program administrators these

[24] Applicants applying within two years of high school graduation can multiply their GPA with 1.08.

Table 1: Descriptive statistics on samples

		Estimation sample	Program cutoff Q1	Q2	Higher priority cutoff Q1	Q2	Simulation sample
GPA	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.

GPAs are subject to error.[25] The corresponding GPA-ranks are imputed by a Random Forrest regression trained on those applicants where both GPA-rank and the alternative GPA is reported. For the subset of applicants where no variant of GPA is observed, I set the value to 0.5.

5.1 Descriptive statistics

Table 1 present descriptive statistics for the samples. The unit 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.[26]

The following 4 columns display the corresponding statistics for the samples used for sampling correction. 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 contains *all* applicants in 2016 which is close to 83 thousand applicants. 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, where the estimation samples drop applicants where I do not observe a realized positive outcome.

[25]For instance, the same applicant can be observed with multiple values of the alternative GPA across applications.
[26]The average exchange rate in 2015 was 6.72 DKK/USD.

Table 2: F-statistics for program specific IV

Method	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.

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.[27] Figures 2a to 2d show balancing of predetermined variables and I find no evidence of effects of sorting around the cutoff. [28]

The first stage, displayed in Figure 2e, 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).[29] 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 4.1. 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.

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

[28]While threshold-crossing effects are valid, the slopes of the polynomials should be interpreted with care as they are affected by composition effects due to the stacking of programs.

[29]The fuzzy RD is estimated with the rdrobust package in Stata.

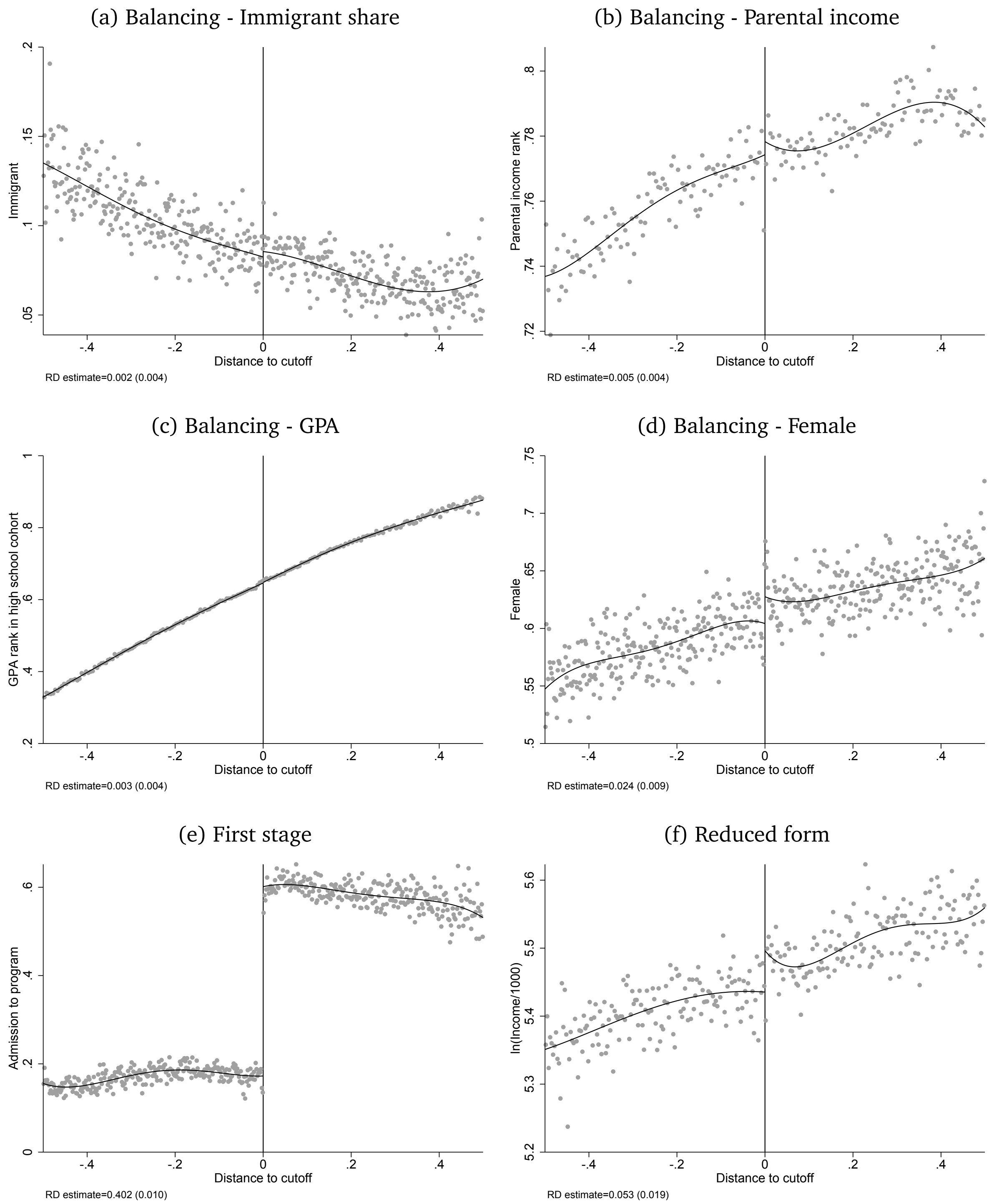


Figure 2: Regression discontinuity plots

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 care. 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.

The distribution of estimated LATEs I now proceed to present the estimates of program specific LATEs from a fuzzy regression discontinuity design. This merely serves as a sanity check and are not the estimates used for predicting outcomes in the simulation sample. As mentioned above, IV estimates provide consistent but noisy measures of the relevant LATE for the first-order effect of a program expansion. To investigate which programs might yield high returns and account for the noise I shrink the IV estimates toward zero using a standard Empirical Bayes approach.[30] Figure 3 shows the programs with the highest returns according to the posterior LATE. The blue squares show the raw IV estimates with corresponding 95 percent confidence intervals. The black circles shows the shrunken estimates with 95 percent confidence intervals calculated using the method developed by Armstrong, Kolesár and Plagborg-Møller (2020). Appendix Figure B.1 contains estimates for all programs where IV is feasible. The average of the posterior returns shrunken to zero is 0.0036, compared to an unweighted average of 0.083. However, programs such as Medicine at University of Aarhus (AU) and Political Science at University of Copenhagen (KU) yield returns of around 10 percent.[31] As is evident both in Figure 3 and Appendix Figure B.1, a large share of program-specific IV estimates of LATEs are very imprecisely estimated and the confidence intervals often cross zero. This makes it problematic in practice to base policy on raw IV estimates. On the other hand, the distribution of the posterior LATEs is quite narrow and the confidence intervals are large making it difficult to conclude which programs yield high returns.

While the program-specific LATEs (and their posteriors) provide insights into which programs policymakers might be interested in expanding, the estimates are based on samples that are stacked across multiple years and might differ substantially from the complier groups of a simulated supply change in a single year. Further, even within the estimation sample, the composition of compliers across programs is likely to differ substantially and it is not obvious that shrinking towards zero (or average LATE) is the correct approach in terms of predicting counterfactual outcomes. To address these worries, the outcomes used in the simulations of supply changes are calculated using the posterior control function parameters and the actual composition of the simulation sample. The LATEs presented here are, therefore, not directly comparable to the simulation results.

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 serve to bolster the credibility of the estimates of β obtained through structural modeling. Figure 4a plots the estimated program-specific LATEs obtained through the control function approach against the corresponding IV estimates using the variation from crossing own-program thresholds. 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

[30]This Empirical Bayes approach shrinks the LATEs as estimated by IV and is not to be confused with the shrinking of the parameters in the outcome equations outlined in Section 4.1. Shrinking towards zero differs from the approaches used by Chetty, Friedman and Rockoff (2014) as they shrink towards an average treatment effect identified through assuming conditional randomness of teacher allocation. In the present context I estimate LATEs through IV. In case of treatment heterogeneity and self-selection into program application, the average treatment effect is not identified. To weight the credibility of the IV estimates I therefore shrink the estimates towards zero.

[31]These are highly prestigious and oversubscribed programs and have existed throughout the time period for which ensures precision in estimation of the LATEs both in terms of sample size and mass around the cutoff.

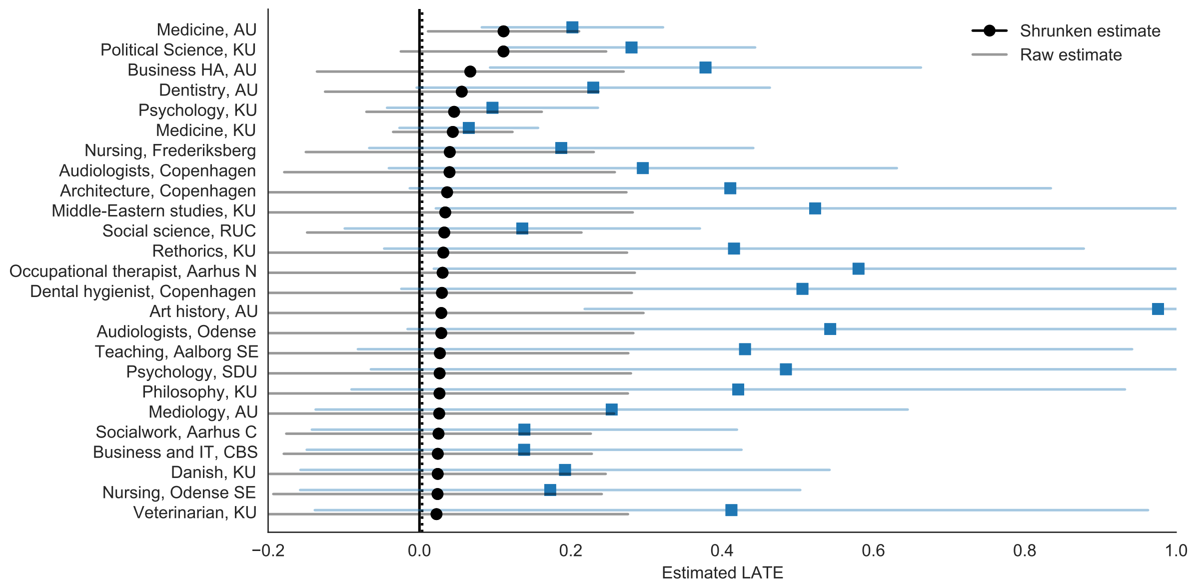


Figure 3: IV estimates of program LATE

The figure shows LATE estimates calculated from IV estimates of a standard fuzzy regression discontinuity design as well as the posteriors of an Empirical Bayes procedure where estimates are shrunken towards zero. The source of variation is crossing own program cutoff in quota 1. Estimates with an F-statistic lower than 10 have been excluded. The bars represent 95 percent confidence intervals. The confidence intervals of the IV estimates are based on robust standard errors and the confidence intervals for the posteriors are calculated using the methods developed by Armstrong, Kolesár and Plagborg-Møller (2020). The dotted lines represent the unweighted mean of program-specific LATEs.

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 4b. 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 4.1. These estimates should be interpreted as weighted means of parameters across programs but masking considerable heterogeneity across programs.

Without controlling for selection, the average marginal return to GPA rank is 0.12, implying that moving up a decile in the GPA distribution 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

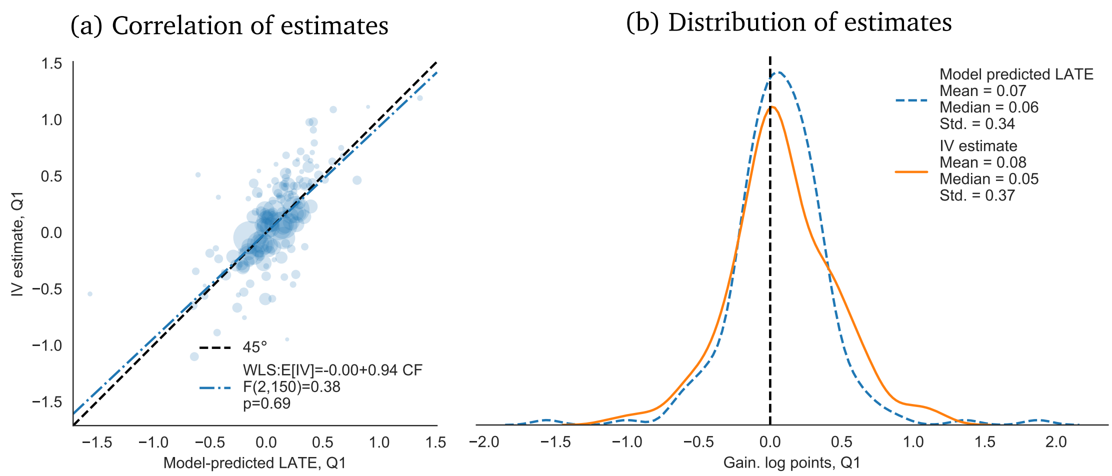


Figure 4: Model-based and IV estimates of LATE

The figure shows LATE estimates calculated from the selection correction model in Equation (13) and IV estimates of a standard fuzzy regression discontinuity design. The source of variation is crossing own program cutoff in quota 1. Figure 4a 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. Estimates with an F-statistic lower than 10 or a p-value larger than 0.7 have been excluded. Figure 4b shows the distribution of LATEs using IV and control functions.

(12) correlates positively with outcomes. As discussed in Section 4, 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 column (5) and (6). The coeffficient 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.[32]

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.

[32]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 (15) estimated from the program level estimates for the estimation methods outlined in Section 4. 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 $E\left[\frac{\partial y}{\partial GPA}\right] = \frac{1}{2}\beta_{GPA} + \frac{1}{3}\beta_{GPA^2}$.

Predicting returns in simulation sample For the simulation sample, containing all applications in 2016, the expected annual income is predicted based on the shrunken parameters, β_p^*. The models contain a linear time trend and I set the year to 2010. The models used for predicting incomes are hierarchical in the sense that the selection correction models are preferred to models based on conditional independence assumptions. The hierarchy and the share predicted by each method is presented in Table 4. 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 predicted incomes in the base level assignment in 2016 is displayed in Figure 5a. 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.

The distribution of expected gains of being admitted into Higher Education is shown in Figure 5b. These figures suggest that obtaining an admission is associated with an annual

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.

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. 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 5c.

7 Clearing markets under counterfactual supply

To investigate the importance of applicant substitution patterns in the admission system I simulate a range of experiments. 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 overflow of applicants from Quota 2, I disable the transfer 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. Each simulation returns a vector of allocations which is merged onto predicted incomes to form a counterfactual income distribution. 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.[33] 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 6a shows that on average when a counterfactual supply induces 10 applicants to change in or out of a program, this is associated with 15 applicants 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. These flows are completely determined in the data independently of the predicted outcomes.

[33] 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.

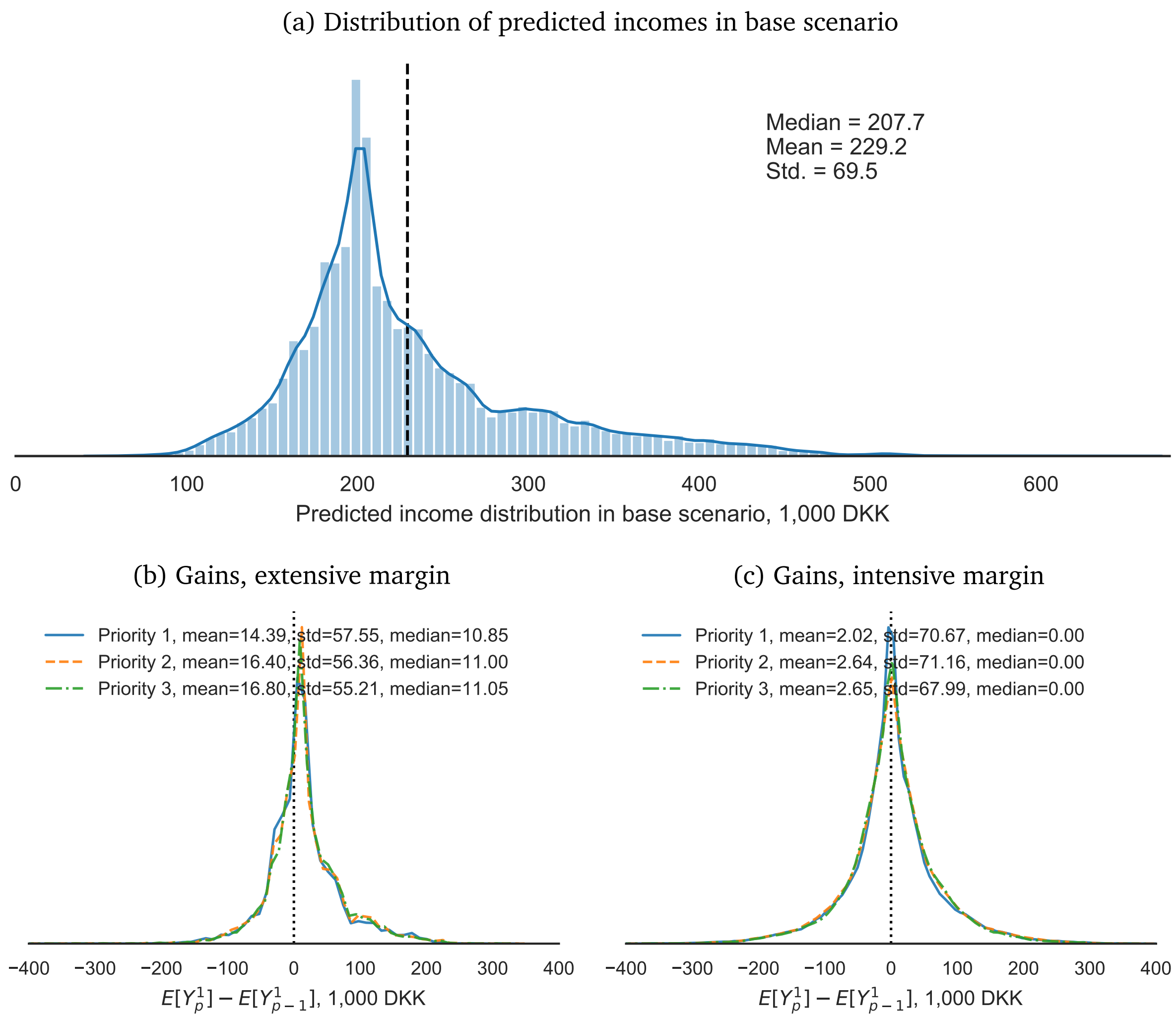


Figure 5: Predicted incomes

Figure 5a displays the distribution of predicted incomes in the simulation sample. The dashed lines show the points of truncation. Figure 5c shows the distribution of predicted gains from moving up on slot up a priority on a rank-order list, and Figure 5b shows the predicted gain from going from non-assigned to assigned. Incomes are in 1.000 DKK in 2010-prices.

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 total variance in the experiments. This implies that 39 percent of the actual vari-

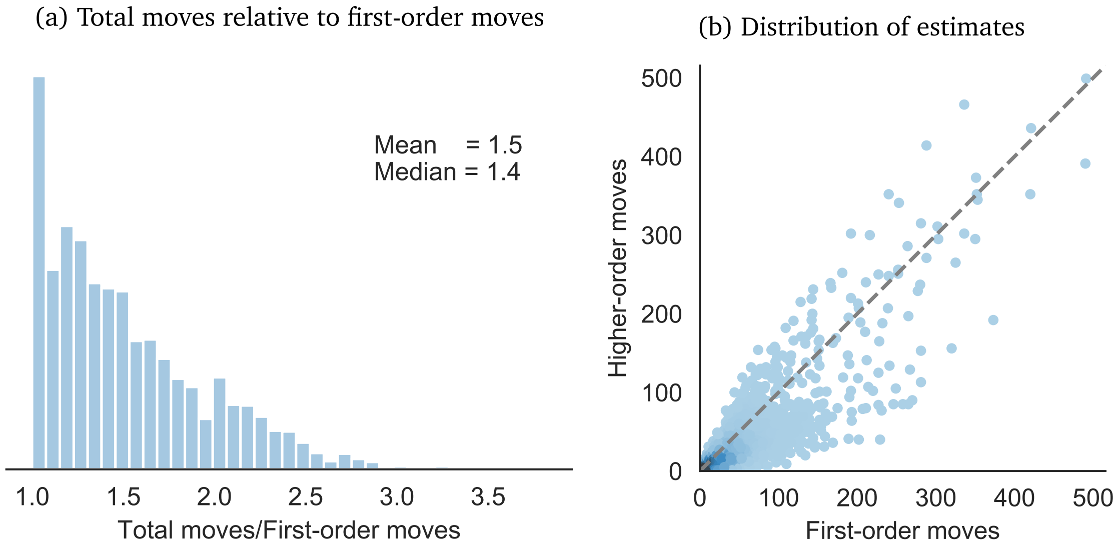


Figure 6: First-order and higher-order moves

Figure 6a plots the ratio of the total number of movers relative to the first order movers. The unit of analysis is a simulation. Figure 6b plots the number of higher-order moves against the number of first-order moves. Points are colored according to the density of points.

ation is missed by not taking higher-order effects into account. The corresponding share for higher-order effects is 9.3 percent and the remaining 30 percent is due to interaction of first- and higher-order effects. 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.

Program-level returns to adding slots The variance decomposition show that higher-order effects are important and positively correlated with first-order effects. However, the variance across experiments is a function of stacking programs of different sizes and large programs may therefore dominate the results. To investigate program level heterogeneity further, I compute implied program level marginal returns to increasing supply. For each program, I perform 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 I regress the first-order (F), higher-order (H), and the aggregate total (T) gains of the experiments on the number of slots changed:

$$\Delta_{pe}^{g} = \gamma_p^g (S_{pe} - S_{p0}) + \varepsilon_{pe}, \quad g \in \{F, H, T\}, \tag{17}$$

where Δ_{pe}^{g} 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_p^g, can be interpreted as the marginal gain from adding a slot in a given program. The usefulness of this measure depends on the underlying linearity

Table 5: Variance decomposition of earning gains in sumulations

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

The table displays the contribution to 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.

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 accommodate a sufficiently large program expansion, meaning that at some point the marginal gain from expansion might drop to zero. Secondly, if the composition of the complier groups, ie. those who take up an offer, depends on the magnitude of supply change, the marginal effect of supply changes might change as a function of changing complier composition. To gauge the validity of the slope as a metric, Figure 7a plots the R^2 from the regressions in Equation (17). 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. This 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. In conclusion, the estimated set of γ_p^g provides valuable information on the effects of changes in supply.

The distribution of γ_p^g is shown in Figure 7b. 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). Figure 7b suggests that there are large differences in the marginal effect of adding slots. 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

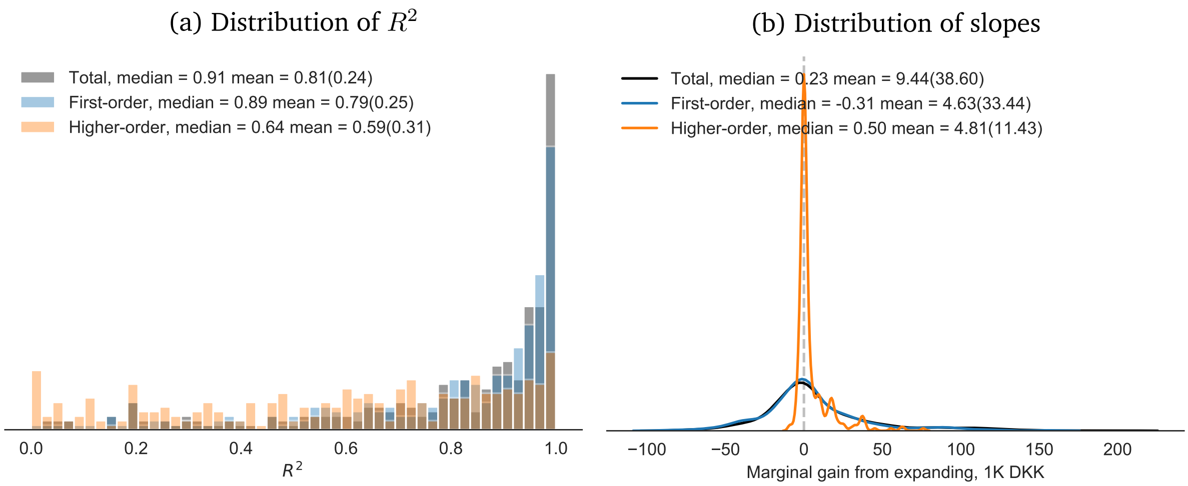


Figure 7: Marginal gain of changing supply

Figure 7a shows the distribution of R^2 from regressing aggregate gains on the number of slots as in Equation (17). Figure 7b shows the distribution of corresponding slopes, γ_p^g.

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 reverses once higher-order gains is 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 8 plots the higher-order gains against first-order gains. The dashed black line represents break-even points where the marginal first-order gain is neutralized by the higher-order losses. The programs are colored according to total gain, and every program to the right of the dashed line exhibits positive total returns. 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 induce few higher-order moves. 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 it similarity in content with the Aarhus program yield much larger higher order gains (I return to this case below). Thirdly, 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 mitigated by higher-order gains.

Table 6: Marginal gains of program expansion

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

The table provides descriptive statistics on the estimated slopes using the model in equation (17). Only non-zero slopes are included.

Summing up, I find that higher-order gains may attenuate, increase or reverse expected gains. The magnitudes and signs of the marginal effects of program expansion vary both within and between fields and program lengths. For research, this implies that even if policy-makers are provided with a well-estimated LATE to a specific program, taking clearing of the market into account can substantially change policy recommendations. However, the first-order and higher-order effect are positively correlated, implying that the sign of the first-order LATE is indicative of the direction of the total effect of a local supply change.

A case study of supply changes in Business programs To illustrate the mechanisms behind the finding above I briefly present a case study with similar business programs (In Danish: Erhvervsøkonomi, HA.), at Copenhagen Business School (CBS) and University of Aarhus (AU). Both programs are large in terms of intake and have medium-high GPA cutoffs, implying the presence of higher-order movers. Figure 9a show the simulation results for Business at CBS, the slopes reported in the previous figures are the linear approximation of the three series of gains. Figure 9b show the corresponding results for Business at the University of Aarhus. Both programs have a slope of the first order gains of 108.49 and 96.76, ie. a marginal gain of approximately 100 thousand DKK pr. additional slot. However, taking into account the higher-order moves, the marginal gain of expanding the Copenhagen program has a marginal gain of 147.63, an increase of 26 percent. In contrast, the Aarhus program has a marginal gain of 98.97, barely higher than the first order gain.

Part of the reason why these programs differ can be seen by inspecting the flows of applicants following a program expansion. Figure 9c and Figure 9d show the transition matrices on field level corresponding to a 10 percent increase in program capacity. The darker the square, the larger the number of movers. Ignoring the grouping of programs within fields, squares within the dashed lines roughly correspond to first-order moves. Comparing the two transition

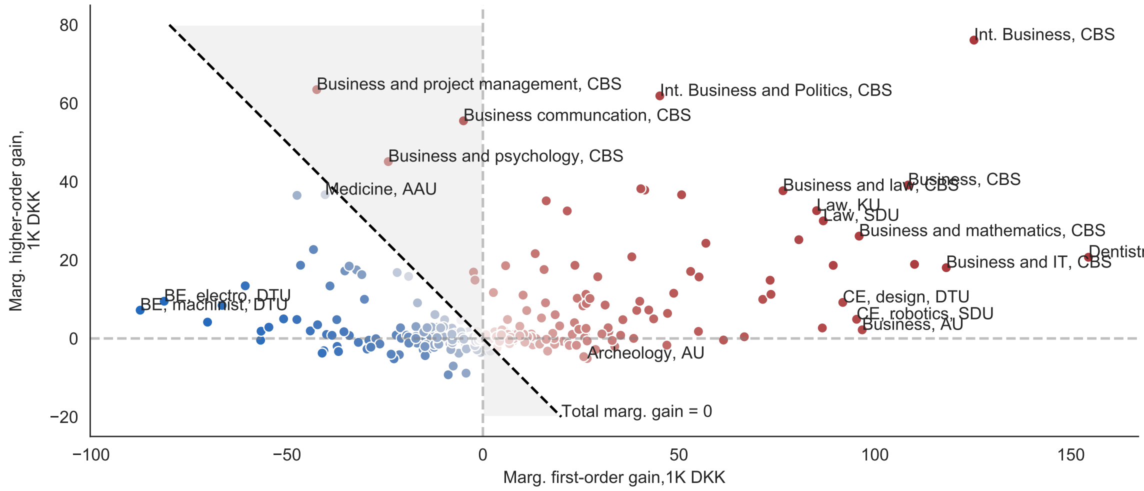


Figure 8: First-order and higher-order gains

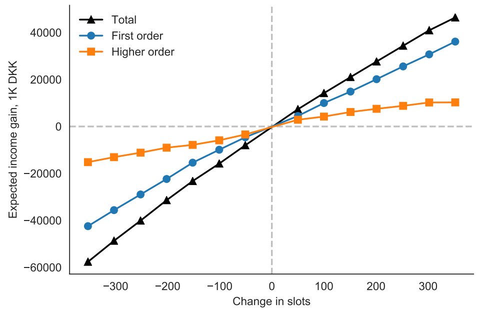
The figure plots estimated higher-order marginal gains against first-order marginal gains. The programs are colored according to their aggregate gains, which can be read off the distance to the dashed line. Selected programs are annotated, CE and BE are abbreviations for civil and bachelor engineering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

matrices a number of differences are apparent. Firstly, an expansion of the Copenhagen program primarily causes an inflow from other Business programs. In Aarhus, on the other hand, the primary inflow is along the extensive margin, which does not induce higher-order moves. Additionally, an expansion in Copenhagen draws in applicants from other long-cycle programs such as social science and humanities. The difference in the gains thus comes from applicants ranking other programs below the Copenhagen program, while applicants to the Aarhus program are more likely to be on the margin of non-admittance.

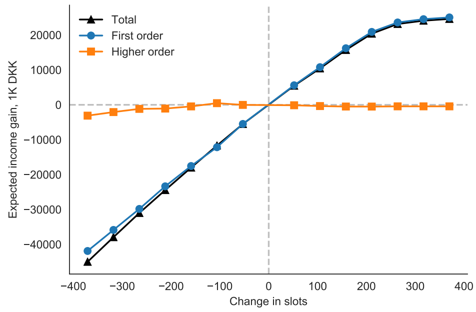
Cutoffs as measure of externality The substitution externalities are due to the rationing of programs. As such, one would expect that the size of the externality is increasing in the degree of rationing and thereby the value of the GPA-cutoff. Figure 10 plots ratio of the absolute value of the higher-order slope over the absolute value of the first-order slope against the GPA cutoff in the year of the simulation-sample.[34] The figure shows that the magnitude of the externality is approximately constant 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 is a good predictor of the magnitudes of substitution effects.

Inequality tradeoff A given aggregate gain from a supply change may be associated with different changes to inequality depending on the number of applicants affected and where they are positioned in the income distribution. Figure 11 plots (one minus) the Atkinson index

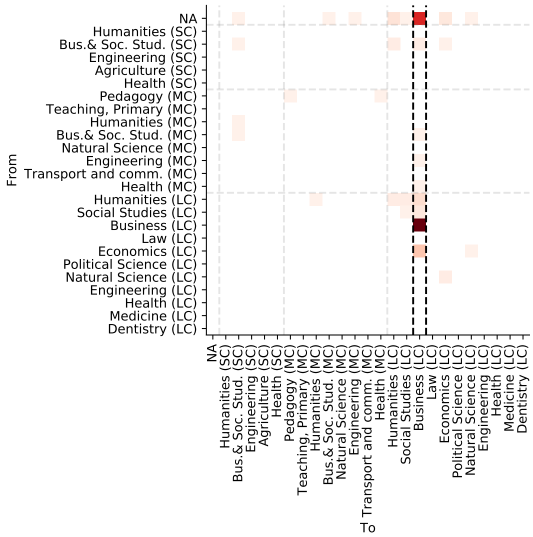
[34]Passed exams are graded on a scale between 2 and 12. In combination with bonus for multiple A-levels and ab early application bonus cutoffs can therefore take values between 2 and slightly over 12.



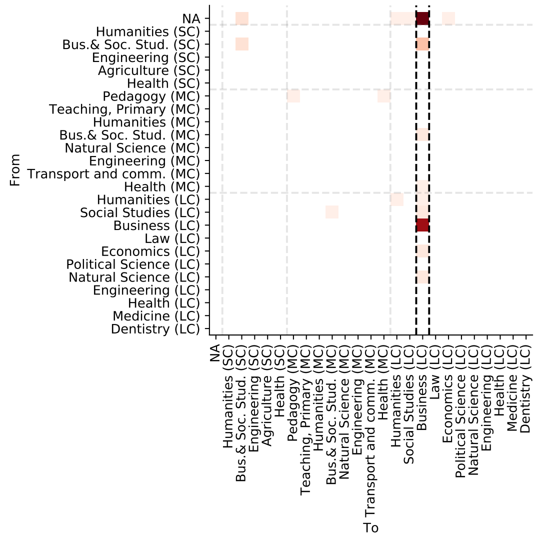
(a) Gains, Business, CBS



(b) Gains, Business, AU



(c) Transitions, Business, CBS



(d) Transitions, Business, AU

Figure 9: Business programs in Copenhagen (KU) and Aarhus (AU)

Figure 9a and Figure 9b show the simulation results for Business at Copenhagen Business School and University of Aarhus respectively. Figures 9c and Figure 9d show the flow of applicants resulting from a single simulation of ten percent in the respective programs. The column within the dashed line signifies the first-order moves.

against the aggregate gains. The higher the value on the y-axis the lower the inequality. The higher the value on the x-axis the higher the aggregate income gain. Thus, a policy maker who values both growth and equality will increase utility by moving towards north-east. The intersection between the vertical and horizontal lines indicates the baseline distribution. From the figure, an overall negative relationship between equality and growth is evident, indicating a growth-inequality trade-off of supply policies. However, the experiments do not all lie in the upper-left or the lower-right quadrant. In other words, the trade-off does not manifest itself for all policies.

To investigate heterogeneity in the growth-inequality trade-off I regress $log(MLD)$ on the log of the aggregate income, thus estimating the elasticity of the trade-off. As the trade-off should reflect the experiments rather than general program level differences I include program fixed effects. Table 7 presents the results. The first column shows that on average a 1 percent increase in aggregate income is associated with a 3.94 percent increase in inequality measured by the mean log deviation index. Column 2 interacts income with program length. The trade-off remains for long-cycle and short-cycle programs, but the sign switches for medium-cycle

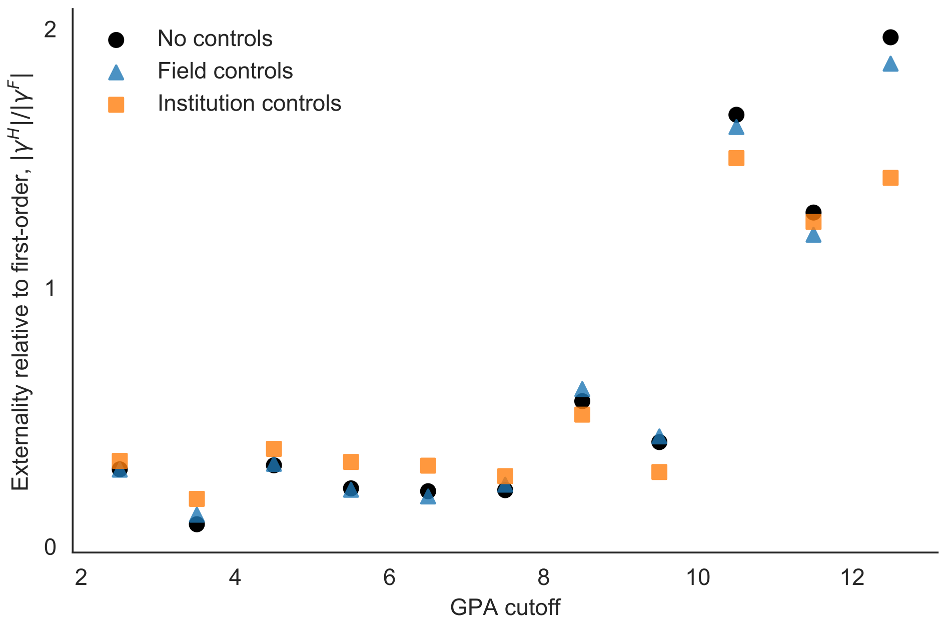


Figure 10: Substitution externality as a function of GPA cutoff

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 is regressed on field and institution dummies respectively. The plot is a binned scatterplot constructed from the binscatter STATA-package (Stepner, 2013).

programs. In other words, on average, the policymaker can both achieve higher incomes and less inequality by changing the supply in medium-cycle programs.

In Column 3 aggregate income is interacted with fields. The trade-off exists for Health, STEM, and social science and business but is not evident for humanities and Teaching. Notice, that teaching on average yields negative returns as evidenced in Table 6. In other words, the missing trade-off shows that it is possible to increase income and decrease inequality by decreasing the number of slots in teaching.[35] Column 4 interacts program length and field to further investigate the trade-off. Besides teaching, the largest elasticities are found for long-cycle health, humanities, and social science programs.

8 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

[35]However, other possible externalities are disregarded. Fewer teachers may lead to worse outcomes for students which is completely disregarded in the present analysis.

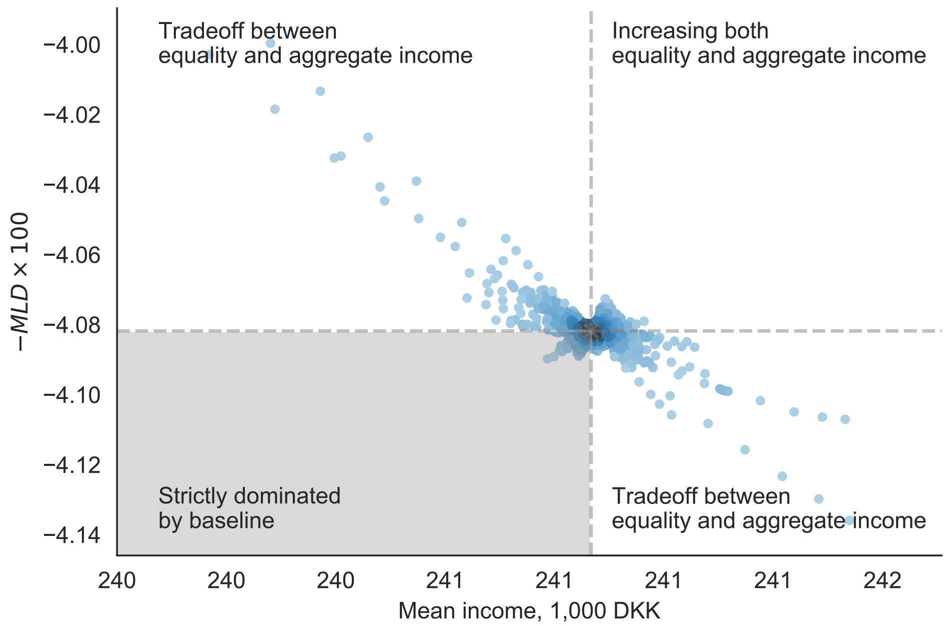


Figure 11: Inequality and mean gains

The figure presents the relationship between aggregate gains and inequality in predicted outcomes. To counter the issue of points overlapping they are colored according to the density of the points. The aggregate income is divided by the number of applicants.

Table 7: Trade-off between growth and inequality

	(1) ln(MLD)	(2) ln(MLD)	(3) ln(MLD)	(4) ln(MLD)
ln(Gain)	3.94 (0.17)			
LC $\times$ ln(Gain)		4.05 (0.00)		
MC $\times$ ln(Gain)		-1.07 (0.00)		
SC $\times$ ln(Gain)		0.81 (0.00)		
Health $\times$ ln(Gain)			2.17 (0.14)	
Humanities and communication $\times$ ln(Gain)			-2.12 (0.00)	
STEM $\times$ ln(Gain)			0.61 (0.07)	
Social science and business $\times$ ln(Gain)			4.40 (0.11)	
Teaching $\times$ ln(Gain)			-5.81 (0.00)	
LC $\times$ Health $\times$ ln(Gain)				2.25 (0.00)
LC $\times$ Humanities and communication $\times$ ln(Gain)				-2.12 (0.00)
LC $\times$ STEM $\times$ ln(Gain)				0.64 (0.00)
LC $\times$ Social science and business $\times$ ln(Gain)				4.47 (0.00)
MC $\times$ Health $\times$ ln(Gain)				-2.17 (0.00)
MC $\times$ STEM $\times$ ln(Gain)				0.38 (0.00)
MC $\times$ Social science and business $\times$ ln(Gain)				-0.72 (0.00)
MC $\times$ Teaching $\times$ ln(Gain)				-5.81 (0.00)
SC $\times$ STEM $\times$ ln(Gain)				1.58 (0.00)
SC $\times$ Social science and business $\times$ ln(Gain)				0.74 (0.00)
Cluster level	Length	Length	Length	Length
Fixed effects	Program	Program	Program	Program
R^2	0.73	0.75	0.80	0.81
N	3,570	3,570	3,570	3,570

The table displays the results of regressing the log of the MLD index on the log of aggregate gains thus providing elasticities of the trade-off. The level of analysis is a single supply change. Standard errors are shown in parentheses.

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 supply changes are 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 in the top of the list. A method to counter this would be to estimate preferences under the assumption outlined by Fack, Grenet and He (2019) and fill out rank-ordered lists. However, to do this 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 the full rank-ordered-list, the substitution chains would likely become longer. The present results are therefore most likely a lower bound on the magnitude of externalities.

In the analysis, I frame the results in terms of earnings and inequality. However, to fully account for market clearing in a cost-benefit analysis, more information and assumptions are needed. If programs are very heterogeneous in terms of costs, this would have to be accounted for in a full cost-benefit analysis. If one would want to undertake such an analysis, one needs a dynamic analysis of how applicants drop out and reapply to measure the accruing costs of provision as well as a mapping from incomes in a given year to life-time earnings trajectories and tax payments, see Hendren and Sprung-Keyser (2020) for an example of such an exercise. This is outside the scope of this paper. Additionally, I assume throughout that outcomes can be accurately described by a set of vectors which are held constant. However, for non-marginal and permanent changes in supply, general equilibrium effects become important and the link between programs and outcomes might change.

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 clearing.

References

Abdulkadiroğlu, Atila, and Tayfun Sönmez. 2003. "School choice: A mechanism design approach." *American economic review*, 93(3): 729–747.

Abdulkadiroğlu, Atila, Joshua D Angrist, Yusuke Narita, and Parag A Pathak. 2017. "Research design meets market design: Using centralized assignment for impact evaluation." *Econometrica*, 85(5): 1373–1432.

Abdulkadiroglu, Atila, Joshua D Angrist, Yusuke Narita, and Parag A Pathak. forthcomming. "Breaking ties: Regression discontinuity design meets market design." *Econometrica*.

Abdulkadiroğlu, Atila, Nikhil Agarwal, and Parag 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, Atila, Parag A. Pathak, Jonathan Schellenberg, and Christopher R. Walters. 2020. "Do Parents Value School Effectiveness?" *American Economic Review*, 110(5): 1502–39.

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

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

Agarwal, Nikhil, and Paulo Somaini. 2018. "Demand analysis using strategic reports: An application to a school choice mechanism." *Econometrica*, 86(2): 391–444.

Altonji, Joseph G, Ching-I Huang, and Christopher R Taber. 2015. "Estimating the cream skimming effect of school choice." *Journal of Political Economy*, 123(2): 266–324.

Altonji, Joseph G, Peter Arcidiacono, and Arnaud 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, 305–396. Elsevier.

Andersen, Mikael Bjørk, Sebastian Hørlück, and Christian Lund Sørensen. 2020. "Effekt af optag på videregående uddannelser for marginale studerende." Working Paper.

Andrews, Rodney J, Jing Li, and Michael F Lovenheim. 2016. "Quantile treatment effects of college quality on earnings." *Journal of Human Resources*, 51(1): 200–238.

Angrist, Joshua D, Peter D Hull, Parag A Pathak, and Christopher R Walters. 2017. "Leveraging lotteries for school value-added: Testing and estimation." *The Quarterly Journal of Economics*, 132(2): 871–919.

Arcidiacono, Peter, Esteban M. Aucejo, and V. Joseph Hotz. 2016. "University Differences in the Graduation of Minorities in STEM Fields: Evidence from California." *American Economic Review*, 106(3): 525–62.

Armstrong, Timothy B, Michal Kolesár, and Mikkel Plagborg-Møller. 2020. "Robust empirical bayes confidence intervals." *arXiv preprint arXiv:2004.03448*.

Autor, David H. 2019. "Work of the Past, Work of the Future." *AEA Papers and Proceedings*, 109: 1–32.

Azevedo, Eduardo M, and Jacob D Leshno. 2016. "A supply and demand framework for two-sided matching markets." *Journal of Political Economy*, 124(5): 1235–1268.

Brinch, Christian N, Magne Mogstad, and Matthew Wiswall. 2017. "Beyond LATE with a discrete instrument." *Journal of Political Economy*, 125(4): 985–1039.

Bucarey, Alonso. 2018. "Who pays for free college? Crowding out on campus." *Job market paper*.

Calsamiglia, Caterina, Guillaume Haeringer, and Flip Klijn. 2010. "Constrained school choice: An experimental study." *American Economic Review*, 100(4): 1860–74.

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014. "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates." *American Economic Review*, 104(9): 2593–2632.

Dale, Stacy Berg, and Alan 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, Moira, Mathias Fjællegaard Jensen, and Daniel Le Maire. 2020. "University admission and preferred field of study." Working Paper.

Fack, Gabrielle, Julien Grenet, and Yinghua He. 2019. "Beyond truth-telling: Preference estimation with centralized school choice and college admissions." *American Economic Review*, 109(4): 1486–1529.

Feller, Avi, Todd Grindal, Luke Miratrix, Lindsay 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, David, and Lloyd S Shapley. 1962. "College admissions and the stability of marriage." *The American Mathematical Monthly*, 69(1): 9–15.

Hastings, Justine S, Christopher A Neilson, and Seth D Zimmerman. 2013. "Are some degrees worth more than others? Evidence from college admission cutoffs in Chile." National Bureau of Economic Research.

Heckman, James J. 1979. "Sample selection bias as a specification error." *Econometrica: Journal of the econometric society*, 153–161.

Heckman, James J, and Edward Vytlacil. 2001. "Policy-relevant treatment effects." *American Economic Review*, 91(2): 107–111.

Heckman, James J, Sergio Urzua, and Edward Vytlacil. 2008. "Instrumental variables in models with multiple outcomes: The general unordered case." *Annales d'Economie et de Statistique*, 151–174.

Heinesen, Eskil, and Christian Hvid. 2019. "Returns To Field Of Study In The Medium Term – Instrumental Variables Estimates Based On Admission Thresholds." Working Paper.

Hendren, Nathaniel, and Ben Sprung-Keyser. 2020. "A unified welfare analysis of government policies." *The Quarterly Journal of Economics*, 135(3): 1209–1318.

Humlum, Anders, and Bjørn Bjørnsson Meyer. 2020. "Artificial Intelligence and College Majors." Working Paper.

Imbens, Guido W, and Joshua D Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(2): 467–475.

Kapor, Adam J, Christopher A Neilson, and Seth D Zimmerman. 2020. "Heterogeneous beliefs and school choice mechanisms." *American Economic Review*, 110(5): 1274–1315.

Kirkeboen, Lars J, Edwin Leuven, and Magne Mogstad. 2016. "Field of study, earnings, and self-selection." *The Quarterly Journal of Economics*, 131(3): 1057–1111.

Kline, Patrick, and Christopher R Walters. 2016. "Evaluating public programs with close substitutes: The case of Head Start." *The Quarterly Journal of Economics*, 131(4): 1795–1848.

Kline, Patrick, and Christopher R Walters. 2019. "On heckits, late, and numerical equivalence." *Econometrica*, 87(2): 677–696.

Manning, Alan, and Barbara Petrongolo. 2017. "How local are labor markets? Evidence from a spatial job search model." *American Economic Review*, 107(10): 2877–2907.

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

Mountjoy, Jack, and Brent 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).

OECD. 2020. *Education at a Glance 2020.* OECD.

Piacquadio, Paolo G. 2020. *https://twitter.com/pgpiacquadio/status/1317848448406855680*.

Robbins, Herbert E. 1992. "An empirical Bayes approach to statistics." In *Breakthroughs in statistics*. 388–394. Springer.

Roy, Andrew Donald. 1951. "Some thoughts on the distribution of earnings." *Oxford economic papers*, 3(2): 135–146.

Sandmo, Agnar. 2015. "Chapter 1 - The Principal Problem in Political Economy: Income Distribution in the History of Economic Thought." In *Handbook of Income Distribution*. Vol. 2 of *Handbook of Income Distribution*, , ed. Anthony B. Atkinson and François Bourguignon, 3–65. Elsevier.

Stepner, Michael. 2013. "BINSCATTER: Stata module to generate binned scatterplots." *Statistical Software Components, Boston College Department of Economics*.

Tanaka, Mari, Yusuke Narita, and Chiaki Moriguchi. 2020. "Meritocracy and Its Discontent: Long-run Effects of Repeated School Admission Reforms." Research Institute of Economy, Trade and Industry (RIETI).

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

Young, Alwyn. 2019. "Consistency without inference: Instrumental variables in practical application." Working Paper.

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

A Appendix - Methods

A.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.2 Empirical Bayes schrinkage

In this section I provide details on the Empirical Bayes shrinkage procedure outlined in Section 4.1. As mentioned in the main text, this procedure follows Abdulkadiroğlu et al. (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 (14) and (15), 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:

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

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 Appendix - Figures

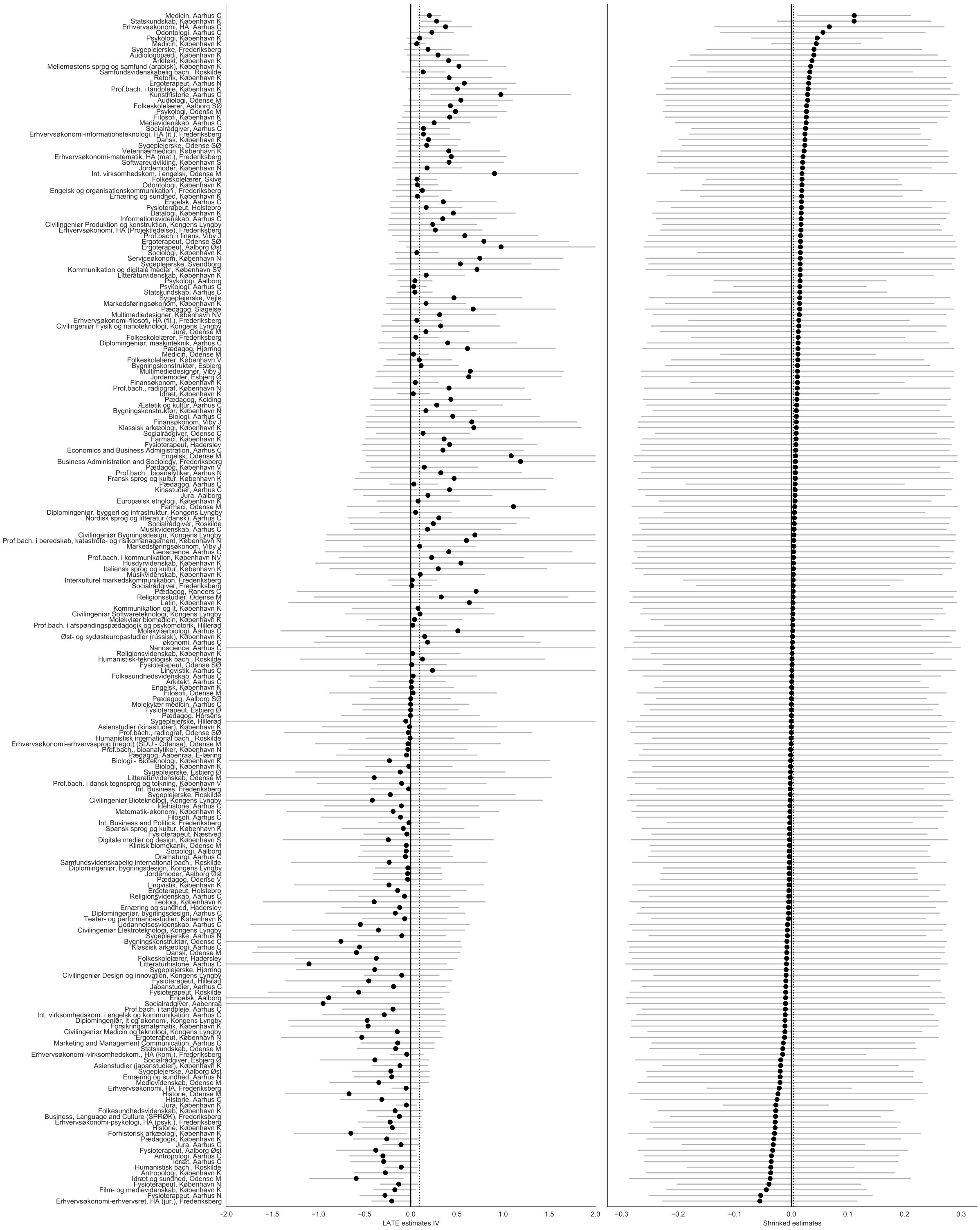


Figure B.1: IV-estimates of LATE

The figure shows LATE estimates calculated from a standard fuzzy regression discontinuity design with a linear spline in the running variable as well as the posteriors of an Empirical Bayes procedure where estimates are shrunken towards zero. The source of variation is crossing own program cutoff in quota 1. Estimates with an F-statistic lower than 10 have been excluded. The bars represent 95 percent confidence intervals. The IV intervals are based on heteroscedastic standard errors. Confidence intervals for the posteriors are calculated using the methods developed by Armstrong, Kolesár and Plagborg-Møller (2020). The dotted lines represent the overall mean of the raw and shrunken effects respectively.