Equilibrium Effects of Higher Education Subsidies to Public Schools*

Vinícius Botelho†

August 13, 2024

[Latest version here]

Abstract

This paper sets up a structural supply and demand model for higher education to understand how capacity constraints change the equilibrium impact of subsidies on enrollment and prices. To do so, we estimate the model using Brazilian data and simulate the effects of counterfactual conditional (only public schools) and unconditional (all schools) subsidies. When we increase unconditional subsidies, private prices go down, and enrollment goes up. We show that capacity constraints in public schools reinforce both the price and enrollment effects through student sorting. When subsidies are conditional and public degrees are not capacity-constrained, we see price-increasing competition; however, it vanishes as the level of capacity utilization in public schools rises. These results suggest that budget-neutral reforms increasing subsidies to private schools would reduce private prices significantly. Using the model, we estimate that a budget-neutral reform increasing the monthly tuition in public degrees by BRL 100.0 and providing a BRL 22.7 monthly unconditional scholarship increases enrollment by 1.1%, reduces the enrollment gap between high-SES and low-SES students by 0.8 p.p., and increases consumer surplus by 0.4% of baseline prices (even though average market prices rise by 9.8%).

***Keywords*:** mixed markets, optimal taxation, public provision, higher education, pricing, equilibrium effects
***JEL Codes*:** H21, H42, H44, H52, L11, L33, I22, I23

*I am extremely indebted to Michael Dinerstein, Fabio Miessi, and Cristine Pinto for their invaluable guidance and feedback. I am grateful, for their helpful comments and suggestions, to Christina Brown, Christopher Campos, Umut Dur, Esperanza Johnson, Jack Mountjoy, Letícia Nunes, Marcelo Sant'Anna, Rodrigo Soares, and seminar participants at the 22^{nd} International Industrial Organization Conference (2024), the Applied Micro Workshops at Insper in 2022, 2023, and 2024, and a workshop to PhD candidates in Economics at the University of Chicago in 2023. All opinions, errors, and omissions are solely attributable to me. This paper uses data from Serviço de Acesso a Dados Protegidos (SEDAP) at Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) through process 23036.000713/2022-85. The Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) supports this work through grants 2022/01333-2 and 2023/00373-3. The views expressed herein are mine and do not necessarily reflect those of Insper, FAPESP, or INEP.

†Insper (Job Market Paper). Email: viniciusob@al.insper.edu.br

1 Introduction

Many goods that generate positive externalities, such as education, healthcare, and childcare, are supplied in mixed markets, where there is private and public provision. For example, more than 95% of the countries have private and public supply in higher education (Levy, 2018). Typically, the public provision is heavily subsidized, in many cases to the point of generating excess demand.

The coexistence of a private market and a public capacity-constrained sector has implications for how markets clear. When shocks increase demand for goods produced by the private sector, price adjustments clear the market. On the other hand, public prices are usually heavily regulated and cannot be adjusted to guarantee that demand always meets capacity. In this case, the goods produced by the public sector need to be rationed. For example, when degrees in higher education are capacity-constrained, they may only admit students whose scores lie above a threshold. Generally speaking, customers are assigned priority levels, and only those whose priority lies above a threshold can access the public provision. These criteria change consumer choices, impacting private prices and optimal policy. Yet, little is known about how mixing market clearing mechanisms affects optimal subsidy design. In this paper, we hope to shed light on this question.

Specifically, we focus on the role of capacity constraints in public schools in the effect of conditional (public) and unconditional (public and private) subsidies on enrollment and private prices, discussing their implications for optimal policy. To do so, we set up and estimate a structural supply and demand model for the higher education market in Brazil. We follow standard discrete choice literature in which degrees have heterogeneous levels of attractiveness, costs, and capacity. Private degrees choose prices, and public degrees choose admission scores. Students have heterogeneous price sensitivities (according to their socioeconomic status and test scores) and choose degrees in a random utility model with heterogeneous choice sets (accounting for the admission process). Using the model, we simulate counterfactual conditional and unconditional subsidies. Then, we decompose the equilibrium impacts of subsidy changes into two components: the direct effect, which measures the impact of subsidy changes when public degrees do not face capacity constraints, and the indirect effect, which isolates the impact of subsidy changes caused by changes in the extent to which public degrees are capacity-constrained. We ask what is the effect of conditional and unconditional subsidies on enrollment and private prices; then, we estimate the direct and indirect effects.

We find a substantial role for capacity constraints in equilibrium outcomes. On the one hand, we document that unconditional subsidies increase enrollment and capacity constraints in public degrees contribute to this through a private price reduction. Noteworthy, the number of students enrolled in private degrees because of the price reduction caused by capacity constraints in public schools is higher than the number of students prevented from enrolling in public schools because of capacity limits. On the other hand, when subsidies are conditional and capacity in public degrees is not constrained, we see price-increasing competition between private and public providers. However, the price-increasing competition disappears as the capacity of public schools gets constrained. We show that a BRL 100 tuition in public schools and a budget-neutral unconditional scholarship would increase enrollment by 1.1%, reduce the enrollment gap between rich and poor students by 0.8p.p., and lower average private prices by 1.0%. Because the simulation involves a rise in public school prices, average market prices rise 9.8%. Even so, consumer surplus rises by 0.4% of average prices in the baseline.

Section 2 describes our theoretical framework, presenting a generalized structural supply and demand model with capacity constraints. It shows how subsidies to private and public players change total consumption and private prices, providing a guide on the equilibrium role of capacity constraints. In section

3, we present our data and Brazil's higher education market, highlighting that it provides an interesting application to our model. Usually, in similar problems, capacity is not observed, and the process through which goods are allocated under excess demand, the rationing rule, is unknown. Education is an interesting market to study capacity constraints because, as discussed in the extensive literature on matching mechanisms between students and schools (Aygün and Bó, 2021; Bó and Hakimov, 2019; Fack et al., 2019; Agarwal and Somaini, 2018; Balinkski and Sönmez, 1999; Gale and Shapley, 1962), the rationing rule is usually a function of observable test scores.

Furthermore, Brazil is an ideal setting for the empirical application for two reasons. First, it has a long tradition of admitting students based only on test scores, the *vestibulares*, and most public schools admit students through a centralized admissions system that ranks candidates using the same high school test (*ENEM*). Second, there is data on capacity utilization for higher education degrees. The data show that capacity utilization near 100% and high oversubscription levels are common in public degrees, suggesting that excess demand plays an important role in the market equilibrium. The law prohibits public schools from charging tuition, making it impossible to clear the market through price adjustments. On the other hand, most private schools display capacity utilization levels that do not suggest capacity constraints. Both types of providers hold sizable market shares in Brazil, suggesting there is room for a strong interaction between them.

Our core dataset is the Brazilian Higher Education Census, which contains information on all higher education schools, degrees, and students in Brazil. At the firm (school) level, we observe ownership (public or private) and the products (degrees) supplied; at the product (degree) level, we observe capacity, location (campus), and major. We enrich this dataset with administrative registries of student loan contracts (*FIES*), in which we observe the prices of most private degrees (public degrees do not charge tuition, so we know their prices). At the student level, the Higher Education Census provides us with the degree choices of all students. We enrich this dataset with the administrative registry containing student-level information on socioeconomic indicators and test scores from all students who took the national high school examination (*ENEM*). The data was merged in a secure room at the official agency for education statistics in Brazil (*INEP*). The dataset we use to estimate our demand function has information on more than 3 million students.

Section 4 presents our empirical model, and section 5 describes our empirical strategy for estimating it. To estimate demand parameters, we follow Berry et al. (2004). To account for the endogeneity between prices and unobserved degree attractiveness, we use a cost shifter: the average wages of higher education employees and college professors in the region where the degree is supplied. Section 6 shows model estimation results. We show that price elasticities are higher for poor students with low test scores, and the distance between students and degrees has a sizable effect on utility. One of the possible explanations for the fact that students with lower test scores have higher price-elasticity even after controlling for their socioeconomic status is that they face lower returns to education, reducing their willingness to pay for college.

Section 7 documents two sets of counterfactual results. First, we evaluate conditional and unconditional subsidy changes. When subsidies are conditional, only students who choose public schools receive them; when subsidies are unconditional, all students in higher education receive them. Counterfactual subsidies are flat, restricting simulations to policies with smaller pass-through (Sahai, 2023). Subsidies are a cash transfer, so they are not limited by tuition. Subsidy changes can be positive (increasing the subsidy from the baseline level) or negative (reducing the subsidy from the baseline level).

We find that increasing unconditional subsidies increases enrollment: a flat monthly scholarship of BRL

200[1] would increase enrollment by 46.6%,[2] with the indirect effect (measuring the role capacity constraints in public schools) responsible for an overall enrollment rise of 1.3%. To understand this result, we must analyze the joint effect of capacity constraints and subsidies on private prices. Simulations show that the same subsidy would reduce private prices by 3.3%, with the direct effect responsible for a price reduction of approximately 2.3% and the indirect effect responsible for a price reduction of approximately 1.0%. Unconditional subsidies make all schools more attractive to students in the outside option, particularly the price-elastic. High-score price-elastic students switching from the outside option to public degrees raise admission scores and force low-score students in public degrees to switch to private (and to the outside option), explaining the indirect effect. After all, low-score students have higher price elasticity, increasing the price elasticity of students in private schools and reducing market prices. On the other hand, the mechanism behind the direct effect is that unconditional subsidies incentivize price-elastic students in the outside option to switch to private schools, increasing the proportion of high-elasticity students attending private degrees and also reducing prices.

When evaluating conditional scholarships, we see a much smaller effect on enrollment. A flat conditional scholarship double the size of the unconditional subsidy we just discussed (BRL 400) would increase enrollment only by 4.1%, mainly because of the mechanical impact of capacity constraints on public school enrollment. When we turn our attention to the effect of conditional subsidies on prices, we find nonlinear results. For capacity utilization levels below the baseline, we find price-increasing competition between public and private schools; for capacity utilization levels above the baseline, we see higher subsidies to public schools reducing private prices. The decomposition is very useful for understanding the nonlinearity. The direct effect shows the price-increasing competition in all simulations: when conditional subsidies are BRL 400, the direct effect points to a 3.5% private price increase. However, the indirect effect forces the total price effect in the opposite way: when conditional subsidies rise, admission scores also rise and force low-score students to switch to private degrees or the outside option, increasing the proportion of price-elastic students attending private schools and reducing prices. The same conditional subsidy of BRL 400 has an indirect effect on prices of -4.0%, with a total reduction in private prices of 0.5%, which is not enough to generate sizable aggregate enrollment effects.

In these exercises, the subsidy expenditures change in every simulation. Therefore, optimal subsidy design would require hypotheses on the effects of raising or reducing taxation to match changing subsidy expenditures and a general equilibrium analysis. So, we build our second set of counterfactuals by joining this concern with the fact that private prices fall when unconditional subsidies rise. We look for a budget-neutral equilibrium in which we charge tuition in public schools and use tuition revenue to fund an unconditional subsidy. We find that charging small tuition (BRL 100 monthly) and providing unconditional subsidies can be a welfare-improving reform: enrollment rises (by 1.1%), the enrollment gap between high-SES and low-SES students falls by 0.8 p.p., and consumer surplus rises by 0.4% of baseline prices, even though market prices rise by 9.8%. As expected, private prices fall (by 1.0%), but profits rise (by 6.6%).

This paper connects to several strands of the literature. The first set of papers looks at subsidies in mixed markets. When private firms have market power, market segmentation into public and private provision may leave private providers with low price-elasticity consumers, generating price-increasing competition, as Atal et al. (2024) has shown for pharmacies in Chile. Our empirical application also shows some level of price-increasing competition between private and public providers, and we discuss the consequences of such a

[1] Approximately USD 40.

[2] Unless stated otherwise, all effects are measured in comparison to baseline levels.

feature on the market equilibrium. Furthermore, there are many examples of mixed markets in which there is rich interaction between private and public providers, like primary and secondary schools (Dinerstein and Smith, 2021), food distribution (Jiménez-Hernández and Seira, 2022; Handbury and Moshary, 2021; Banerjee et al., 2019), retail stores (Busso and Galiani, 2019), healthcare (Curto et al., 2019), and financial services (Fonseca and Matray, 2022). We add to this literature by showing how capacity constraints change the impact of subsidies on overall consumption and private prices in mixed markets, discussing the implications of these findings to optimal policy. We also discuss the conditions under which it would be optimal to design the subsidies in a way that causes excess demand for public provision.

The second set of papers looks at the public provision of goods. In a very influential paper, Besley and Coate (1991) provide a framework in which the public provision of a high-quality good is Pareto-dominated by the public provision of a low-quality good, which serves as a tool to redistribute resources from the rich to the poor. In their setting, providing high quality at zero price would make everyone choose and pay for high quality, even those who value it below marginal cost. In our framework, the public provision is of high quality but capacity-constrained. Therefore, it cannot be universal and is not Pareto-dominated by the provision of low-quality goods.

We give two contributions to this literature. The first is to discuss the equilibrium effects of capacity constraints when the public provision is high quality, which sheds light on an alternative approach to Besley and Coate (1991). The second is to discuss the case of an indivisible good whose consumption level is important to the policymaker because of market failures. This is the case for many real-world public provision schemes, like education, health, and childcare markets, and it matters to the optimal design of public provision. We still present a discussion in which the market failures are not strong enough to make the public provision efficient, but we discuss how to redistribute resources in a way that maximizes the consumption of the type of goods being provided. Our results are robust to welfare-maximizing metrics (consumer surplus and profits).

The third set of papers looks at the higher education market. There is a long literature on the assignment of students to capacity-constrained degrees (Aygün and Bó, 2021; Bó and Hakimov, 2019; Fack et al., 2019; Agarwal and Somaini, 2018; Balinkski and Sönmez, 1999; Gale and Shapley, 1962) and another on higher education subsidies, discussing their importance (Dynarski et al., 2022; Chetty et al., 2020), optimal design (Epple et al., 2017), pass-through (Kargar and Mann, 2022; Dobbin et al., 2022; Lucca et al., 2018; Turner, 2017; Cellini and Goldin, 2014; Turner, 2012; Singell and Stone, 2007), and side effects (Cohodes and Goodman, 2014; Cellini, 2009; Long, 2004; Peltzman, 1973). We use the literature that discusses the assignment of students to capacity-constrained degrees to inform our empirical model, and we make two main contributions to the literature on higher education subsidies. First, we show a tractable way of incorporating a matching mechanism between students and schools in a Nash equilibrium that allows the interaction of profit-maximizing private schools with their public counterparts, as happens in many real-world markets (Levy, 2018). Second, we document price-increasing competition between private and public schools in a context where many public schools are high-quality and capacity-constrained. Consequently, a budget-neutral reform that would reduce subsidies to public degrees and increase subsidies to private schools increases enrollment and lowers the gap between poor and non-poor students.

2 Theoretical Framework

Our model assumes the existence of one public and one private supplier, producing differentiated goods. Producers are indexed by j and consumers are indexed by i. Consumer i pays p_j, and the indirect utility function of consumer i from choosing alternative j is given by Equation (1). V_{ij} is the individual-specific non-price component of utility from the consumption of good j, α_i is an individual-specific price sensitivity coefficient, and ϵ_{ij} is an extreme value type I shock (i.i.d across consumers and alternatives).

$$u_{ij} = V_{ij} - \alpha_i p_j + \epsilon_{ij} \tag{1}$$

We assume that each customer is assigned to a priority index $a \in [0, 1]$. When the capacity of the public firm is constrained, only those whose priority index is above a threshold λ have access to the goods produced by the public sector. In other words, the goods produced from the public firm are in the choice set of a customer whose priority index is a if, and only if, $a \geq \lambda$. We assume that there are no ties in a and there is a mass one of consumers, allowing us to match each individual to a priority level. Therefore, we can write functions V_j and α as shown in Equation (2).

$$\begin{aligned} V_{ij} &= V_j(a) \\ \alpha_i &= \alpha(a) \end{aligned} \tag{2}$$

Customers' choice sets are heterogeneous and change according to a and λ. Consumers are in one of two possible states ($a \geq \lambda$ or $a < \lambda$), and we have to distinguish the demand for both types of goods in each of them: s_j^U is the unrestricted demand (the demand for goods from producer j when consumers have both alternatives in their choice set); s_{pri}^R is the restricted demand (the demand for goods from the private firm when consumers do not have the goods from the public firm in their choice set). The individual demand functions are shown in Equation (3).

$$\begin{aligned} s_j^U(a, p_j, p_{-j}) &= \frac{exp(V_j(a) - \alpha(a)p_j)}{1 + \sum_k exp(V_k(a) - \alpha(a)p_k)} \text{ if } a \geq \lambda \\ s_{pri}^R(a, p_{pri}) &= \frac{exp(V_{pri}(a) - \alpha(a)p_{pri})}{1 + exp(V_{pri}(a) - \alpha(a)p_{pri})} \text{ if } a < \lambda \end{aligned} \tag{3}$$

To obtain aggregate demand, we sum individual demand over the space of students' characteristics. To do so, we rely on Assumption (1), which states there exists a partition of the space of priority indexes in which all consumers whose priority lies in the same interval have the same observable preference parameters.[3] Nevertheless, the demand function is still very flexible and represents a wide range of applications in discrete choice. After all, there is no bound to the difference between preference parameters among intervals and there is no limit to the number of intervals in which we partition the space of priority indexes.

Assumption 1. *$\exists \mathcal{P}$ such that $\forall a_1, a_2 \in P_k$: $V_j(a_1) = V_j(a_2)$ and $\alpha(a_1) = \alpha(a_2)$, where $\mathcal{P} = \{P_1, ...P_n\}$ is a partition of $[0, 1]$ into intervals.*

Consumption for each type of good and total consumption (S) are shown in Equation (4).

[3] The assumption is essential to allow for integration. We use the Riemann integral.

$$
\begin{aligned}
S_{pub}(\lambda, p_{pub}, p_{pri}) &= \int_{\lambda}^{1} s^{U}_{pub}(a, p_{pub}, p_{pri})da \\
S_{pri}(\lambda, p_{pub}, p_{pri}) &= \int_{0}^{\lambda} s^{R}_{pri}(a, p_{pri})da + \int_{\lambda}^{1} s^{U}_{pri}(a, p_{pub}, p_{pri})da \\
S(\lambda, p_{pub}, p_{pri}) &= S_{pub}(\lambda, p_{pub}, p_{pri}) + S_{pri}(\lambda, p_{pub}, p_{pri})
\end{aligned}
\tag{4}
$$

We assume there are flat subsidies to public and private firms. Hence, there is a distinction between the price consumers pay for good j (p_j) and the price supplier j charges for each unit produced (p^S_j), which depends on the subsidy to the firm of type j (t_j). The relationship between these variables is given by Equation (5).

$$
p_j = p^S_j - t_j \tag{5}
$$

We assume all costs for public provision are infrastructure costs, so the marginal cost of public firms is zero, and public firms charge the marginal cost ($p^S_{pub} = c_{pub} = 0$). Furthermore, policymakers set their priority thresholds (λ) to maximize the amount of goods they supply. In other words, they adjust the threshold to make demand as close as possible to their capacity level (K), as stated in Equation (6).

$$
\lambda(p_{pub}, p_{pri}, K) = \min\{\lambda : S_{pub}(\lambda, p_{pub}, p_{pri}) \leq K\} \tag{6}
$$

Private firms, on the other hand, choose prices (p^S_{pri}) maximizing profits. When doing so, they take thresholds in the public sector as given. We assume they face a fixed marginal cost (c_{pri}). The profit-maximization problem and the pricing equation are shown in (7).

$$
\begin{aligned}
&\max_{p^S_{pri}} S_{pri}(\lambda, p_{pub}, p_{pri})(p^S_{pri} - c_{pri}) \\
&p^S_{pri} = c_{pri} + \frac{S_{pri}(\lambda, p_{pub}, p_{pri})}{\left|\frac{\partial S_{pri}(\lambda, p_{pub}, p_{pri})}{\partial p_{pri}}\right|}
\end{aligned}
\tag{7}
$$

2.1 Optimal policy

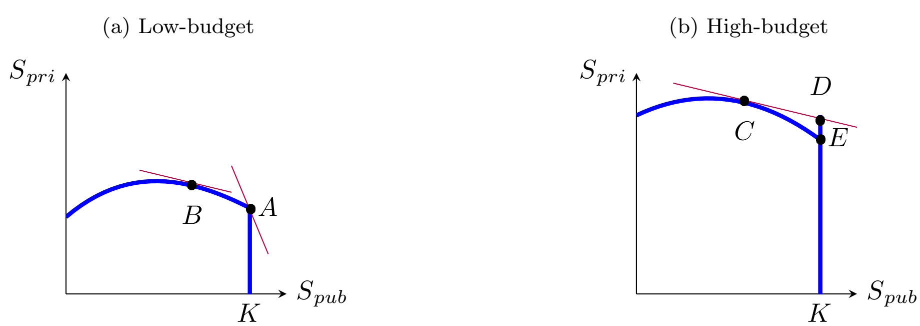
In our model, the final goods provide a positive externality, so the policymaker is concerned with their consumption levels. We allow different weights for goods produced by public and private firms: policymaker's preferences are equal to a weighted sum of private and public consumption in which public consumption is weighted by $\phi_{pub} > 0$. The objective of the policymaker is to set subsidies to maximize total weighted consumption, given a budget level (T) and public capacity (K). The problem is shown in (8).

$$
\begin{aligned}
&\max_{t_{pub},t_{pri}} S_{pri}(\lambda, p_{pub}, p_{pri}) + \phi_{pub} S_{pub}(\lambda, p_{pub}, p_{pri}) \\
&\text{s.t.} \\
&T = t_{pub} S_{pub}(\lambda, p_{pub}, p_{pri}) + t_{pri} S_{pri}(\lambda, p_{pub}, p_{pri}) \\
&S_{pub}(\lambda, p_{pub}, p_{pri}) \leq K
\end{aligned} \tag{8}
$$

The solution to the Problem (8) is shown in Figure (1). In Panel (1a), we write the combination of S_{pub} and S_{pri} compatible with a fixed budget level, in blue. The curve represents all feasible combinations of S_{pub} and S_{pri}. There is a positively-inclined region in the blue curve in which it is impossible to have an optimal combination of S_{pri} and S_{pub} (unless we allowed ϕ_{pub} to be negative). In this region, the subsidy to public firms is negative (a tax) and reducing the tax rate (increasing the subsidy) is simultaneously increasing the demand for goods produced by the public sector and increasing total revenue, allowing for a simultaneous increase in the subsidy to private players and in the demand for goods produced by the private sector.[4]

Then, we draw the policymaker indifference curves, in purple, and look for the highest feasible one. In Panel (1a), we plot two possible indifference curves, one assuming $\phi_{pub} = 1.0$, which gives point A as the optimal, another assuming $\phi_{pub} = 0.1$, which gives point B as the optimal. Optimal subsidy levels are implicitly determined.

Figure 1: Capacity constraints and optimal policy



Note: simulation results. Half the customers have preferences are given by $V_{i,pub} = V_{i,pri} = -1$ and $\alpha_i = 6$, and all of them have higher priority than the other half, whose preferences are given by $V_{i,pub} = V_{i,pri} = 4$ and $\alpha_i = 2$. Capacity: $K = 0.1$. Marginal cost: $c_{pri} = 1$. Budget: $T = 0.30$ (low) or $T = 0.45$ (high). At point A, $\phi_{pub} = 1.0$, at points B, C, and D $\phi_{pub} = 0.1$. All axes have the same sizes.

Even though policymaker's preferences play a role in determining whether it is optimal to be capacity-constrained, the shape of the corner at A implies that there is a wide range of parameters for ϕ_{pub} for which it is optimal to keep capacity constrained in public firms. In other words, when policy is set optimally, the existence of capacity-constrained providers in the public sector is informative about the behavior of private firms. For example, when private subsidies increase prices (e.g., the pass-through is low), the blue curve becomes less inclined, and capacity constraints in public provision are more likely. Therefore, if policy is optimal, the occurrence of capacity constraints in the public provision in one market can be informative about the pass-through of private firms.

Panel (1b) shows simulation results assuming a higher budget (increased from 0.30, in Panel (1a) to 0.45

[4] The mechanism is the same of the Laffer curve.

in Panel (1b)) and $\phi_{pub} = 0.1$. Now, the policymaker is indifferent between C and D. However, they represent very different allocations: C generates idle capacity and D represents excess demand in public firms. As the budget level changes, the equilibrium jumps from idle capacity to excess demand.

Noteworthy, not all markets exhibit a segment like $\overline{DE}$. When they do, a wide range of policymaker preferences is compatible with optimal excess demand in public firms and optimal policy chooses either idle capacity or excess demand: regardless of policymaker preferences, full capacity utilization without excess demand would never be optimal.

The existence of a segment like $\overline{DE}$ depends on the interaction between the price response of firms and the shape of the demand curve. It can also occur when prices are set in perfect competition.

In the example, subsidies to public (private) firms are higher (lower) at any point in $\overline{DE}$ than at E[5] and it is not surprising that this combination of subsidies generates excess demand in public firms. What is surprising is that, in this region, subsidies to public firms have a positive effect on the demand for privately-produced goods and such effect is even stronger than that of the subsidies to private firms themselves. After all, increasing subsidies to public firms and reducing subsidies to private firms is increasing private demand at $\overline{DE}$.

To understand the phenomenon, we write the effect of public subsidies on private demand in Equation (9).[6]

$$\frac{dS_{pri}}{dt_{pub}} = \left[-\frac{\partial S_{pri}}{\partial p_{pub}} - \frac{\partial S_{pri}}{\partial \lambda}\frac{\partial \lambda}{\partial p_{pub}} \right] + \frac{\frac{dp_{pri}}{dt_{pub}}}{\frac{dp_{pri}}{dt_{pri}}}\frac{dS_{pri}}{dt_{pri}} \tag{9}$$

There are two components with opposing signs inside the brackets in Equation (9), both related to price effects. When prices in public firms falls, some of the customers who have access to both types of goods switch to public, reducing private consumption (measured by $-\frac{\partial S_{pri}}{\partial p_{pub}}$) and increasing the threshold (measured by $-\frac{\partial \lambda}{\partial p_{pub}}$). When the threshold rises, some customers who had access to goods produced by the private and the public sector now can only consume privately-produced goods, increasing private demand (measured by $\frac{\partial S_{pri}}{\partial \lambda}$).

For the overall effect inside brackets to the positive, the number of consumers who switch from private to public because prices changed must be lower than the number of customers who switch from public to private because they do not have access to the goods produced by the public firm anymore. This happens when the threshold increase caused by the fall in public prices is driven by consumers switching from the outside option and most of the costumers forced to switch from the public option choose the private alternative instead of the outside good.

The second component in Equation (9) is the ratio of the impact of public and private subsidies on private

[5]Without further assumptions, it is not impossible to assert that this is always the case when there is a segment like $\overline{DE}$. The only combination of subsidy levels which is impossible in a segment like $\overline{DE}$ is the one in which subsidies for both private and public firms are higher under D when compared to E, because it would violate budget neutrality: higher subsidies to public when public capacity is constrained necessarily increase the expenditure with this kind of subsidy, and if subsidies to the private firm are higher under D, which has a higher private demand than E, expenditure with private subsidies would rise as well. So, D and E would necessarily involve different expenditure levels.

[6]Noteworthy, the private demand function ($S_{pri}(p_{pri}, p_{pub}, \lambda)$) is continuous, but not differentiable at all points. To show that, take a such that $\alpha(a-\epsilon) \neq \alpha(a+\epsilon)$ and $V_j(a-\epsilon) \neq V_j(a+\epsilon)$, for any $\epsilon > 0$ (i.e., a is at the border of any of the intervals defined in Assumption 1). Then, it is clear that the partial derivative of S_{pri} w.r.t. λ does not exist at a because $\lim_{\epsilon \to 0^-} \frac{S_{pri}(p_{pri},p_{pub},a+\epsilon)-S_{pri}(p_{pri},p_{pub},a)}{\epsilon} \neq \lim_{\epsilon \to 0^+} \frac{S_{pri}(p_{pri},p_{pub},a+\epsilon)-S_{pri}(p_{pri},p_{pub},a)}{\epsilon}$. Nevertheless, the derivatives are still informative about the function's behavior because they do not exist on only a finite number of points and S_{pri} is continuous. The non-existence of the derivatives at these points is important for optimization because the thresholds at the borders of the intervals in Assumption 1 are strong candidates for optimal thresholds. After all, choosing the threshold at one of these points is equivalent to choosing the types of customers to access the goods produced by the public firms.

prices and measures the effect of price responses on private demand growth. As we will discuss shortly, it follows from Proposition (1 and Equation (10)) that $\frac{dS_{pri}}{dt_{pri}}$ and $\frac{dp_{pri}}{dt_{pri}}$ have opposite signs. Therefore, when there is price-increasing competition (i.e., subsidies to public firms raise private prices), $\frac{dp_{pri}}{dt_{pub}} > 0$ and price responses reduce private demand; when subsidies to public firms reduce private prices, price responses increase private demand. To understand the relationship between $\frac{dS_{pri}}{dt_{pri}}$ and $\frac{dp_{pri}}{dt_{pri}}$, Equation (10) shows the impact of private subsidies on private demand.

In Equation (10), the substitution induced by prices and the threshold is very similar to the mechanisms described in Equation (9). However, now it is possible to show that the substitution effect from prices is stronger than the one induced by threshold changes, as stated by Proposition (1). As a consequence, the quotient between $\frac{dS_{pri}}{dt_{pri}}$ and $\frac{dp_{pri}}{dt_{pri}}$ is always negative, as we have just discussed. The result implies that the sign of the effect of subsidies to private firms on private prices is always determined by the sign of the private price pass-through.

$$\frac{dS_{pri}}{dt_{pri}} = \left[\frac{\partial S_{pri}}{\partial p_{pri}} + \frac{\partial S_{pri}}{\partial \lambda}\frac{\partial \lambda}{\partial p_{pri}}\right]\frac{dp_{pri}}{dt_{pri}} \tag{10}$$

Proposition 1. $\frac{\partial S_{pri}}{\partial p_{pri}} + \frac{\partial S_{pri}}{\partial \lambda}\frac{\partial \lambda}{\partial p_{pri}} < 0$

Proof. In Appendix D □

Going back to the example simulated in Figure (1), the high-priority group has high price elasticity ($\alpha_i = 6$) and low propensity to choose the inside good ($V_{i,pub} = V_{i,pri} = -1$), and the low-priority group has low price sensitivity ($\alpha_i = 2$) and high propensity to choose the inside good ($V_{i,pub} = V_{i,pri} = 4$). Consequently, a very high subsidy would be necessary to induce the high-priority consumers to choose any inside good. As the very high subsidy makes the public firms capacity-constrained, low-priority costumers are forced to switch from public to private or from public to the outside option. Because they have a high preference for the inside good, they switch to private. The net effect implies that a very high subsidy to the public firm increases overall consumption.

We can interpret our subsidy changes as a treatment. In this case, the optimal solution in the example involved increasing the subsidy incidence to compliers (consumers who choose the inside option only if the price is low), displacing always-takers to the private sector.

All in all, the optimal level of subsidies to maximize the objective function of the policymaker is an empirical question. In the next section, we discuss the institutional framework in which we apply this model.

3 Institutional Background

In this section, we discuss examples of markets displaying capacity constraints and high subsidy levels. Then, we present the institutional framework of our empirical model, the higher education market in Brazil. We present our data and discuss how the data informs our model choices regarding the problem of private and public schools. We also discuss descriptive evidence regarding the distribution of subsidies and enrollment between students from high and low socioeconomic statuses (SES).

3.1 Capacity Constraints and Public Provision

Capacity constraints manifest themselves differently depending on the market. In healthcare, they cause longer waiting times; in education, higher admission scores; in childcare, unmet needs. Even though price increases would be an effective tool to clear markets and reduce the level of excess demand, policymakers choose to keep these goods heavily subsidized even when this results in demand exceeding capacity limits.

According to OECD (2020), people face zero or low co-payments for health services in OECD countries and, as a result, waiting times, usually caused by capacity constraints and inefficiencies, are an important policy issue in most of them. The report shows results from a survey with policymakers from 24 countries asking whether waiting times in health services are an issue across several general and disease-specific dimensions: in 10 of them, waiting times were considered an issue across at least 75% of the possible dimensions (only Japan did not consider waiting times an issue in any dimension). Another survey in the same report shows that 17 out of 34 countries considered waiting times as a high priority issue. Using hard data, OECD (2020) showed that waiting times for elective surgery can be ten-fold higher in some countries, highlighting how capacity constraints are relevant in many contexts. Moreover, the report highlights how many countries were not successful in reducing capacity constraints by increasing supply.

When capacity is constrained in health, urgent treatment is prioritized, in detriment of elective procedures. Consequently, it is difficult to establish a rationing rule that allows the estimation of a demand model and simulation of counterfactual policies.

According to Eurostat (2016), the proportion of households in the European Union reporting unmet need for formal childcare services is 12%. By country, the proportion varied from 3.0% (Bulgaria) to 22.2% (United Kingdom).The two main reasons why the respondents answered that were not making more use of formal services were not being able to afford it (42.7%) and finding no places available (13.7%). More people responded that there were no places available than that they could not afford the services in Finland (28.4% vs. 19.2%) and Poland (17.0% vs. 19.0%); furthermore, the proportion of people who responded that there were not plaices available was higher 10% in 17 out of 33 countries. On average, 44% of the respondents who pay for childcare services face a reduced price. The country-level correlation between the proportion of people who answered that the government pays for childcare services and the proportion of people who found no places available is 0.314, indicating that, at some extent, public provision and low capacity seem to be occurring in the same places.

Reinforcing this conclusion, Eurydice (2019) argues that demand for ECEC (early childhood education and care) facilities is higher than supply in many countries, such as Italy, Romania, Slovakia, Montenegro, and Turkey; for children less than 3 years-old, the phenomenon is even more general, affecting also Portugal, Spain, France, and Germany. Given that public and publicly subsidied supply is a very common practice, this also raises the concern about subsidies to excess demand. More than half of the countries classified by the report as having very low fees for children under three years show, according to it, more demand than supply for this same age group (Bulgary, Lithuania, Romania, Montenegro, North Macedonia, and Serbia).

In education, there is a widespread practice of using matching mechanism to assign students and schools, extensively studied in the literature (Aygün and Bó, 2021; Bó and Hakimov, 2019; Fack et al., 2019; Agarwal and Somaini, 2018; Balinkski and Sönmez, 1999; Gale and Shapley, 1962). The purpose of these mechanisms is to guarantee ideal properties for the allocation of students to schools when capacity is limited. However, the fact that individual schools are capacity-constrained may result from their demand level (e.g., they may be high-quality), and not necessarily from an excessive subsidy to all schools in the market. Fortunately, the survey on access to services in Eurostat (2016) sheds light in this question. In fact, in 5 out of 31 countries for which we have data more than 10% of the respondents who do not participate in formal education declared that they did so because they were not admitted to the course to which they applied. The country-level correlation between the proportion of people who reported that they did not pay tuition and people reporting they were not in formal education because they were not admitted is 0.160.

So, education is a setting in which capacity constraints are important and heavily studied. Furthermore, the fact that the priority indexes, test scores, are observed, makes the market an ideal case to study the problem.

3.2 Higher Education in Brazil

Public higher education in Brazil is high-quality, selective, and heavily subsidized (free). In 2016, there were approximately 15 candidates per seat in public degrees, while the proportion of candidates per seat in private schools was approximately one. Nevertheless, private schools hold a sizable share of the market. Admission to public degrees is based only on test scores, which makes the priority index observable.

Schools supply degrees, to which students apply. When choosing degrees, students simultaneously choose a major, a school campus, and the time of day classes will be held, such as morning, afternoon, or evening.

The market has private and public schools, which can be federal, state, or municipal. By law, public schools do not charge tuition and are high-quality: according to the U.S. News Ranking for 2022-2023, there are 16 Brazilian public universities among the best 30 universities in Latin America. The combination of high quality and no tuition makes the public option attractive to all students, raising concerns about whether subsidies effectively increase student enrollment in higher education.

Yearly, Brazil holds a national high school examination called ENEM (*Exame Nacional do Ensino Médio*). It is a non-mandatory test students can take in high school or after graduation, commonly used for admission in public schools. The test is divided into four subjects (Mathematics, Language, Natural Sciences, and Social Sciences) and an essay.

Most public schools admit students using a centralized system, SISU (*Sistema de Seleção Unificada*), which ranks students according to ENEM scores, reported preferences, and affirmative action eligibility. SISU is similar to the Iterative Deferred Acceptance Mechanism, extensively described in Bó and Hakimov (2019).[7] It opens twice yearly, and admissions follow a cutoff score mechanism. Only test scores and affirmative action eligibility are taken into account in the admission process.

3.2.1 Data and Descriptive Statistics

Our core student and degree data come from three sources.

The first dataset is the 2015 ENEM (INEP, 2015), which is the administrative registry with the data from all test-takers in the 2015 edition of ENEM. This data provide us test scores, socioeconomic status, affirmative action eligibility, and location of all test-takers. We divide students into groups according to the percentile of their test scores, the quintile of their socioeconomic status, their affirmative action eligibility, and the combined statistical area in which they took ENEM.[8] Socioeconomic status is defined as family *per capita* income.[9] We only kept in the sample ENEM test takers in 2015 who already had a high school diploma or were graduating that year.

[7]In contrast to the standard Iterative Deferred Acceptance Mechanism, SISU allows students to make two applications (informing their first and second choices), is designed for only four rounds, and allows students to change their choices anytime.

[8]The income levels and test scores that correspond to each group can be found in the Appendix.

[9]Details about how the family *per capita* income and affirmative action eligibility were calculated can be found in the Appendix.

The second dataset is the 2016 Higher Education Census (INEP, 2016), which is an administrative registry containing data from an annual survey on all higher education institutions in Brazil and their students. The data provide us with students' degree choices (the degrees in which these students enrolled as first-year students in 2016), degree capacity, degree ownership status (public or private), and which degrees are free. We dropped information on distance learning degrees and other education modalities beyond high school, such as technical degrees.

The third dataset is the 2016 FIES dataset,[10] which is an administrative registry containing data from all student loan contracts. The data provide us private prices.

These datasets were accessed, merged, and enriched in a secure room at INEP (*Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira*), the official agency for education statistics in Brazil. To reduce the number of choices in the model, we assume that degrees are the combination of major, school, and municipality in which classes are held. This means aggregating degrees in the same municipality, school, and major, regardless of the specific campus or timing in which classes take place. Noteworthy, graduation diplomas usually clearly specify only the school and major in which students graduate.[11] Descriptive statistics of student and degree data are shown in Tables (1) and (2).

Table 1: Descriptive statistics: students

	Mean	Std.Dev.
College enrollment	0.39	0.49
Not eligible to affirmative action	0.34	0.47
Number of observations	5,480,902	

Source: Higher Education Census (INEP, 2016) and ENEM (INEP, 2015).

Table 2: Descriptive statistics: degrees

	Mean	Std.Dev.
Prices	489.15	584.830
Proportion of public	0.27	0.445
Capacity utilization	0.51	0.326
Number of observations	21,232	

Source: Higher Education Census (INEP, 2016) and FIES (MEC, 2016).

To measure higher education costs, we use RAIS (*Relação Anual de Informações Sociais*), an administrative registry containing information on the labor contracts of all companies in Brazil. It contains data on the industry, wages, occupation, and hours contracted of all formal labor market contracts in 2016. We restrict the sample to employees in higher education institutions who signed 20, 40, or 44-hour contracts, which are the most common workweeks in the data and represent 65% of the contracts in higher education institutions. In Table (3) we show average wages and hours contracted of all employees and college professors, by department (as officially stated in the labor contract).[12]

[10] FIES is Brazil's largest student loan program.

[11] For further reference, see Brazilian Ministry of Education (2018).

[12] Mathematics and Statistics, Architecture and Engineering, Biology and Health, Pedagogy, Language and Literature, Humanities, and Economics, Business, and Accounting.

Table 3: Descriptive statistics: labor contracts

Variable	Mean	Std.Dev	N
Wages (Mathematics and Statistics)	9,171.99	5,712.24	7,333
Wages (Architeture and Engineering)	10,718.14	6,454.42	10,798
Wages (Biology and Health)	9,497.45	6,217.29	21,587
Wages (Pedagogy)	7,977.41	6,133.14	52,728
Wages (Language and Literature)	8,639.11	5,188.05	4,989
Wages (Humanities)	8,752.22	6,325.41	16,186
Wages (Economics, Business, and Accounting)	8,138.82	6,629.28	10,810
Wages (all contracts)	5,957.00	5,615.16	552,083
Hours contracted (Mathematics and Statistics)	38.54	6.17	7,333
Hours contracted (Architeture and Engineering)	38.04	6.40	10,798
Hours contracted (Biology and Health)	36.66	8.24	21,587
Hours contracted (Pedagogy)	37.12	8.58	52,728
Hours contracted (Language and Literature)	38.08	6.64	4,989
Hours contracted (Humanities)	37.47	7.63	16,186
Hours contracted (Economics, Business, and Accounting)	35.67	9.74	10,810
Hours contracted (all contracts)	39.79	6.07	552,083

Source: RAIS (2016). **Notes:** wages and hours contracted of professors in each of the seven listed subjects, as well as wages and hours contracted of all employees. Prices in BRL.

3.2.2 Markets

As shown in Table (4), most schools are private. Among public schools, federal and state schools are the most common types.

Table 4: Higher education supply (2016)

	Private	Public (federal)	Public (state)	Public (municipal)
Number of Degrees	15,476	3,747	1,823	186
Number of Colleges	2,005	107	121	43
Number of Majors	79	85	74	37

Source: Higher Education Census (INEP, 2016).

Among students who go to college, 77% go to private degrees, and 23% go to public degrees, as shown in Table (5). Private schools have fewer students per school than their public counterparts.

Table 5: Market size (2016)

	Students	Market Share
Private	1,657,868	0.77
Public (federal)	327,818	0.15
Public (state)	144,673	0.07
Public (municipal)	13,359	0.01

Source: Higher Education Census (INEP, 2016).

We assume a market is a state, reducing the number of degree choices from 21,232 to 786 (on average). This restriction makes choice sets more realistic and reduces the computational burden of estimating and simulating the model. Moreover, only 3.3% of college students choose degrees in other states, meaning that the cost of imposing this restriction does not seem to be large. We also restrict the size of the market to 500,000 students, for computational constraints. So, we exclude the two biggest states (São Paulo and Minas Gerais), which correspond to 30.8% of the sample.

The market is the set of all ENEM test takers. We drop degrees with less than 5 enrollments, or with missing data. We also drop students who chose degrees in other states, or whose data is missing.

When all restrictions are taken into account, the number of students is reduced from 5,480,902 to 3,472,119 (−36.7%) and the number of degrees is reduced from 21,232 to 8,960. Most of the reduction in the number of students comes from dropping the two largest states (30.8%) and the students who chose degrees in other markets (3.3%), meaning that the other sample restrictions have little impact.

As shown in Table (6), the degrees in the sample are similar to those in the population: even though most differences are statistically significant, they are not sizable.

Table 6: Comparison of means: population and sample (degrees)

	Population	Sample	p-value
Prices	489.15	495.93	0.39
Proportion of public	0.27	0.34	0.00
Capacity utilization	0.51	0.56	0.00
Number of observations	21,232	8,960	

Even though degrees in the population and in the sample are very close, the sample has a higher proportion of low-income and low-score students, as shown in Table (7).

Table 7: Comparison of means: population and sample (students)

	Population	Sample	p-value
Proportion of students in the 1st SES Quintile	0.20	0.26	0.00
Proportion of students in the 2nd SES Quintile	0.20	0.22	0.00
Proportion of students in the 3rd SES Quintile	0.20	0.19	0.00
Proportion of students in the 4th SES Quintile	0.20	0.15	0.00
Proportion of students in the 5th SES Quintile	0.20	0.18	0.00
Proportion of students in the 1st Score Quintile	0.20	0.24	0.00
Proportion of students in the 2nd Score Quintile	0.20	0.21	0.00
Proportion of students in the 3rd Score Quintile	0.20	0.18	0.00
Proportion of students in the 4th Score Quintile	0.20	0.19	0.00
Proportion of students in the 5th Score Quintile	0.20	0.18	0.00
Proportion of students not eligible to affirmative action	0.34	0.29	0.00
Proportion of students enrolling in college	0.39	0.29	0.00
Number of observations	5,480,902	3,472,119	

3.2.3 Capacity Utilization and Oversubscription in Public and Private Schools

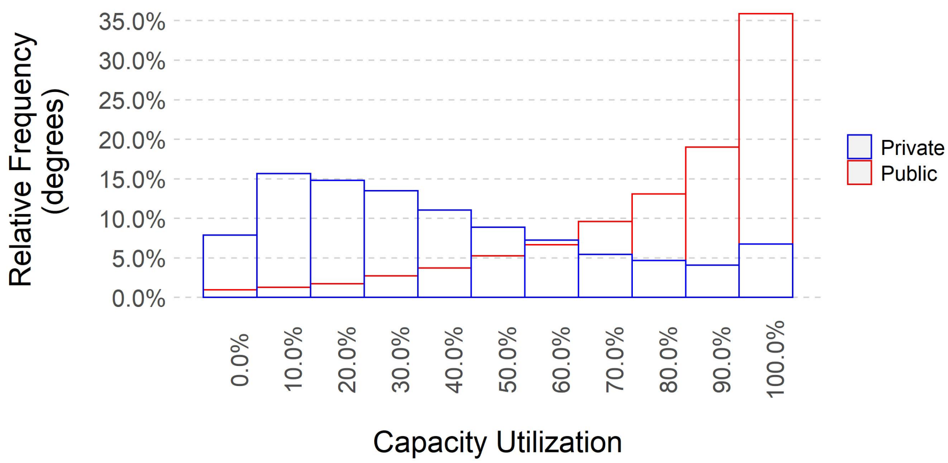
Public schools are competitive and capacity-constrained, but private degrees display widespread idle capacity, as shown in Figure (2).

Furthermore, the enrollment, capacity, and application data in the higher education census suggest that there are approximately 15 candidates per seat in public schools but only 1 candidate per seat in private schools.

In other words, capacity utilization and oversubscription data suggest that public schools are selective, but private schools are not. So, we assume public schools have an active admission process, and private schools admit all students. Therefore, determining the admission process for public schools implies that we know the admission process for the entire market because private schools are not selective.[13]

[13] Previous papers adopted a similar hypothesis, such as Dobbin et al. (2022).

Figure 2: Capacity utilization: public vs. private (2016)



Source: Higher Education Census (INEP, 2016). **Notes:** Proportion of degrees in each capacity utilization group by ownership (public or private) status. Capacity utilization is enrollment divided by the number of reported seats.

3.2.4 Admission in Public Schools

The data suggest that most admissions in public schools are based exclusively on test scores. The higher education census data show that approximately 59% of the students used ENEM scores in their admission process, and less than 8% of all admissions are not based on test scores (ENEM or *vestibulares*). Noteworthy, ENEM is a good proxy for the admission criteria of schools that apply their tests: Estevan et al. (2018) have studied one of the most prestigious Brazilian public universities and shown that scores from *vestibulares* correlate with those from ENEM. Therefore, we use ENEM scores as a proxy for the admission criteria in all schools.

3.3 SES Gap, Subsidy Incidence, and Affirmative Action

Public schools are capacity-constrained, selective, and high-quality. Consequently, they are very attractive, and their admission requirements prevent them from enrolling poor students, who typically have lower scores.

To address this concern, since 2015 the law has required federal public degrees to reserve 50% of their seats to students from public high schools, with sub quotas for student subgroups according to race, income, and disabilities.[14] In fact, public high schools in Brazil are low-quality and typically consumed by poorer students. By focusing on poorer students, subsidies are better targeted to credit-constrained families.

Nevertheless, there is still a sizable gap in the enrollment of high-SES and low-SES students.

To document the gap, we estimate the model in Equation (11) using 2016 data (after the affirmative action reform). In the reduced form, we explain the enrollment of student i in higher education ($Enrollment_i$) by test score dummies ($\tau_{SCORE(i)}$), SES dummies ($\tau_{SES(i)}$), location dummies ($\tau_{LOC(i)}$), and an error term (ϵ_i). We define the SES enrollment gap as the difference between the highest (5th) and lowest (1st) SES dummy estimates. Estimation results are shown in Table (8) and suggest a sizable gap between the two groups, of 19.4 percentage points.

[14] Seats not reserved by affirmative action are called open seats, and the others are called reserved seats.

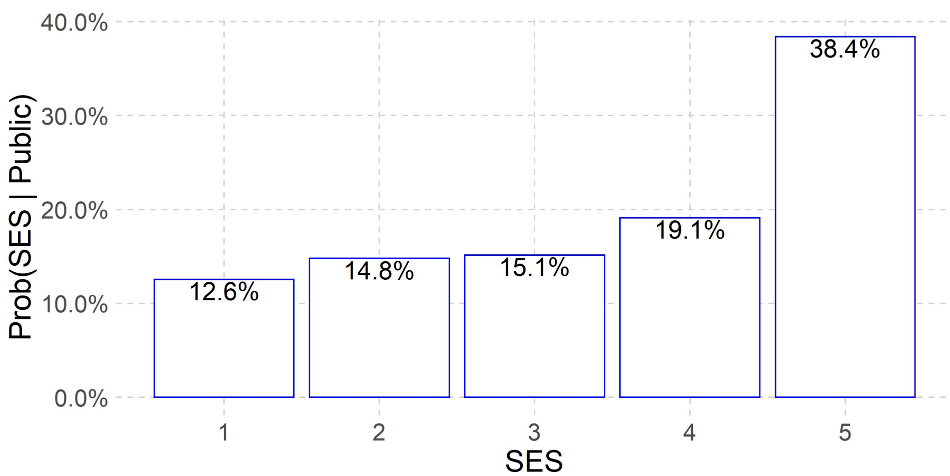
$$Enrollment_i = \tau_{SCORE(i)} + \tau_{SES(i)} + \tau_{LOC(i)} + \epsilon_i \tag{11}$$

Table 8: SES Enrollment Gap

	Model 1	Model 2	Model 3
High-Low SES Gap	0.389	0.253	0.194
Score FE	No	Yes	Yes
Location FE	No	No	Yes

The previous exercise shows a SES-gap for overall enrollment, including private and public schools. In Figure (3) we plot the incidence of subsidies in public higher education by SES group. Results suggest that subsidies reinforce the inequalities the enrollment estimates have shown. Therefore, from a policy perspective, a stronger focus on low-SES students could improve equity in the market.

Figure 3: Subsidy incidence: public schools (2016)



Source: Higher Education Census (INEP, 2016) and ENEM (INEP, 2015).

4 Model

The national market consists of a set of regional markets ($t \in T$) in which colleges ($j \in J_t$) supply degrees ($d \in D_t$) to students ($i \in I_t$). To simplify the exposition, we drop the market index and not present the model with the counterfactual subsidy that we introduce in the section that discusses the policy counterfactuals.

Schools can be private or public. Private schools do not face capacity constraints and admit all students. They hold a portfolio of degrees and choose the prices of all degrees simultaneously, maximizing profits. On the other hand, public schools do not charge tuition, face capacity constraints, and admit students using cutoff scores. Federal public schools have an affirmative action policy, so they have one cutoff for open seats and another for reserved seats. State and municipal schools do not follow affirmative action and have only one cutoff score. Cutoffs are adjusted to make demand equal to the number of available seats.

Students are high school graduates characterized by their socioeconomic status quintile (x_i^{SES}), test scores percentile (x_i^{SCORE}), their distance to all degrees in the market ($\{d_{id}\}_{\forall d}$), and whether they are eligible for reserved seats ($x_i^{AA} \in \{0, 1\}$). When $x_i^{AA} = 1$, student i is eligible for reserved seats; when

$x_i^{AA} = 0$, she is not. Students choose degrees in a random utility model with heterogeneous choice sets: the choice set for each student depends on her test scores and eligibility for affirmative action. When choosing degrees, students observe the prices and cutoffs of all options available in their market.

The equilibrium is a Nash equilibrium on prices and admission scores.

4.1 Preferences

Individual's i indirect latent utility from choosing degree d (u_{id}) is shown in equation (12). It depends on students' characteristics (X_i),[15] degree prices (p_d), unobserved degree attractiveness (ξ_d), demand parameters (θ), and an extreme value type I random variable that represents a taste shock, i.i.d. across students and degrees (ϵ_{id}).

Price sensitivity depends on SES (α^j_{SES}) and test scores (α_{SCORE}). The specification aims to capture the effect of ability (test scores) and credit constraints (SES status) on price elasticities. Students with lower SES will be more likely to be credit-constrained, so we expect they are more price-elastic. On the other hand, students with higher test scores will be more likely to experience higher returns from higher education, so we expect that they are less price-elastic.[16]

Students demand higher education by choosing one of the degrees available or the outside option (not going to college, indexed by $d = 0$). We assume that $p_0 = 0$, $d_{i0} = 0$, and $\xi_0 = 0$, so the utility of the outside option is given by $u_{i0} = \epsilon_{i0}$.

$$\begin{aligned}
u_{id} &= V_d(X_i, p_d; \xi_d, \theta) + \epsilon_{id} \\
V_d(X_i, p_d; \xi_d, \theta) &= \left[-\sum_{j=1}^{5} \alpha^j_{SES} \mathbb{1}_{(x_i^{SES}=j)} - \alpha_{SCORE} x_i^{SCORE} \right] p_d - \beta d_{id} + \xi_d \\
\theta &= (\{\alpha^j_{SES}\}_{j=1}^5, \alpha_{SCORE}, \beta)
\end{aligned} \tag{12}$$

4.2 Choice sets

We now discuss students' choice sets.

There is a large literature about application-admission mechanisms (Gale and Shapley, 1962; Balinkski and Sönmez, 1999; Agarwal and Somaini, 2018; Bó and Hakimov, 2019). Usually, these papers focus on evaluating algorithms that match students and schools. For example, the algorithm starts with students stating their preferences and colleges making initial admission decisions based on a priority index. Schools are limited by capacity and only one student is assigned per school. Because of capacity constraints, many students are not admitted to their first option, so the mechanisms' rules determine how colleges' and students' decisions are updated and the stop criteria. After some steps, the mechanism usually converges and admission decisions are fully determined.

In this paper, we are not discussing mechanisms that make the admission process converge; we assume that admissions are based on cutoff scores (λ_d), chosen by public schools, and we are looking for the equilibrium

[15] The set of individual characteristics is represented by $X_i : (x_i^{SES}, x_i^{SCORE}, x_i^{AA}, \{d_{id}\}_{\forall d})$. $x_i^{SES} \in \{1, 2, 3, 4, 5\}$, $x_i^{SCORE} \in \{0, 1, ...100\}$, $x_i^{AA} \in \{0, 1\}$, and $\{d_{id}\}_{\forall d}$ is finite because distances are calculated between combined statistical areas. Because we restrict the market to a state, distances between students and all degrees available typically assume less than 10 values for each student. Therefore, X_i is discrete.

[16] The relationship between price sensitivity and test scores is linear to avoid identification issues that could arise from the fact that test scores are also being used to determine choice sets (Fack et al., 2019).

cutoff scores that maximize the utility of public schools given the prices of private schools and students' decisions. Degrees with affirmative action policy choose two cutoffs ($\lambda_d = (\lambda_d^0, \lambda_d^1)$)): λ_d^0 for open seats and λ_d^1 for reserved seats. Degrees without affirmative action policy choose only the cutoff for open seats and $\lambda_d^1 = \lambda_d^0$.[17]

To implement the admissions policy, degrees choose a function for admissions $\mu_d : \mathbb{X} \rightarrow \{0, 1\}$ that assumes value 0 if the student is not admitted by degree d and 1 otherwise. Admissions are based only on observable characteristics, and cutoffs depend on affirmative action status: admission score for open seats is λ_d^0, and for reserved seats is λ_d^1. To simplify the notation, let us define $\lambda_d = (\lambda_d^0, \lambda_d^1)$. Therefore, the function for admissions follows equation (13).

$$\mu_d(X_i) = \begin{cases} \mu(X_i; \lambda_d^0) = \mathbb{1}(x_i^{SCORE} \geq \lambda_d^0) \text{ if } x_i^{AA} = 0 \\ \mu(X_i; \lambda_d^1) = \mathbb{1}(x_i^{SCORE} \geq \lambda_d^1) \text{ if } x_i^{AA} = 1 \end{cases} \tag{13}$$

Therefore, the choice set of student i is given by equation (14).

$$C(X_i) = \{d : \mu_d(X_i) = 1\} \tag{14}$$

4.3 Demand

Given admission scores and prices, observed by the students, the probability that degree d is chosen by student i (s_d) follows equation (15).[18] [19]

$$s_d(X_i, p_d, p_{-d}, \lambda_d, \lambda_{-d}; \xi, \theta) = \frac{\mu_d(X_i; \lambda_d^{x_i^{AA}}) exp(V_d(X_i, p_d; \xi_d, \theta))}{1 + \sum_k \mu(X_i; \lambda_k^{x_i^{AA}}) exp(V_k(X_i, p_k; \xi_k, \theta))} \tag{15}$$

Defining $M(X_i)$ as the proportion of students with characteristics X_i, market shares (S_d) follow equation (16).

$$S_d(p_d, p_{-d}, \lambda_d, \lambda_{-d}; \xi, \theta, M) = \sum_{X_i} M(X_i) s_d(X_i, p_d, p_{-d}, \lambda_d, \lambda_{-d}; \xi, \theta) \tag{16}$$

4.4 Applications

Student choices are defined by n_{id}, shown in equation (17). n_{id} equals 1 when student i chooses degree d and 0 otherwise. Because choices are linked to realized utility levels, n_{id} is also a function of ϵ_i. Students can only choose schools that would accept them.

$$n_{id}(\lambda_d^{x_i^{AA}}, p_d, X_i, \epsilon_i; \lambda_{-d}^{x_i^{AA}}, p_{-d}, \xi, \theta) = \mu(X_i; \lambda_d^{x_i^{AA}}) \mathbb{1}(u_{id} > u_{iw} \, \forall w \neq d \,|\, \mu(X_i; \lambda_w^{x_i^{AA}}) = 1) \tag{17}$$

[17] Because $x_i^{SCORE} \in \{0, 1, 2, ..., 100\}$, $\lambda_d^0 \in \{0, 1, 2, ..., 100\}$ and $\lambda_d^1 \in \{0, 1, 2, ..., 100\}$.

[18] The formulation is equivalent to heterogeneous and exogenous choice sets, following Fack et al. (2019).

[19] $\xi = \{\xi_d\}_{\forall d}$.

Cutoff scores for degree d depend on the number and characteristics of students choosing it. We define the set of enrollments in degree d by A_d, the set of enrollments in degree d from students not eligible for affirmative action by A_d^0, and the set of enrollments in degree d from students eligible for affirmative action by A_d^1. The definitions are shown in equation (18).[20] The fact that n_{id} depends on ϵ_i implies that enrollment sets depend on a demographic variable, $\mathbb{M} = (X_i, \epsilon_i)_{\forall i}$.

$$\begin{aligned} A_d^0 &= \{i : n_{id}(\lambda_d^0, p_d, X_i, \epsilon_i; \lambda_{-d}^0, p_{-d}, \xi, \theta) = 1 \wedge x_i^{AA} = 0\} \\ A_d^1 &= \{i : n_{id}(\lambda_d^1, p_d, X_i, \epsilon_i; \lambda_{-d}^1, p_{-d}, \xi, \theta) = 1 \wedge x_i^{AA} = 1\} \\ A_d &= A_d^0 \bigcup A_d^1 \end{aligned} \tag{18}$$

Using the definitions from equation (18), we calculate enrollment by group of affirmative action eligibility (N_d^0 and N_d^1) and total enrollment (N_d), as shown in equation (**??**).

$$\begin{aligned} N_d^0(\lambda_d^0, p_d; \lambda_{-d}^0, p_{-d}, \xi, \theta, \mathbb{M}) &= |A_d^0| \\ N_d^1(\lambda_d^1, p_d; \lambda_{-d}^1, p_{-d}, \xi, \theta, \mathbb{M}) &= |A_d^1| \\ N_d(\lambda_d, p_d; \lambda_{-d}, p_{-d}, \xi, \theta, \mathbb{M}) = N_d^0(\lambda_d^0, p_d; \lambda_{-d}^0, p_{-d}, \xi, \theta, \mathbb{M}) &+ N_d^1(\lambda_d^1, p_d; \lambda_{-d}^1, p_{-d}, \xi, \theta, \mathbb{M}) \end{aligned} \tag{19}$$

4.5 Public schools

Following Balinkski and Sönmez (1999), we assume that public schools prefer students with higher test scores. In other words, comparing any two students (i and j), public degree preferences imply that schools prefer i to j if, and only if, $x_i^{SCORE} > x_j^{SCORE}$. We also assume that public degrees prefer any student to no student. Therefore, admission cutoffs are the lowest scores that satisfy two criteria. First, they must satisfy the capacity constraint: enrollment must not exceed degree capacity (K_d). Second, a fixed proportion of seats (Ψ_d) must be reserved for students eligible for affirmative action.

Following the standard in the literature that discusses mechanisms for assigning students and degrees, we write the model assuming that cutoff scores are calculated using the realized utility of applying students (including ϵ_{id}). This is important because capacity constraints limit the realized (not the expected) number of enrolled students. However, when we simulate the model, we assume the market is large enough to ensure that each degree's realized number of students converges to its expected value.

In our model, we show that the only information needed for public degrees to choose the optimal cutoff scores is the set of their applicants' characteristics. Therefore, it is the only information public schools require to make their optimal admission score choices. Noteworthy, we assume that students know the cutoffs of all degrees when making the application decisions, so our model does not specify a mechanism that assigns students and degrees and students make their choices following a standard discrete choice model with heterogeneous choice sets. In other words, students only apply to the degrees they enroll in, and, consequently, we use the terms application and enrollment interchangeably.

The objective of this section is to show how cutoff scores are determined, given prices and students' choices. So, to simplify the exposition, we will express the functions in this subsection as a function of only cutoff scores.

[20] $|A|$ represents the number of elements in A.

4.5.1 Preferences

Following Balinkski and Sönmez (1999), we assume that public schools prefer students with higher test scores and prefer any student to no student. We assume that colleges are strategic players instead of objects to be consumed. These preferences are represented by the objective function shown in equation (20), which we assume is the objective function of public schools.

$$O(\lambda_d^0, \lambda_d^1) = \sum_{i \in A_d} x_i^{SCORE} \tag{20}$$

Public schools maximize their objective function subject to two constraints.

4.5.2 Capacity constraint

The capacity constraint requires that the number of enrolled students (N_d) is smaller than the number of seats (i.e., capacity, K_d). It is defined in equation (21).

$$N_d^0(\lambda_d^0) + N_d^1(\lambda_d^1) \leq K_d \tag{21}$$

4.5.3 Affirmative action constraint

The affirmative action constraint reserves a fixed proportion (Ψ_d) of seats (K_d) for eligible students. To implement affirmative action, degrees are allowed to have a separate cutoff (λ_d^1) for eligible students. To write the affirmative action restriction, we first define a function that provides the actual proportion of students eligible for affirmative action enrolled in degree d (ψ_d), shown in equation (22).

$$\psi_d(\lambda_d^1) = \frac{N_d^1(\lambda_d^1)}{K_d} \tag{22}$$

There are two caveats when introducing the affirmative action constraint into the model. The first is that the model must allow for the proportion of students eligible for affirmative action to be higher than the quota, so $\psi_d(\lambda_d^1) \geq \Psi_d$. When the affirmative action restriction is not binding (i.e., $\psi_d(\lambda_d^1) > \Psi_d$), eligible students are admitted to non-reserved seats, meaning that they face the same cutoff score as the one for students not eligible for affirmative action. In other words, in this situation, all students see the same cutoff score regardless of their affirmative action status. The second is that it may not be feasible to satisfy the affirmative action restriction if the demand for degree d from students eligible for affirmative action is low. In other words, there may be no λ_d^1 that makes $\psi_d(\lambda_d^1) \geq \Psi_d$. So, we rewrite the affirmative action condition stating that the proportion of students eligible for affirmative action must be higher than the minimum value between Ψ_d and the highest possible proportion of affirmative action students $\max_\lambda(\psi_d(\lambda))$. Because ψ_d is strictly decreasing on λ, we know that $\max_\lambda(\psi_d(\lambda)) = \psi_d(0)$.

Equation (23) defines the affirmative action constraint.

$$\min(\Psi_d, \psi_d(0)) \leq \psi_d(\lambda_d^1) \tag{23}$$

Noteworthy, when degrees do not face the affirmative action constraint, $\Psi_d = 0$, and equation (23) is satisfied for all possible parameter values.

4.5.4 Optimization problem

The problem public schools solve is shown in 24.

$$\begin{aligned}
\max_{\{\lambda_d^0,\lambda_d^1\}} & O(\lambda_d^0, \lambda_d^1) \\
\text{s.t.} & \\
N_d^0(\lambda_d^0) + N_d^1(\lambda_d^1) &\leq K_d \\
\min(\Psi_d, \psi_d(0)) &\leq \psi_d(\lambda_d^1)
\end{aligned} \tag{24}$$

4.5.5 Cutoff scores

The solution of the optimization problem shown in 24 is given by equation (9). The proof can be found in Appendix C.

Table 9: Determining cutoff scores

$N_d(0,0)$	$\psi_d(0)$	λ_d^1	λ_d^0	λ_d^R
$< K_d$		0	0	
$\geq K_d$	$< \Psi_d$	0	$N_d(\lambda_d^0, 0) = K_d$	
$\geq K_d$	$\geq \Psi_d$	$\min(\lambda_d^0, \lambda_d^R)$	$N_d(\lambda_d^0, \min(\lambda_d^0, \lambda_d^R)) = K_d$	$\Psi_d = \psi_d(\lambda_d^R)$

Because private schools admit all students, the choice set of student i combines all private degrees available in the market with all public degrees that would admit her. So, to model students' choice sets, we only need to model how public schools admit students.

4.6 Private Schools

We assume private schools admit all students because they do not have capacity constraints: $\mu(X_i; \lambda_d) = 1 \, \forall X_i$. Marginal costs ($c_d$) are fixed, and the profit maximization problem is static, as shown in (25). Private schools choose the prices of all their degrees (D_j) simultaneously, maximizing profits.

The objective of this subsection is to show how prices are determined, given prices of other schools, cutoff scores, and students' choices. So, to simplify the exposition, we will drop the non-price arguments of all demand functions.

$$\max_{\{p_d\}_{d \in D_j}} \sum_{k \in D_j} S_k(p_k)(p_k - c_k) \tag{25}$$

The first-order condition that arises from the profit-maximizing problem in (25) is given by equation (26). We assume that private schools observe and take as given the admission scores and prices of other schools. Although they observe cutoff scores, private schools do not observe the application to public degrees, and choose prices maximizing expected profits calculated using the demand curve given by equation (16). So, the behavior of private schools follows the standard procedure in the literature about price setting under monopolistic competition with heterogeneous goods. Therefore, private schools choose prices before students choose private degrees, but at the same time public schools choose admission scores, and students apply to public schools.

$$S_d(p_d) + \sum_{k \in D_j} (p_k - c_k) \frac{\partial S_k(p_k)}{\partial p_d} = 0 \tag{26}$$

4.7 Equilibrium

The equilibrium is a pure strategy Nash equilibrium on prices and admission scores. The equilibrium is a vector of prices ($\mathbb{P} = \{p_d\}_{\forall d}$) and cutoff scores ($\Lambda = \{\lambda_d\}_{\forall d}$) in which students maximize utility, as shown in (12), private schools maximize profits, as shown in (25), and public schools maximize utility taking into account capacity constraints and affirmative action requirements, choosing the cutoff scores as shown in Table (11).

The market timing is divided into three stages: first, ϵ_{id} is realized; second, private schools choose p_d, public schools choose λ_d, and students enroll in public schools, simultaneously; third, students enroll in private schools.

4.7.1 Properties of the student-college match

Now, we discuss some of the properties of the matching between students and schools using the framework from the literature about mechanisms that assign students to colleges. Our objective is to define a class of mechanisms compatible with our model. To do so, we follow Balinkski and Sönmez (1999).

Definition 1 *[From Balinkski and Sönmez (1999)] The matching is fair if students with higher scores get assigned to better schools.*

Definition 2 *[From Balinkski and Sönmez (1999)] The matching is individually rational if no student is assigned to a college worse than the no-college option.*

Definition 3 *[From Balinkski and Sönmez (1999)] A blocking pair exists if (1) at least one student prefers a not-capacity-constrained college to their actual assignment, or (2) at least one student α is assigned to college A, but prefers college B, and college B prefers α to at least one of its students.*

Lemma 1 *[From Balinkski and Sönmez (1999)] The matching is stable if it is individually rational and if there is no blocking pair.*

Lemma 2 *[From Balinkski and Sönmez (1999)] The matching is stable if it is individually rational and if there is no blocking pair.*

Proposition 1 *The model is fair and individually rational*

Proposition 2 *The model is stable*

Proposition 3 *There is only one mechanism that yields the same outcome as the model*

Although we are not discussing the mechanism that schools and students use to converge to the equilibrium, the outcome is the only stable equilibrium possible. From Lemma 1 in Balinkski and Sönmez (1999), we know that admissions based on cutoff scores are fair when students choose degrees maximizing utility. Moreover, because the outside option is available to all students, we know that the resulting student-college equilibrium is individually rational (i.e., no student is assigned to a college worse than the no-college option). Last, the Nash equilibrium on prices and admission scores implies that the resulting equilibrium has no blocking pairs. Therefore, from Lemma 2 in (Balinkski and Sönmez, 1999) we also know that the equilibrium is stable. Because all schools rank students using the same scores, Lemma 2.1 implies that the resulting equilibrium is the only stable equilibrium possible.

4.8 Takeaways from the model

When public degrees reduce prices, public schools increase their market share and capacity-constrained degrees increase admission scores. Several mechanisms affect the optimal pricing of private schools in response to this shock. First, because public schools reduced their prices, private schools are incentivized to reduce their prices in response (direct effect), reducing private markups. If consumers were homogeneous and there were no capacity constraints, the model would imply that lower prices in public schools cause a reduction in private prices as well. Second, price-elastic students switch from private to public schools, reducing the price-elasticity of students attending private degrees. The reduction in the price elasticity of students attending private degrees tends to increase private prices (composition effect). Third, under capacity constraints in public schools, the reduction in public prices increases the admission scores of public degrees, reducing the composition effect because it prevents low-score students (many of them with high price sensitivity) from switching from private to public. Fourth, under capacity constraints in public schools, the rise in admission scores reduces competition between private and public schools (public schools are no longer in the choice set of low-score students), increasing private prices (a choice set effect).

The choice set effect is similar to the effect found in (Fillmore, 2023), in which the information about the number of schools to which students were applying affected the prices private schools charged the students. The composition effect discussed in the previous paragraph is similar to the effects discussed in (Dobbin et al., 2022) and in many other papers that discuss similar topics on different markets.

4.9 Limitations of the model

Students observe cutoff scores. We assume that students and schools know cutoff scores. In practice, the centralized admissions system follows the iterative acceptance algorithm. Students make one or two applications (first and second choices) and choose the type of seat for which they are competing. Then, cutoff scores are calculated and shown to students on a daily basis for (usually) four days. During this period, students can change their choices. After the centralized system closes, students enroll, and the remaining seats are allocated using waitlists. Most seats are allocated through waitlists, suggesting that the actual uncertainty of students choosing degrees is much higher than the uncertainty assumed by the model. The major implication of that is that the choice set students actually consider may be lower than the choice set we assume they see.

Reserved seats first. In the model, we assume that schools fill reserved seats first, and open seats later. It is a very common practice in the literature of affirmative action and student matching. However, the implementation of the affirmative action law and the centralized admission system is different from that. In practice, assignment to open and reserved seats occurs separately. In practice, this creates two unexpected problems. First, students competing for open seats may face lower cutoff scores than those competing for reserved seats, as shown by Aygün and Bó (2021). This is the case when demand from students eligible for affirmative action is much higher than demand from students not eligible for affirmative action. Second, when students competing for one type of seat have low demand, degrees may end up with idle capacity.

Since 2024, a change in the law has made students who chose to compete for reserved seats compete for open seats first, eliminating the possibility that they face higher cutoff scores than the students who chose to compete for open seats. However, if there is low demand from students eligible for affirmative action, their seats cannot be occupied by students who are not eligible. In other words, changes in the law made the implementation of affirmative action closer to the mechanism of filling reserved seats first, but there are still important operational differences.

We chose not to model the admission process separately for two reasons.

First, it would add unnecessary complexity to the model. Students eligible for affirmative action could apply for open or reserved seats, and students not eligible for affirmative action could only apply for open seats. Then, assuming that students have no specific preference for being admitted through quotas or not, given that students know the cutoff scores in our model they would choose the type of seat with a lower cutoff and the final allocation would be identical to that of our model.

Second, it would generate idle capacity in public schools. In this case, the affirmative action restriction would be rewritten as $N_d^0(\lambda_d^0) \leq (1-\Psi_d)K_d$ and the capacity restriction would still be $N_d^0(\lambda_d^0)+N_d^1(\lambda_d^1) \leq K_d$. Then, a very small demand from students eligible for affirmative action (i.e., $N_d^1(\lambda_d^1) < \Psi_d K_d \ \forall \lambda_d^1$) will not be compensated by a lower cutoff score for open seats and the capacity constraint will not be binding in the equilibrium $(\lambda_d^0, \lambda_d^1)$.

First, our choice for not incorporating into the model the fact that the admission processes are separated does not change substantially how cutoff scores are estimated. For almost capacity-constrained degrees (capacity utilization between 0.75 and 1.00), making $N_d^0(\lambda_d^0) \leq (1-\Psi_d)K_d$ or $\psi_d(\lambda_d^1) \geq \min(\psi_d(0), \Psi_d)$ does not affect the estimator of the cutoff score for open or reserved seats. However, if demand from affirmative action students is very low and demand from students not eligible for affirmative action is very high, we may end up in a situation in which we assume that cutoff scores for open and reserved seats is 0 (because capacity is not constrained) and the cutoff score for open seats is positive. Even though this possibility is theoretically possible, it does not seem to be empirically relevant.

However, making $N_d^0(\lambda_d^0) \leq (1-\Psi_d)K_d$ would affect the counterfactual simulations. When counterfactual policies change the relative demand between students eligible and not eligible for affirmative action, making $N_d^0(\lambda_d^0) \leq (1-\Psi_d)K_d$ would make public schools keep idle seats even though there is demand for them. So, our counterfactual simulations must be interpreted as in a model in which affirmative action is implemented as reserved seats first.

5 Empirical Strategy

In this section we show the strategy to estimate cutoff scores (λ_d), demand parameters (θ and ξ_d), and marginal costs (c_d).

5.1 Capacity

To do so, we must observe the sample analog of the set of the characteristics of students enrolled in each degree (A_d), whether the degree is capacity-constrained ($K_d = N_d$) or not ($K_d > N_d$), total enrollment (N_d), the proportion of enrolled students eligible for affirmative action (ψ_d), and the required proportion of students eligible for affirmative action (Ψ_d).

We observe A_d, N_d, and ψ_d. For state and municipal schools, we assume there are no reserved seats ($\Psi_d = 0.0$), but 50% of seats are reserved in federal schools ($\Psi_d = 0.5$), as stated by the affirmative action law (that imposes the restriction only to federal degrees). We also observe the number of available seats in each degree (Q_d).

We assume the existence of market frictions that allow degrees to be capacity-constrained when their capacity utilization level is below 100%. So, we now discuss the relationship between Q_d (reported capacity) and K_d (actual capacity).

We assume that a degree is capacity-constrained when there is excess demand. It is possible that capacity utilization is below 100% and there is excess demand if the mechanism matching students and degrees is not efficient enough to guarantee that all students who would like to choose a degree to which they would be admitted do enroll.

This discussion is important because oversubscription and capacity utilization data suggest different levels of capacity constraints in public schools. Even though many public degrees display capacity utilization levels very close to 100%, only one-third operate at the maximum level, and a possible explanation for this is the existence of market frictions that prevent all seats in public schools from being filled.

To analyze the likelihood of these frictions, we analyze data from the centralized system for admission in federal schools, SISU.

One of the main weaknesses of SISU is that it does not encompass all schools in the market. Therefore, many students who report a degree as their first choice may prefer a non-listed degree. Consequently, many admitted students do not enroll and college admissions depend on waitlists. However, finding waitlisted students can be challenging and generate market frictions, as (Kapor et al., 2022) documented for the Chilean case.

Table (10) reports data about the SISU admission process during the beginning of 2018.[21] Five facts can be inferred from the table. First, SISU is sizable: the number of applying students during the first semester of 2018 is close to the total number of students enrolling in college in 2016 (Table 5), and the first edition of SISU in 2018 accounts for more than 40% of enrollment in state and federal degrees in 2016. Second, most students make both first and second choices. After all, the number of applications is nearly double that of students. Third, waitlists are an important admission mechanism, accounting for more than 86% of all admissions. Fourth, oversubscription is considerable: for each admitted student,[22] a degree has, on average, 9.6 students marking it as their first option. Fifth, despite the large oversubscription, many admitted students do not enroll, and many public degrees show some idle capacity (47% of degrees do not fill all their open seats).

Table 10: SISU Statistics (1st round, 2018)

	Data
Applications	3,876,778
Students	1,990,607
Admitted (regular)	220,970
Admitted (waitlist)	1,396,168
Enrollment	193,917
Applications (only 1st choice) / Regular admissions, by degree (mean)	9.655
Proportion of degrees that do not fill all open seats	0.475

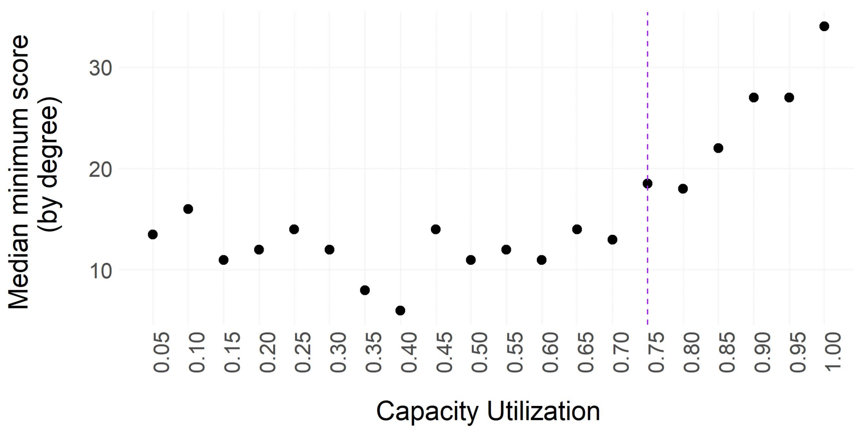
Source: Ministry of Education (2018).

[21] We do not have the complete data for waitlist admissions in the 2016 and 2017 editions.

[22] Regular admissions, not counting admissions from the waitlists.

Therefore, oversubscription is strong even in degrees with capacity utilization below 100%. To find a level above which it would be reasonable to assume that capacity is constrained, we explore the fact that capacity constraints increase admission cutoffs. The higher the excess demand, the higher the cutoff. The higher the cutoffs, the higher should be the lowest score among admitted students. In Figure (4), we plot the median of the minimum score among admitted students by groups of capacity utilization.[23] Figure (4) suggests that the lowest score among admitted students gets higher when capacity utilization reaches 0.75.

Figure 4: Median minimum scores, by degree, in each capacity utilization group



Therefore, we assume that market frictions make public degrees capacity-constrained when their capacity utilization level reaches 75%. To incorporate the assumption in the model, we proceed as follows. Being Q_d the capacity reported by degree d in the higher education census and N_d the number of students enrolled in d, we assume that capacity (K_d) is given by Equation (27), where κ is the level of capacity utilization above which capacity is constrained. When $N_d > \kappa Q_d$, $K_d = N_d$, meaning that capacity utilization is 100% and the degree is capacity-constrained. When $\kappa = 1$, $K_d = Q_d$ and only degrees reporting full capacity utilization are capacity-constrained; when $\kappa = 0$, $K_d = N_d$ and all degrees are capacity-constraned. Therefore, we make $\kappa = 0.75$.

$$K_d(Q_d, \kappa) = \max(\kappa Q_d, N_d) \tag{27}$$

5.2 Cutoff scores

Cutoff scores could be consistently estimated by making $\lambda_d^j = \min\{x_i^{SCORE} : X_i \in A_d^j\}$. However, we the solution to the problem of public schools in (**??**) gives information about the relationship between the number of enrolled students, N, capacity, K, the affirmative action requirement, Ψ, and the proportion of students eligible for affirmative action, ψ.

Moreover, we show how public schools can determine cutoff scores based only on A_d, which they observe, and still follow the solution from (**??**).

Now, we describe how to estimate scores based on these information and the solution to the problem of public schools.

The higher the cutoff scores, the fewer students enroll. Therefore, we define feasible cutoffs as those for which enrollment is lower than capacity. Because public schools prefer any student to no student, their

[23]Each group corresponds to a 5p.p. interval of the capacity utilization rate.

preferences imply they choose the lowest feasible cutoffs, to maximize the number of enrolling students.

Consequently, when capacity is not constrained, degrees admit all students. Because the minimum score is 1 (equivalent to the first percentile of the test score distribution), degrees not capacity constrained make $\lambda_d^0 = \lambda_d^1 = \min\{x_i^{SCORE}\} = 1$.

When capacity is constrained, degrees do not admit all students, and admission scores depend on affirmative action policy. A useful way to explain the cutoff scores in each case is to discuss the lowest score a new student must have to be admitted.

When capacity is constrained and $\Psi_d = 0$, the fact that public schools prefer students with higher test scores makes scores the only admission criterion. Therefore, they make $\lambda_d^0 = \lambda_d^1$. In this case, an applying student would only be admitted if she had higher scores than the lowest score among the admitted students. After all, schools could refuse admission to the lowest-ranking student and admit her, increasing their utility. Therefore: $\lambda_d^0 = \lambda_d^1 = \min\{x_i^{SCORE} : X_i \in A_d\}$.

When capacity is constrained and $\Psi_d > 0$, admission is based on test scores and affirmative action eligibility. First, let us consider the cutoff for reserved seats. We assume that affirmative action policy requires that an applying student eligible for affirmative action always be admitted to a degree that is capacity-constrained and has a proportion of enrolled students eligible for affirmative action (ψ_d) lower than the objective (Ψ_d). Therefore, the cutoff score for reserved seats when $\psi_d < \Psi_d$ is the lowest score possible. In other words: $\lambda_d^1 = \min\{x_i^{SCORE}\} = 1$. On the other hand, if the proportion of enrolled students eligible for affirmative action (ψ_d) is equal to or higher than the objective (Ψ_d), an applying student eligible for affirmative action must have scores higher than the lowest-ranking enrolled student to be admitted. After all, when the affirmative action objective is achieved ($\psi_d \geq \Psi_d$), the admission of another student eligible for affirmative action would still guarantee that $\psi_d \geq \Psi_d$, so schools choose students based only on test scores (their preferences). In this case, schools could deny admission to the lowest-ranking enrolled student (regardless of her affirmative action status) and grant admission to the new candidate, with higher scores, increasing its utility. Therefore, the cutoff score in this situation is $\lambda_d^1 = \min\{x_i^{SCORE} : X_i \in A_d\}$.

Now, let us consider the cutoff for open seats when capacity is constrained and $\Psi_d > 0$. We assume that when the current proportion of enrolled students eligible for affirmative action is equal to or lower than the requirement ($\psi_d \leq \Psi_d$), an applying student not eligible for affirmative action cannot take the seat of a student eligible for affirmative action, because it would bring the degree farther from the objective. So, to be admitted, she has to have scored higher than the lowest-ranking student among those not eligible for affirmative action. In other words: $\lambda_d^0 = \min\{x_i^{SCORE} : X_i \in A_d^0\}$. However, when the proportion of enrolled students eligible for affirmative action is higher than the requirement ($\psi_d > \Psi_d$), a student not eligible for affirmative action can take the seat of a student eligible for affirmative action without violating the affirmative action objective. So, in this case, to be admitted an applying student must have scored higher than the lowest-ranking student in the class: $\lambda_d^0 = \min\{x_i^{SCORE} : X_i \in A_d\}$. Noteworthy, in this case $\lambda_d^0 = \lambda_d^1$.

Table (11) summarizes the discussion about cutoff scores in each scenario. Noteworthy, schools only need to observe the applications they receive (A_d) to determine their optimal cutoffs in every scenario, and we assume they do.

Therefore, if schools observe the set of applying students, A_d, they can use K_d and Ψ_d to calculate N_d and ψ_d. Then, they can determine cutoff scores based on Table (11). In other words, the only information schools need to choose scores is the set of applying students and we use the sample analog of the formulas in Table (11) to estimate cutoff scores.

Table 11: Determining cutoff scores

N_d	Ψ_d	ψ_d	λ_d^0	λ_d^1
$< K_d$			$min(x_i^{SCORE})$	$min(x_i^{SCORE})$
$= K_d$	$= 0$		$min(x_i^{SCORE} : X_i \in A_d)$	$min(x_i^{SCORE} : X_i \in A_d)$
$= K_d$	> 0	$> \Psi_d$	$min(x_i^{SCORE} : X_i \in A_d)$	$min(x_i^{SCORE} : X_i \in A_d)$
$= K_d$	> 0	$= \Psi_d$	$min(x_i^{SCORE} : X_i \in A_d^0)$	$min(x_i^{SCORE} : X_i \in A_d)$
$= K_d$	> 0	$< \Psi_d$	$min(x_i^{SCORE} : X_i \in A_d^0)$	$min(x_i^{SCORE})$

5.3 Demand parameters

The demand parameters are θ and ξ. Identification of demand parameters in a discrete choice model with cutoffs determining heterogeneous choice sets follows Fack et al. (2019) and estimation follows Berry et al. (2004): for each value of θ, there is a unique value of ξ for which the market shares from the estimated model match their sample analogs. So, we estimate θ using a two-step non-linear GMM.

Using the estimated cutoff scores, we assume that the choice set for student i is $\{d : x_i^{SCORE} \geq \hat{\lambda}_d^{x_i^{AA}}\}$.

The first set of moments used for estimation is an exclusion restriction that explores the properties of our cost shifter. To do so, we incorporate fixed effects for major ($r_d \in R$)[24] and ownership status ($h_d \in H$)[25] on degree attractiveness. We do so by calculating the residual degree attractiveness ($\Delta\xi_d$), as shown in Equation (28).

$$\xi_d = h_d + r_d + \Delta\xi_d \tag{28}$$

Because the econometrician does not observe $\Delta\xi_d$, but schools and students do, estimating θ requires instrumental variables that deal with the potential correlation between p_d and $\Delta\xi_d$. Our instrumental variables strategy for degree d relies on the regional variation of wages associated with occupations that are important inputs to higher education supply. We use eight cost shifters representing the average hourly wages paid by higher education institutions in the same location[26] of degree d. The first instrument uses all occupations in the sample. Each of the other seven instruments uses the hourly wages of faculty in each of the following departments: (1) Mathematics, Statistics, and Computer Science; (2) Architeture, Urbanism, Engineering, Geology, and Geophysics; (3) Biology and Health; (4) Pedagogy; (5) Language and Literature; (6) Humanities; (7) Economics, Business, and Accounting.

We include average wages by department to allow our instrument to isolate differences in the hourly cost of hiring professors in each region. We implicitly assume that those costs are driven by supply-side restrictions, such as the availability of professors in each location.

Defining Z as the matrix in which columns index cost shifters and rows index degrees, and $\Delta\xi$ as a vector for residual degree attractiveness, the moment condition for the cost shifter is summarized in Equation (29).[27]

[24] R is the set of majors.
[25] $H = \{Public, Private\}$.
[26] Combined statistical area.
[27] $\overrightarrow{0}$ is a vector of zeros.

$$E[Z^T \Delta\xi] = \overrightarrow{0} \tag{29}$$

Following Berry et al. (2004), the second set of moments used for estimation makes the model and the sample closer in several dimensions: the proportion of students in each SES group (by degree), the proportion of students eligible for affirmative action (by degree), the average students' location (by degree), and the average test scores (by degree). These dimensions are represented by g[28] and their corresponding moment conditions are shown in Equation (30).

$$\begin{aligned} E\left[\mathbb{1}_{(r_d=r)}\left[\frac{\sum_{X_i} M(X_i)s_d(X_i,p_d,p_{-d},\lambda_d,\lambda_{-d};\xi,\theta)g(X_i)}{S_d(p_d,p_{-d},\lambda_d,\lambda_{-d};\xi,\theta,M)} - \frac{\sum_{X_i\in A_d} g(X_i)}{\sum_{X_i\in A_d} 1}\right]\right] = 0 \ \forall g, r \\ E\left[\mathbb{1}_{(h_d=h)}\left[\frac{\sum_{X_i} M(X_i)s_d(X_i,p_d,p_{-d},\lambda_d,\lambda_{-d};\xi,\theta)g(X_i)}{S_d(p_d,p_{-d},\lambda_d,\lambda_{-d};\xi,\theta,M)} - \frac{\sum_{X_i\in A_d} g(X_i)}{\sum_{X_i\in A_d} 1}\right]\right] = 0 \ \forall g, h \end{aligned} \tag{30}$$

There are two major threats to our identification strategy. The first is that local wages may not be strongly correlated with degree prices. The second is that our instrument may be correlated with demand shocks. For example, if some regions have higher total factor productivity, they may simultaneously have higher wages and higher education demand. In this case, $E[Z^T\Delta\xi]$ would be positive (instead of zero).

We estimate a homogeneous demand model to evaluate these threats. To do so, we reestimate the model assuming $u_{id} = \alpha p_d + h_d + r_d + \Delta\xi_d + \epsilon_{id}$. We do so for two reasons: first, the homogeneous demand model can be written in a reduced form that allows the comparison of OLS and 2SLS results; second, first-stage estimates provide useful information about the inclusion restrictions. Results (reported in the Appendix) show that the OLS estimate for the price sensitivity component is not statistically significant, but the 2SLS estimate is negative and statistically different from zero, ruling out the possibility that $E[Z^T\Delta\xi] > 0$. Moreover, first-stage results suggest that our instrument satisfies the inclusion restriction.

5.4 Marginal costs

Marginal costs are estimated using the demand curve and the first-order condition of private degrees, shown in Equation (**??**). We do not estimate marginal costs for public schools, assuming that the costs for the public sector depend on capacity, not enrollment. Consequently, the cost for public schools is the same in the baseline and all counterfactual simulations, since our counterfactual policies do not change public school capacity.[29]

$$S_d(p_d,p_{-d},\lambda_d,\lambda_{-d};\xi,\theta,M) + \sum_{k\in D_j}(p_k - c_k)\frac{\partial S_k(p_k,p_{-k},\lambda_k,\lambda_{-k};\xi,\theta,M)}{\partial p_d} = 0 \tag{31}$$

6 Results

In this section, we discuss model estimates for preference and supply parameters.

[28] $g(X_i)$: $\mathbb{1}_{(x_i^{SES}=j)}$, x_i^{AA}, x_i^{SCORES}, or d_{id}

[29] Some of our simulations assume public schools have no capacity constraints, but they are only used to decompose the overall effect of counterfactual policies.

6.1 Preferences

Table (12) shows demand estimates. All coefficients are statistically significant and have the expected signs. As expected, the price sensitivity coefficient becomes closer to zero as the socioeconomic status increases (as shown by $\hat{\alpha}^1_{SES}$ to $\hat{\alpha}^5_{SES}$). Even though the point estimates suggest that the fourth socioeconomic status ($\hat{\alpha}^4_{SES}$) has a smaller price sensitivity than the fifth ($\hat{\alpha}^5_{SES}$), the differences are not statistically significant. Higher scores also imply a lower price sensitivity ($\hat{\alpha}_{SCORE}$). Moreover, the distance between students and degrees strongly negatively affects indirect utility ($\hat{\beta}$).

Table 12: Demand estimates

	Estimates
$\hat{\alpha}^1_{SES}$	6.1389^{***}
	(0.0361)
$\hat{\alpha}^2_{SES}$	5.4576^{***}
	(0.0365)
$\hat{\alpha}^3_{SES}$	5.1869^{***}
	(0.0405)
$\hat{\alpha}^4_{SES}$	4.3425^{***}
	(0.0396)
$\hat{\alpha}^5_{SES}$	4.4627^{***}
	(0.0388)
$\hat{\alpha}_{SCORE}$	-2.4111^{***}
	(0.0407)
$\hat{\beta}$	19.4714^{***}
	(0.1604)
N	8960

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors in parantheses. Test Scores vary between 0 and 1. Distances measured in thousand kilometers. Subsample: 25 states (out of 27)

Equation (32) shows the own-price elasticity of student i (η_d). Figure (5) shows the mean of the price elasticities of students at each point of the SES distribution. Even though $\hat{\alpha}^4_{SES}$ is slightly lower than $\hat{\alpha}^5_{SES}$, the average own-price elasticity of students from the 4^{th} SES group is higher than that of the students from the 5^{th}, as expected.

$$\eta_d(X_i, p_d, p_{-d}, \lambda_d, \lambda_{-d}; \xi, \theta) = \alpha_{SES}^{x_i^{SES}}\, p_d(1 - s_d(X_i, p_d, p_{-d}, \lambda_d, \lambda_{-d}; \xi, \theta)) \tag{32}$$

6.2 Markups

Markup for degree d (b_d) follows Equation (33), and Figure (6) shows markup estimates. The median markup is 0.44.

$$b_d = \frac{p_d - c_d}{p_d} \tag{33}$$

Figure 5: Mean own-price elasticity by SES group

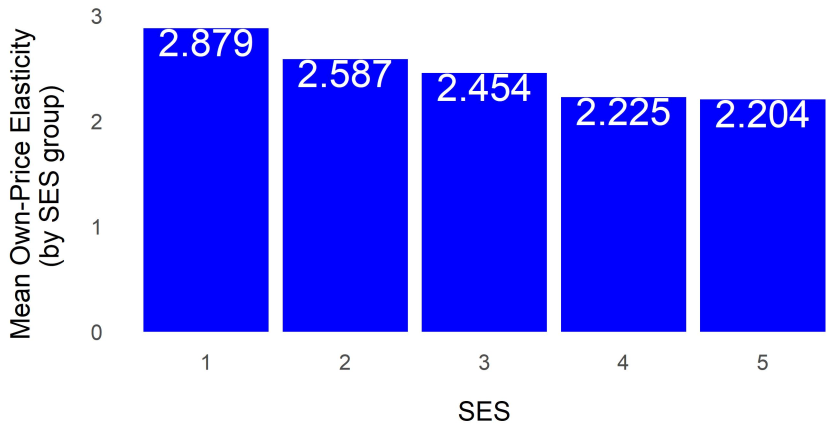
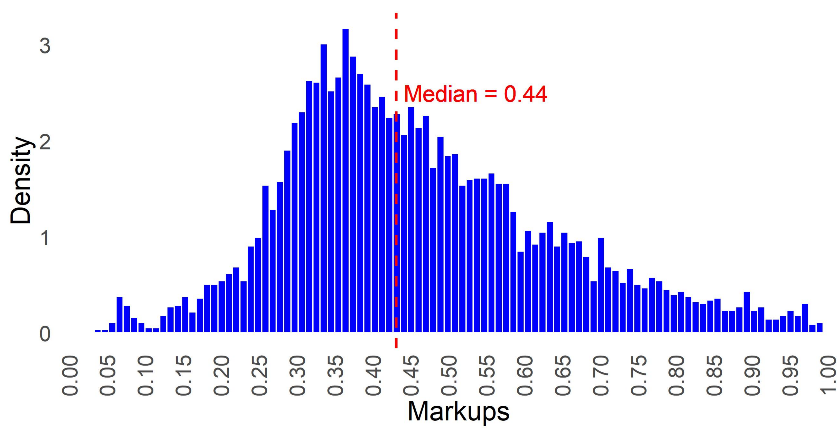


Figure 6: Histogram of markup estimates (by degree)



7 Counterfactuals

Using the model, we simulate two types of counterfactual policies: conditional and unconditional subsidies. Furthermore, subsidies are flat to reduce the pass-through of unconditional subsidies to private prices (Sahai, 2023).

A subsidy design w is defined by the vector $\Theta_w = (T_w, P_w, t_w)$. T_w is the stipend value, P_w is the set of subsidized schools, and t_w is the price of public schools.

Under the subsidy design w, degree d charges $\tilde{p}_{dw}$.[30] On the other hand, the price students pay to choose degree d is $\tilde{p}^{Std}_{dw}$. Prices chosen by schools do not depend on students' characteristics because we assume no price discrimination. The relationship between the price students pay and the price schools receive is given by equation (34).

The stipend depends on two components. First, the subsidy value: T_w, fixed for all recipients. Second, the eligibility status for degrees: when subsidies are unconditional, the set of subsidized degrees is the set with all degrees in the market ($P_w = U = \{d : \forall d\}$); when subsidies are conditional, only public degrees receive the subsidy ($P_w = C = \{d : h_d = Public\}$). In the baseline, $w = 0$ ($T_0 = 0, P_0 = C, t_0 = 0$).

[30] Variables with a tilde are those calculated under a counterfactual subsidy design.

$$\tilde{p}_{dw}^{Std}(\tilde{p}_{dw}) = \tilde{p}_{dw} - B_w \mathbb{1}(d \in P_w) + b_w \mathbb{1}(d \in \text{Public}) \tag{34}$$

Market share for degree d under subsidy design w is given by equation (35).

$$\begin{aligned}
&\tilde{S}_{dw}(\tilde{p}_{dw}, \tilde{p}_{-dw}, \tilde{\lambda}_{dw}, \tilde{\lambda}_{-dw}; \xi, \theta, M) = \sum_{X_i} M(X_i)\tilde{s}_{dw}(X_i, \tilde{p}_{dw}, \tilde{p}_{-dw}, \tilde{\lambda}_{dw}, \tilde{\lambda}_{-dw}; \xi, \theta) \\
&\text{Where:} \\
&\tilde{s}_{dw}(X_i, \tilde{p}_{dw}, \tilde{p}_{-dw}, \tilde{\lambda}_{dw}, \tilde{\lambda}_{-dw}; \xi, \theta) = s_d(X_i, \tilde{p}_{dw}^{Std}(\tilde{p}_{dw}), \tilde{p}_{-dw}^{Std}(\tilde{p}_{-dw}), \tilde{\lambda}_{dw}, \tilde{\lambda}_{-dw}; \xi, \theta)
\end{aligned} \tag{35}$$

To evaluate equity in counterfactual equilibria, we use the high-low SES gap ($\tilde{G}_w$), which is the difference in enrollment between students in the highest SES quintile and students in the lowest SES quintile, as shown by equation (36).

$$\begin{aligned}
\tilde{G}_w = &\frac{\sum_{X_i : x_i^{SES}=5} M(X_i)\tilde{s}_{dw}(X_i, \tilde{p}_{dw}, \tilde{p}_{-dw}, \tilde{\lambda}_{dw}, \tilde{\lambda}_{-dw}; \xi, \theta)}{\sum_{X_i : x_i^{SES}=5} M(X_i)} - \\
&\frac{\sum_{X_i : x_i^{SES}=1} M(X_i)\tilde{s}_{dw}(X_i, \tilde{p}_{dw}, \tilde{p}_{-dw}, \tilde{\lambda}_{dw}, \tilde{\lambda}_{-dw}; \xi, \theta)}{\sum_{X_i : x_i^{SES}=1} M(X_i)}
\end{aligned} \tag{36}$$

Following Small and Rosen (1981), we use equation (37) to compare the consumer surplus changes from each policy reform ($\Delta\tilde{C}_w$).

$$\begin{aligned}
\Delta\tilde{C}_w &= \sum_{X_i} \frac{M(X_i)}{\alpha(X_i; \theta)} \Delta\tilde{c}_w(X_i) \\
\Delta\tilde{c}_w(X_i) &= ln\left(\sum_d \mu(X_i; \tilde{\lambda}_{dw}) \exp(V_d(X_i, \tilde{p}_{dw}^{Std}(\tilde{p}_{dw}); \xi_d, \theta))\right) - \\
&\quad - ln\left(\sum_d \mu(X_i; \tilde{\lambda}_{d0}) \exp(V_d(X_i, \tilde{p}_{d0}^{Std}(\tilde{p}_{d0}); \xi_d, \theta))\right)
\end{aligned} \tag{37}$$

7.1 Evaluating Counterfactuals

Now, we discuss the numerical procedure for calculating counterfactual results. To do so, we use the first-order condition for private schools, shown in equation (**??**), and the optimal choice of cutoff scores, shown in equation (**??**). Then, from a vector of prices and cutoff scores in iteration h ($\mathbf{p^h}, \lambda^{\mathbf{h}}$), we proceed as follows:[31]

1 For each private supplier j, we use the first-order condition for private schools to calculate p_d^{h+1} for all $d \in D_j$, given prices and cutoff scores in iteration h ($\mathbf{p^h}, \lambda^{\mathbf{h}}$)

[31] There is no proof for the uniqueness of the equilibrium. However, simulating the baseline from random starting points for cutoffs and prices suggests the equilibrium is unique. The cutoffs are the same in all random simulations. The difference between prices in each random simulation is never above BRL 2.00, which comes from our convergence parameters. Our threshold for price convergence is BRL 1.00; therefore, the distance between the price outcome of any simulation and the true value is always below BRL 1.00 and the distance between any two simulations should not exceed BRL 2.00 (as it is the case).

2 For each public degree, we use the equation for the optimal choice of cutoff scores to calculate λ_d^{h+1}, given prices and cutoff scores in iteration h ($\mathbf{p^h}, \lambda^{\mathbf{h}}$)

3 Update prices and cutoff scores with $\mathbf{p^{h+1}}$ and $\lambda^{\mathbf{h+1}}$

4 Calculate $e^h = \sum_d |p_d^h - p_d^{h+1}| + \sum_d |\lambda_d^h - \lambda_d^{h+1}|$

5 If $e^h \geq \epsilon^{convergence}$, go back to step 1 and begin a new iteration; if $e^h < \epsilon^{convergence}$, ($\mathbf{p^{h+1}}, \lambda^{\mathbf{h+1}}$) is the new equilibrium

7.2 Capacity Constraints in the Baseline

The first set of counterfactual exercises evaluates the effect of removing capacity constraints in the baseline. The exercise is not a counterfactual policy, but, as we will show, it is an important tool for understanding the role of capacity constraints in the model.

We decompose the impact of subsidies on equilibrium outcome y for degree d under subsidy w ($\tilde{y}_{dw}$) into a direct and an indirect effect, using equation (38). Variables with a *NoCC* superscript are calculated assuming public degrees do not face capacity constraints.[32] The decomposition allows us to understand the role of two mechanisms behind counterfactual results.

The first of the two components is the direct effect (D^y_{dw}), which measures changes in the outcome y for degree d under subsidy w comparing counterfactual equilibrium outcomes without capacity constraints in public degrees to a counterfactual baseline in which public schools do not have capacity constraints. Therefore, the direct effect measures the pure impact of the subsidy, without capacity constraint effects.

The second is the indirect effect. To understand the indirect effect, it is important to define I^y_{dw}, which is how capacity constraints change equilibrium outcome y for degree d under subsidy w $\left(I^y_{dw} = \frac{\tilde{y}_{dw}}{\tilde{y}^{\text{NoCC}}_{dw}}\right)$. The indirect effect is given by $\frac{I^y_{dw}}{I^y_{d0}}$.

$$
\begin{aligned}
&(1 + \text{Effect}) = \frac{\tilde{y}_{dw}}{\tilde{y}_{d0}} = D^y_{dw} \frac{I^y_{dw}}{I^y_{d0}} \\
&D^y_{dw} = \frac{\tilde{y}^{\text{NoCC}}_{dw}}{\tilde{y}^{\text{NoCC}}_{d0}} = 1 + \text{Direct} \\
&\frac{I^y_{dw}}{I^y_{d0}} = 1 + \text{Indirect} \\
&I^y_{dw} = \frac{\tilde{y}_{dw}}{\tilde{y}^{\text{NoCC}}_{dw}} \\
&\text{Effect} = (1 + \text{Direct})(1 + \text{Indirect}) - 1
\end{aligned}
\tag{38}
$$

Removing capacity constraints from the baseline is equivalent to finding I^y_{d0}. Noteworthy, I^y_{d0} is the same for every possible subsidy w. Therefore, understanding the role of capacity constraints in the baseline is important to understanding the indirect effects from subsidy changes. Results from the simulation that removes capacity constraints in public degrees are shown in Table (13).

As expected, removing capacity constraints in public degrees significantly increases public school enrollment (by 114%). Furthermore, many students switch from the outside option, causing total enrollment in higher education to increase by 23%. The expansion of public school capacity also increases public subsidies

[32]We show a multiplicative decomposition, but a straightforward additive decomposition can also be used.

Table 13: Effect of Removing Capacity Constraints

Enrollment	23.42%
SES Gap	5.49pp
Consumer Surplus	25.84
Public Share	9.09pp
Private Share	-2.29pp
Outside Share	-6.80pp
Public Enrollment	113.63%
Private Enrollment	-10.88%
Average Market Prices	-26.43%
Average Private Prices	1.88%
Profit	-9.00%

Notes: Consumer surplus in BRL per student in the market. All results compared to baseline levels.

because public schools do not charge tuition (the amount of the subsidy per student would remain the same, but many more students would benefit). Therefore, it is unsurprising that the enrollment gap for students in the highest and lowest SES groups falls, and that average market prices fall.

Even though public subsidies would be much higher without capacity constraints in public degrees, private prices would rise by 1.9%. The result suggests that the composition effect determines price dynamics: more subsidies to public schools increase the average price elasticity of students attending public degrees and reduce the average price elasticity of those attending private schools. Therefore, more subsidies to public degrees would increase private schools' prices (even though their profit falls by 9.0%). Although the markup of private schools increases, their profits fall, suggesting that capacity constraints in public schools, as expected, increase the profit of their private counterparts.

Results also show an important impact of capacity constraints on consumer surplus. The result is driven by the reduction in average market prices and the increase in students' choice sets (capacity restrictions in public schools were lifted, so now all students are admitted). Therefore, policies that reduce subsidies in public schools would have two main effects: the first, a lower consumer surplus due to higher average prices; the second, a higher consumer surplus due to a smaller level of capacity constraints in public degrees. In the next subsection, we explore this result and analyze the effects of counterfactual policies.

7.3 Counterfactual Policy Outcomes

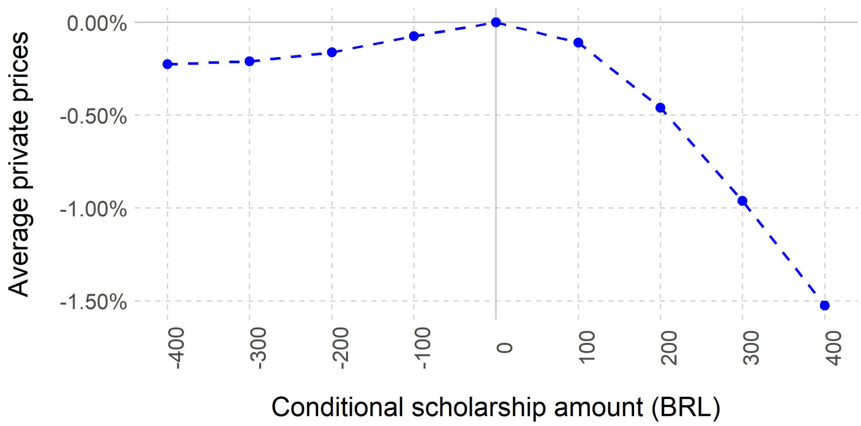
The second set of counterfactual policies analyzes the impact of varying T_w (the scholarship amount) for both types of subsidy: conditional and unconditional. We focus on the price outcomes. In the simulations shown in this subsection, we make $t_w = 0$ (public school prices vary only in the budget-neutral exercises).

7.3.1 Conditional subsidies

First, we focus on conditional subsidies ($P_w = C$). Figure (7) shows the impact of conditional subsidies on private prices, varying T_w. All results are compared to the baseline unless stated otherwise. Two main facts can be inferred from the graph: the impact of changing the subsidy is very close to zero and non-monotonic: public subsidies increase private prices until private prices start going down.

To better understand this effect, Figure (8) shows the relationship between capacity constraints, conditional subsidies, and private prices. Increasing public subsidies increases private prices when the proportion

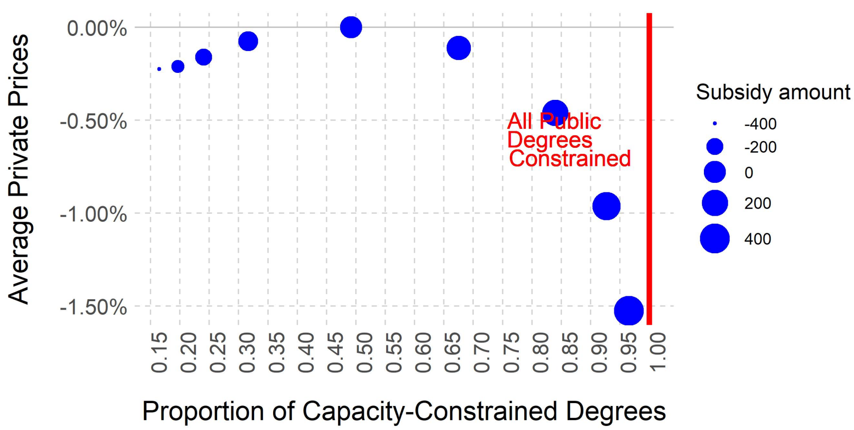
Figure 7: Conditional subsidies and private prices



Notes: median impact on private prices in each counterfactual simulation for conditional subsidies. In the x-axis, we show several possible values for the monthly scholarship; in the y-axis, we show the increase in average private prices.

of capacity-constrained degrees is low. However, the relationship is reversed as the proportion of capacity-constrained degrees becomes closer to its maximum (100%). In other words, capacity constraints are crucial in explaining why the relationship between public subsidies and private prices is not monotonic.

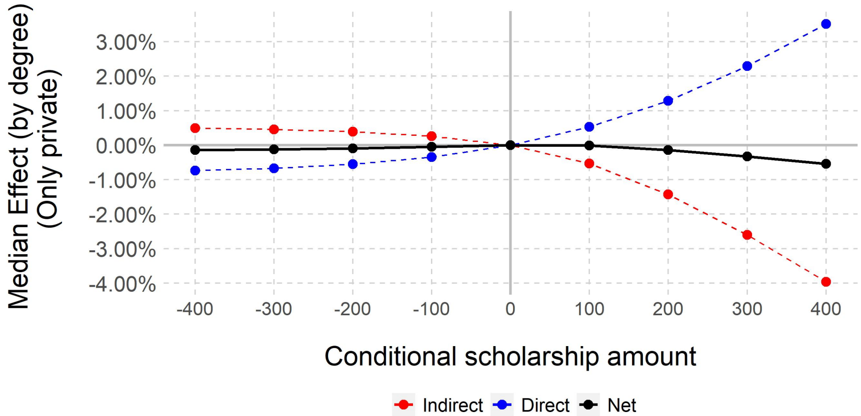
Figure 8: Conditional subsidies, private prices, and capacity constraints



Notes: change in average prices in each counterfactual simulation for conditional subsidies. In the x-axis, we show several possible values for the proportion of capacity-constrained public degrees; in the y-axis, we show the increase in average prices paid by the students in each scenario; the size of the bullets represents the subsidy amount.

The decomposition of price effects into direct and indirect effects can cast more light on the mechanics of the equilibrium results. Results from the decomposition are shown in Figure (9). The direct effect (when public schools have no capacity constraints) from increasing conditional subsidies makes private prices go up. However, the indirect effect goes in the opposite direction. The mechanics behind the result are as follows: more subsidies in public schools increase the level of capacity constraints, raising admission scores and forcing low-score students to switch from public to private schools; because students with lower scores are more price-elastic, this increases the price elasticity of students attending private degrees and reduces prices. Noteworthy, when subsidies to public schools are high, the indirect effect dominates the direct effect and changes the median pass-through direction.

Figure 9: Pass-through decomposition: conditional subsidies



Notes: median impact on private prices in each counterfactual simulation for conditional subsidies. In the x-axis, we show several possible values for the monthly scholarship; in the y-axis, we show the median price increase in each scenario. The figure reports the direct effect (when public schools do not face capacity constraints) and the indirect effect (from changes in the level of capacity constraints in public schools). The direct and indirect effects are defined in equation (38).

We can apply a similar procedure to decompose the effect of conditional subsidies on the average price sensitivity ($PS_{d,w}$) of students attending private degrees.[33] The average price sensitivity of students attending degree d under subsidy design w is given by equation (39). The objective of understanding the behavior of the average price sensitivity under each counterfactual subsidy simulation is to understand to which extent the price effects are being driven by the composition effect. In the literature about the equilibrium effects of subsidies, there are two important drivers of the effect of subsidies on prices: the individual effect (how the price elasticity of the consumers of good A is changed by policy reform) and the composition effect (how more or less price-elastic consumers consume more or less of good A after the reform).

If the composition effect is an important driver of our price responses, we will see the price sensitivity of students attending private degrees and private prices move in opposite directions. Until now, we discussed our equilibrium effects emphasizing the relevance of the composition effect, and the price sensitivity decomposition offers evidence that the composition effect is indeed the main driver of the price equilibrium responses.

$$
\begin{aligned}
&PS_{d,w} = \frac{\sum_{X_i} M(X_i)\tilde{s}_{dw}(X_i, \tilde{p}_{dw}, \tilde{p}_{-dw}, \tilde{\lambda}_{dw}, \tilde{\lambda}_{-dw}; \xi, \theta)\alpha(X_i;\theta)}{\tilde{S}_{dw}(\tilde{p}_{dw}, \tilde{p}_{-dw}, \tilde{\lambda}_{dw}, \tilde{\lambda}_{-dw}; \xi, \theta, M)} \\
&\text{Where:} \\
&\alpha(X_i;\theta) = \sum_{j=1}^{5} \alpha^j_{SES}\mathbb{1}_{(x_i^{SES}=j)} + \alpha_{SCORE}x_i^{SCORE}
\end{aligned}
\tag{39}
$$

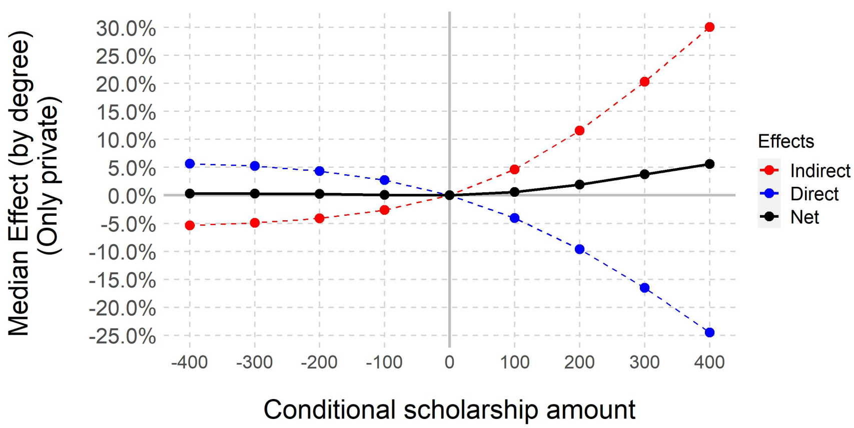
As shown in Figure (10) the direct and indirect effects from increasing conditional subsidies impact the price sensitivity of students in private degrees in opposite ways. On the one hand, the direct effect

[33]Now, the decomposition is additive.

of increasing subsidies makes prices go up and the price sensitivity of students attending private degrees go down, providing evidence that the composition effect is an important driver of the price responses our model generates. Subsidizing public schools incentivizes students with higher price sensitivity to switch from private to public degrees; those who still choose private degrees after the higher subsidy in public schools have lower price sensitivity.

On the other hand, the indirect effect of increasing subsidies makes prices go down and the price sensitivity of students attending private degrees go up, reinforcing the evidence that the composition effect is an important driver of the price responses our model generates. Results suggest that capacity constraints play a vital role in understanding the overall composition effect under conditional subsidies: the sign of the net effect on the average price sensitivity comes from the indirect component in the simulations with higher scholarship amounts. Moreover, the changes in the average price sensitivity of students attending private degrees explain the direction of the changes in private prices.

Figure 10: $PS_{d,w}$ decomposition for each w: conditional subsidies



Notes: median impact on the average price sensitivity of students attending private degrees in each counterfactual simulation for conditional subsidies. In the x-axis, we show several possible values for the monthly scholarship; in the y-axis, we show the median price sensitivity changes in each scenario. The figure reports the direct effect (when public schools do not face capacity constraints) and the indirect effect (from changes in the level of capacity constraints in public schools). The direct and indirect effects are defined in equation (38), but are additive (instead of multiplicative).

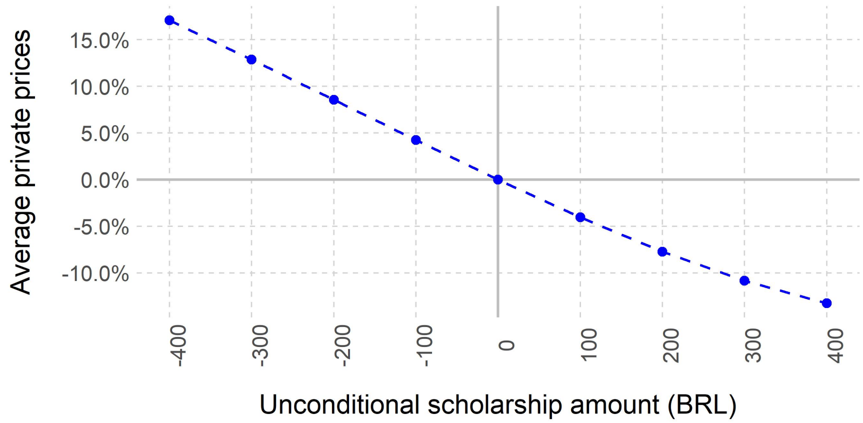
7.3.2 Unconditional subsidies

The effects of unconditional subsidies ($P_w = U$) on average private prices are shown in Figure (11). When subsidies are unconditional, they reduce private prices. For all simulated policies, the relationship between these two variables is monotonic.

As shown in Figure (12), the higher subsidy levels are not enough to make all degrees capacity-constrained.

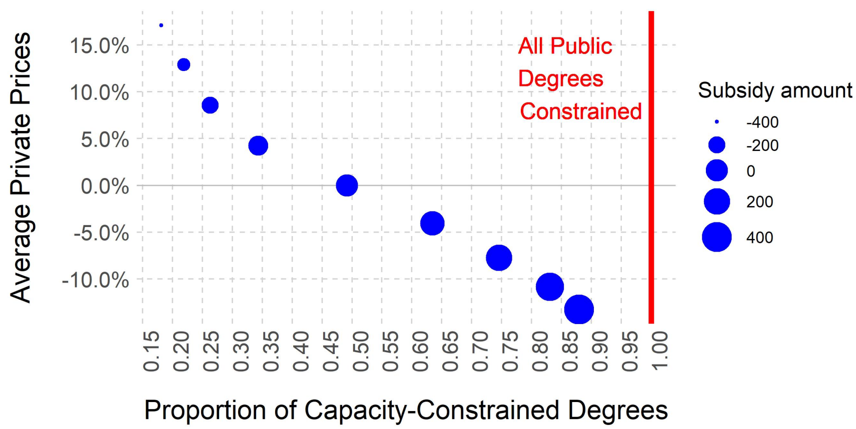
In Figure (13), we decompose the effect of subsidies on the median pass-through using equation (38). Similarly to what was observed under conditional subsidies, the indirect effect of increasing subsidies reduces private prices. On the other hand, when subsidies increase, the direct effect reduces private prices, opposing the result of the conditional subsidy simulations. In a world without capacity constraints, unconditional subsidies reduce private prices and conditional subsidies increase them. Noteworthy, the effect of capacity constraints becomes sizable as the subsidy grows.

Figure 11: Unconditional subsidies and private prices



Notes: median impact on private prices in each counterfactual simulation for unconditional subsidies. In the x-axis, we show several possible values for the monthly scholarship; in the y-axis, we show the increase in average private prices in each scenario.

Figure 12: Unconditional subsidies, private prices, and capacity constraints

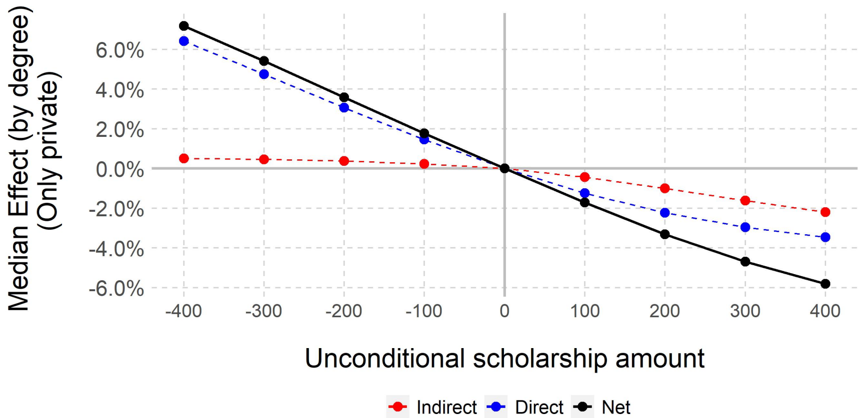


Notes: change in average prices in each counterfactual simulation for unconditional subsidies. In the x-axis, we show several possible values for the proportion of capacity-constrained public degrees; in the y-axis, we show the increase in average prices paid by the students in each scenario; the size of the bullets represents the subsidy amount.

Using the definition of average price sensitivity level in equation (39) and the additive version of the decomposition outlined in equation (38), we decompose the average price sensitivity of students attending private degrees under unconditional subsidies. Results are shown in Figure (14). The direct effect of increasing unconditional subsidies is an increase in the price sensitivity of students attending private schools, in line with the fact that the same effect makes prices go down and reinforcing the role of the composition effect to understanding the price responses our model generates. On the other hand, the indirect effect of increasing unconditional subsidies also increase the price sensitivity of students attending private degrees because of the impact of subsidies on admission requirements, forcing low-score students to switch from public to private schools. Results suggest that capacity constraints play a key role to estimating the composition effect under unconditional subsidies.

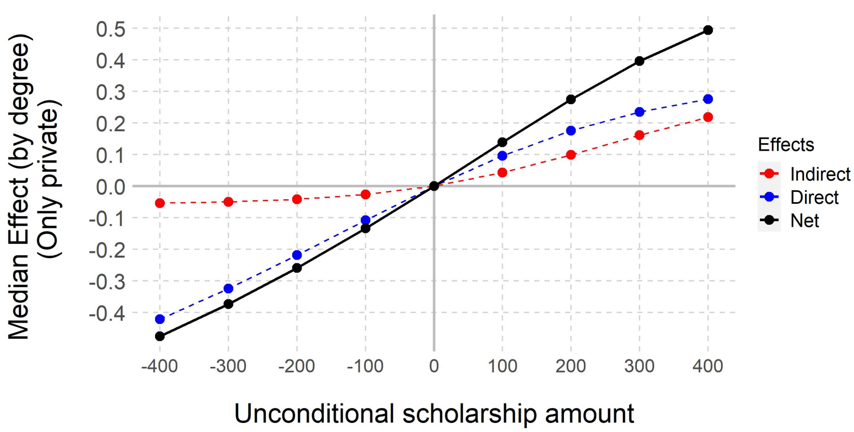
All in all, results show that a flat and unconditional subsidy does not increase private prices. Such a

Figure 13: Pass-through decomposition: unconditional subsidies



Notes: median impact on private prices in each counterfactual simulation for unconditional subsidies. In the x-axis, we show several possible values for the monthly scholarship; in the y-axis, we show the median price increase in each scenario. The figure reports the direct effect (when public schools do not face capacity constraints) and the indirect effect (from changes in the level of capacity constraints in public schools). The direct and indirect effects are defined in equation (38).

Figure 14: $PS_{d,w}$ decomposition for each w: unconditional subsidies



Notes: median impact on the average price sensitivity of students attending private degrees in each counterfactual simulation for unconditional subsidies. In the x-axis, we show several possible values for the monthly scholarship; in the y-axis, we show the median price sensitivity changes in each scenario. The figure reports the direct effect (when public schools do not face capacity constraints) and the indirect effect (from changes in the level of capacity constraints in public schools). The direct and indirect effects are defined in equation (38), but are additive (instead of multiplicative).

possibility reduces the concerns about the pass-through of private subsidies to private prices. Since these simulations are not budget-neutral, analyzing their welfare impacts in equilibrium would require knowledge about the marginal costs and benefits of increasing or reducing taxation. So, we redesign previous experiments to make them budget-neutral and look for welfare-improving policies that increase unconditional subsidies: the next set of counterfactuals charge tuition in public schools and use the tuition revenue to pay a scholarship to low-income students.

7.4 Counterfactual Budget-Neutral Policies

The third set of counterfactuals are the budget-neutral policies that increase unconditional subsidies charging tuition in public schools.

To simulate budget-neutral policies, $t_w > 0$ varies exogenously for public degrees. In each simulation, all public degrees charge the same tuition.

The budget-neutrality condition adjusts T_w to guarantee that the revenue from tuition in public schools matches the aggregate scholarship expenditures. The budget-neutral T_w, $\tilde{T}_w(t_w)$, is shown in equation (40)

$$\tilde{T}_w(t_w) = \frac{\sum_{d:h_d=Public} \tilde{S}_{dw}(\tilde{p}_{dw}, \tilde{p}_{-dw}, \tilde{\lambda}_{dw}, \tilde{\lambda}_{-dw}; \xi, \theta, M) t_w}{\sum_d \sum_{X_i} M(X_i) \tilde{s}_{dw}(X_i, \tilde{p}_{dw}, \tilde{p}_{-dw}, \tilde{\lambda}_{dw}, \tilde{\lambda}_{-dw}; \xi, \theta) \mathbb{1}_{(x_i^{AA}=1)}} \tag{40}$$

Results from budget-neutral simulations are shown in Table (14). The only simulations that raise enrollment are the ones in which tuition is positive and small, up to BRL 200. When tuition is low, aggregate enrollment rises and the enrollment gap between high-SES and low-SES students falls. The small tuition scenario is the only one in which consumer surplus rises, which is a surprising result given that average market prices rise between 9.76% and 17.75%: most of the effect comes from the reduction in overcapacity in public schools (enrollment in public schools falls significantly in all scenarios), increasing the choice set of low-score students. In all simulations, private enrollment and profits rise. However, private prices decrease because the price elasticity of students attending private degrees rises.

Rising tuition reduces capacity constraints in public degrees, making the positive effects of reducing admission scores in public schools outweigh the negative effects of higher prices in the market.

Table 14: Budget neutral simulations

Public tuition	-400.00	-300.00	-200.00	-100.00	100.00	200.00	300.00	400.00
Monthly subsidy (BRL)	-176.86	-119.82	-71.97	-31.88	22.73	36.35	42.85	44.32
Enrollment	−21.68%	−14.72%	−8.45%	−3.41%	1.10%	0.54%	−0.95%	−2.95%
SES Gap	2.98pp	1.94pp	1.23pp	0.60pp	-0.75pp	-1.39pp	-1.78pp	-1.97pp
Consumer Surplus	-26.40	-18.06	-10.60	-4.41	1.95	1.77	0.19	-2.22
Average Market Prices	-50.24%	-35.29%	-22.39%	-10.65%	9.76%	17.75%	23.84%	28.33%
Public Enrollment	25.71%	23.65%	19.61%	11.81%	-16.58%	-33.67%	-48.64%	-60.96%
Private Enrollment	-39.70%	-29.31%	-19.12%	-9.19%	7.82%	13.54%	17.18%	19.11%
Profit	-35.10%	-25.62%	-16.52%	-7.86%	6.56%	11.34%	14.42%	16.13%
Average Private Prices	6.84%	4.63%	2.81%	1.27%	-1.00%	-1.65%	-1.98%	-2.07%

Notes: Subsidy and tuition in monthly BRL. Consumer surplus in BRL per student in the market. All results compared to baseline levels.

Therefore, increasing subsidies to private schools can improve the market equilibrium even when public schools are of high quality. Results show that a budget-neutral reform that charges a small tuition fee in public schools and provides an unconditional scholarship can improve market efficiency by increasing enrollment, consumer surplus, and profits. The same reform can increase equity, reducing the gap between low-SES and high-SES students. So, the reform has positive effects on the higher education market, improving the welfare of its participants, and its effects have the potential to expand the positive externalities of higher education.

8 Conclusion

We estimated and simulated a model to compare the effects of conditional and unconditional subsidies on higher education. Our model incorporates degree and student heterogeneity, highlighting the role of capacity constraints and admission scores in the market equilibrium.

The effect of any reform on the price elasticity of the students attending a private degree is an empirical question: even if the price elasticity of every consumer in the market falls, the average price elasticity of consumers attending some degrees may rise depending on the switching behavior of students with higher and lower price elasticities (the composition effect). Because of the composition effect, private and public subsidies can reduce or increase private prices, depending on the context.

We show that the price elasticity of students attending private degrees is lower when public schools are heavily subsidized, increasing private prices. Moreover, if public schools are capacity-constrained and selective, heavily subsidizing public degrees makes them artificially selective, increasing admission requirements and reducing the admissions of low-score students. Since low-score students typically have higher price sensitivity, capacity constraints affect the average price elasticity of students attending private degrees. Simulations showed that capacity constraints change the direction of the impact of public subsidies on private prices: without capacity constraints, heavily subsidizing public schools raises private prices; with capacity constraints, subsidies to public schools pass through admission scores and reduce private prices.

Unconditional subsidies reduce private prices because price-elastic students switch to private degrees, and admission scores in public schools rise. So, market characteristics reduce the concern that private schools would raise prices if subsidized, and it may be possible to design a reform that increases subsidies to private schools and improves welfare.

Then, we simulate budget-neutral policies that charge tuition in public schools and use the tuition revenue to give an unconditional scholarship. Those simulations found that charging small tuition in public schools (below BRL 200 monthly) and distributing budget-neutral unconditional stipends raises enrollment by up to 1.1%, reducing the enrollment gap between high-SES and low-SES students. The same reform also raises consumer surplus, even though average prices rise more between 9.76% and 17.75%. Profits rise, private prices fall, and private enrollment rises.

References

Agarwal, N. and Somaini, P. (2018). Demand analysis using strategic reports: an application to a school choice mechanism. *Econometrica*, 86.

Atal, J. P., Cuesta, J. I., González, F., and Otero, C. (2024). The Economics of the Public Option: Evidence from Local Pharmaceutical Markets. *American Economic Review*, 114.

Aygün, O. and Bó, I. (2021). College Admission with Multidimensional Privileges: The Brazilian Affirmative Action Case. *American Economic Journal: Microeconomics*, 13.

Balinkski, M. and Sönmez, T. (1999). A Tale of Two Mechanisms: Student Placement. *Journal of Economic Theory*, 84.

Banerjee, A., Hanna, R., Kyle, J., Olken, B., and Sumarto, S. (2019). Private Outsourcing and Competition: Subsidized Food Distribution in Indonesia. *Journal of Political Economy*, 127.

Berry, S., Levinsohn, J., and Pakes, A. (2004). Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market. *Journal of Political Economy*, 112.

Besley, T. and Coate, S. (1991). Public Provision of Private Goods and the Redistribution of Income. *The American Economic Review*, 81.

Brazilian Ministry of Education (2018). Regulation 1,095.

Busso, M. and Galiani, S. (2019). The Causal Effect of Competition on Prices and Quality: Evidence from a Field Experiment. *American Economic Journal: Applied Economics*, 11.

Bó, I. and Hakimov, R. (2019). Iterative versus standard deferred acceptance: experimental evidence. *The Economic Journal*, 130.

Cellini, S. (2009). Crowded Colleges and College Crowd-Out: The Impact of Public Subsidies on the Two-Year College Market. *American Economic Journal: Economic Policy*, 1.

Cellini, S. and Goldin, C. (2014). Does Federal Student Aid Raise Tuition? New Evidence on For-Profit Colleges. *American Economic Journal: Economic Policy*, 6.

Chetty, R., Friedman, J., Saez, E., Turner, N., and Yagan, D. (2020). Income Segregation and Intergenerational Mobility Across Colleges in the United States. *The Quarterly Journal of Economics*, 135.

Cohodes, S. and Goodman, J. (2014). Merit Aid, College Quality, and College Completion: Massachussetts' Adams Scholarship as an In-Kind Subsidy. *American Economic Journal: Applied Economics*, 6.

Curto, V., Einav, L., Finkelstein, A., Levin, J., and Bhattacharya, J. (2019). Health Care Spending and Utilization in Public and Private Medicine. *American Economic Journal: Applied Economics*, 11.

Dinerstein, M. and Smith, T. (2021). Quantifying the Supply Response of Private Schools to Public Policies. *American Economic Review*, 111(10):3376–3417.

Dobbin, C., Barahona, N., and Otero, S. (2022). The Equilibrium Effects of Subsidized Student Loans.

Dynarski, S., Page, L., and Scott-Clayton, J. (2022). College Costs, Financial Aid, and Student Decisions.

Epple, D., Romano, R., Sarpça, S., and Sieg, H. (2017). A general equilibrium analysis of state and private colleges and access to higher education in the us. *Journal of Public Economics*, 155.

Estevan, F., Gall, T., and Morin, L.-P. (2018). Redistribution without distortion: evidence from an affirmative action programme at a large Brazilian university. *The Economic Journal*, 129.

Eurostat (2016). European Union Statistics on Income and Living Conditions: Module "access to services".

Eurydice (2019). Key Data on Early Childhood Education and Care in Europe: 2019 edition.

Fack, G., Grenet, J., and He, Y. (2019). Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions. *American Economic Review*, 109.

Fillmore, I. (2023). Price Discrimination and Public Policy in the US College Market. *Review of Economic Studies*, 90:1228–1264.

Fonseca, J. and Matray, A. (2022). Financial Inclusion, Economic Development, and Inequality: Evidence from Brazil.

Gale, D. and Shapley, L. (1962). College Admissions and the Stability of the Marriage. *The American Mathematical Monthly*, 69(1):9–15.

Handbury, J. and Moshary, S. (2021). School Food Policy Affects Everyone: Retail Responses to the National School Lunch Program.

INEP (2015). Exame Nacional do Ensino Médio.

INEP (2016). Censo da Educação Superior.

Jiménez-Hernández, D. and Seira, E. (2022). Should the Government Sell You Goods? Evidence from the Milk Market in Mexico.

Kapor, A., Karnani, M., and Neilson, C. (2022). Aftermarket Frictions and the Cost of Off-platform Options in Centralized Assignment Mechanisms.

Kargar, M. and Mann, W. (2022). The Incidence of Student Loan Subsidies: Evidence from the PLUS Program. *The Review of Financial Studies*, 36.

Levy, D. (2018). Global private higher education: an empirical profile of its size and geographical shape. *Higher Education*, 76.

Long, B. (2004). Does the Format of a Financial Aid Program Matter? The Effect of State In-Kind Tuition Subsidies. *Review of Economics and Statistics*, 86.

Lucca, D., Nadauld, T., and Shen, K. (2018). Credit Supply and the Rise in College Tuition: Evidence from the Expansion in Federal Student Aid Programs. *The Review of Financial Studies*, 32.

OECD (2020). Waiting Times for Health Services: Next in Line.

Peltzman, S. (1973). The Effect of Government Subsidies-in-Kind on Private Expenditures: The Case of Higher Education. *Journal of Political Economy*, 81.

Sahai, H. (2023). The Welfare Implications of School Voucher Desing: Evidence from India.

Singell, L. and Stone, J. (2007). For whom the Pell tolls: The response of university tuition to federal grants-in-aid. *Economics of Education Review*, 26.

Small, K. and Rosen, H. (1981). Applied Welfare Economics with Discrete Choice Models. *Econometrica*, 49.

Turner, L. (2017). The Economic Incidence of Federal Student Grant Aid.

Turner, N. (2012). Who benefits from student aid? The economic incidence of tax-based federal student aid. *Economics of Education Review*, 31.

Appendix A SES and Affirmative Action Eligibility Data

Per capita income is the quotient between total family income and the number of family members. The number of family members is observable (the sixth question in ENEM's socioeconomic questionnaire, varying from 1 to 20). However, total family income (the fifth question in ENEM's socioeconomic questionnaire) is measured using 17 possible income ranges. Since the range does not allow for a direct calculation of *per capita* income, we assume family income is in the midpoint of each range. The last range has no maximum, so we extrapolate the income growth from the third to the second to last ranges and assume the last range is equivalent to 21.5 times the minimum wage $(17.5 + (17.5 - 13.5))$, as shown in Table (15).

Table 15: Total family income

Range (ENEM's questionnaire)	Imputation
No income	0.00
Up to 1.0 minimum wage	0.50
From 1.0 to 1.5 minimum wage	1.25
From 1.5 to 2.0 minimum wages	1.75
From 2.0 to 2.5 minimum wages	2.25
From 2.5 to 3.0 minimum wages	2.75
From 3.0 to 4.0 minimum wages	3.50
From 4.0 to 5.0 minimum wages	4.50
From 5.0 to 6.0 minimum wages	5.50
From 6.0 to 7.0 minimum wages	6.50
From 7.0 to 8.0 minimum wages	7.50
From 8.0 to 9.0 minimum wages	8.50
From 9.0 to 10.0 minimum wages	9.50
From 10.0 to 12.0 minimum wages	11.00
From 12.0 minimum wages to 15.0 minimum wages	13.50
From 15.0 minimum wages to 20.0 minimum wages	17.50
More than 20.0 minimum wages	21.50

Affirmative action eligibility comes from students' answers to the 47th question in ENEM's socioeconomic questionnaire. Only students who answered that they were continuously enrolled in public high schools were eligible for affirmative action policies.

A.1 SES and Test Score Distributions

Table 16: Minimum, maximum, and mean per capita income in each SES quintile

SES quintile	Min	Max	Mean
1	0.000	0.250	0.123
2	0.250	0.417	0.319
3	0.417	0.625	0.524
4	0.625	1.037	0.799
5	1.037	21.500	2.054

Table 17: Minimum, maximum, and mean test scores in each test score percentile

Test score percentile	Min	Max	Mean
1	258.520	371.940	346.756
2	371.940	390.620	382.508
3	390.620	401.200	396.270
4	401.200	408.840	405.178
5	408.840	414.960	411.999
6	414.960	420.160	417.628
7	420.160	424.740	422.497
8	424.740	428.860	426.836
9	428.860	432.600	430.751
10	432.600	436.100	434.355
11	436.100	439.360	437.747
12	439.360	442.400	440.897
13	442.400	445.280	443.867
14	445.280	448.040	446.676
15	448.040	450.680	449.379
16	450.680	453.180	451.945
17	453.180	455.620	454.413
18	455.620	457.960	456.790
19	457.960	460.260	459.112
20	460.260	462.500	461.379
21	462.500	464.640	463.565
22	464.640	466.740	465.683
23	466.740	468.800	467.765
24	468.800	470.820	469.794
25	470.820	472.720	471.752
26	472.720	474.640	473.671
27	474.640	476.480	475.566
28	476.480	478.260	477.392
29	478.260	480.000	479.179
30	480.000	481.780	480.888
31	481.780	483.340	482.556
32	483.340	485.000	484.228
33	485.000	486.720	485.864
34	486.720	488.120	487.446
35	488.120	489.780	488.956
36	489.780	491.120	490.468
37	491.120	492.800	491.969
38	492.800	494.000	493.420
39	494.000	495.560	494.823
40	495.560	497.000	496.290
41	497.000	498.300	497.642

42	498.300	499.820	499.050
43	499.820	501.000	500.458
44	501.000	502.440	501.779
45	502.440	503.920	503.137
46	503.920	505.000	504.465
47	505.000	506.460	505.768
48	506.460	508.000	507.208
49	508.000	509.060	508.531
50	509.060	510.580	509.869
51	510.580	512.000	511.254
52	512.000	513.120	512.536
53	513.120	514.700	513.932
54	514.700	516.000	515.370
55	516.000	517.340	516.671
56	517.340	518.920	518.097
57	518.920	520.000	519.423
58	520.000	521.520	520.799
59	521.520	523.000	522.291
60	523.000	524.500	523.677
61	524.500	526.000	525.299
62	526.000	527.580	526.718
63	527.580	529.140	528.408
64	529.140	531.000	530.094
65	531.000	532.520	531.694
66	532.520	534.160	533.378
67	534.160	536.000	535.141
68	536.000	537.940	536.883
69	537.940	539.800	538.763
70	539.800	541.600	540.613
71	541.600	543.480	542.482
72	543.480	545.520	544.486
73	545.520	547.760	546.591
74	547.760	549.980	548.789
75	549.980	552.220	551.076
76	552.220	554.700	553.485
77	554.700	557.120	555.911
78	557.120	559.900	558.493
79	559.900	562.640	561.177
80	562.640	565.520	564.043
81	565.520	568.620	567.040
82	568.620	571.820	570.166
83	571.820	575.180	573.496
84	575.180	579.000	577.072

85	579.000	582.860	580.816
86	582.860	587.000	584.913
87	587.000	591.320	589.157
88	591.320	596.000	593.631
89	596.000	601.160	598.548
90	601.160	606.980	604.010
91	606.980	613.180	610.002
92	613.180	619.960	616.499
93	619.960	627.440	623.612
94	627.440	636.020	631.624
95	636.020	645.920	640.874
96	645.920	657.320	651.511
97	657.320	671.400	664.130
98	671.400	689.160	679.852
99	689.160	715.320	701.113
100	715.320	866.480	742.437

Appendix B Instruments

Two major threats to our identification strategy exist. The first is that our instrument, local wages, has little variation and is not strongly correlated with prices. The second is that our instrument is correlated with demand shocks. For example, if some regions have higher total factor productivity, they may simultaneously have higher wages and higher education demand. In this case, $E[Z^T \Delta\xi] > 0$.

We estimate a demand model without consumer heterogeneity to discuss the relevance of these two threats by comparing the results of OLS and 2SLS estimates and reporting first-stage results. The estimated model assumes that $\alpha(X_i, \theta) = -\alpha$, resulting in the reduced form shown in Equation (41).

$$log(s_d) - log(s_0) = \alpha p_d + h_d + r_d + \Delta\xi_d \tag{41}$$

OLS and 2SLS estimates are shown in Table (18). In our preferred specification (with fixed effects), OLS estimates are statistically insignificant and very close to zero, while 2SLS estimates are statistically significant and negative. These results are robust to the model without fixed effects. If $E[Z^T \Delta\xi] > 0$, the 2SLS estimates would be higher than the OLS estimates, meaning that our results seem to reduce our concerns with the identification threats we presented.

Table 18: Homogeneous demand

	OLS	OLS (FE)	2SLS	2SLS (FE)
Price	0.108^{***}	-0.024	-2.070^{***}	-2.385^{***}
	(0.018)	(0.028)	(0.232)	(0.272)
Major FE	No	Yes	No	Yes
Ownership FE	No	Yes	No	Yes
Num. obs.	8960	8960	8960	8960

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors in parantheses.

Moreover, we use the simplified model to evaluate the correlation between prices and our instruments using the first stage of the 2SLS procedure. Results are shown in Table (19). Because public school prices are always zero (exogenously determined), we also estimate a first-stage regression including only private degrees in the sample. The F-statistic for the regression with all schools is 21.6 and 61.0 when only private schools are included, suggesting that regional wages are correlated with degree prices.

The signs of some estimates are negative because of the correlation between our explanatory variables. We estimate alternative versions of our first-stage results, including only one instrument at a time in a sample that has only private degrees. Results are reported in Table (20).[34] All estimates have positive signs and are statistically significant. Furthermore, in Table (21) we report the correlation between our instruments: many of them are higher than 0.50.

[34]Group 1: Mathematics, Statistics, and Computer Science. Group 2: Architecture, Urbanism, Engineering, Geology, and Geophysics. Group 3: Biology and Health. Group 4: Pedagogy. Group 5: Language and Literature. Group 6: Humanities. Group 7: Economics, Business, and Accounting.

Table 19: First-stage results (homogeneous demand)

	First-stage	Only private degrees
Mathematics, Statistics, and Computer Science	-0.011	-0.289^{***}
	(0.082)	(0.081)
Architecture, Urbanism, Engineering, Geology, and Geophysics	0.371^{***}	0.709^{***}
	(0.116)	(0.112)
Biology and Health	-0.159	-0.409^{***}
	(0.114)	(0.109)
Pedagogy	0.677^{***}	1.072^{***}
	(0.077)	(0.074)
Language and Literature	0.002	-0.049
	(0.095)	(0.091)
Humanities	0.049	0.319^{***}
	(0.101)	(0.099)
Economics, Business, and Accounting	0.144^{***}	0.204^{***}
	(0.053)	(0.049)
All employees	-0.227	-0.133
	(0.158)	(0.157)
Major FE	Yes	Yes
Ownership FE	Yes	Yes
Sample	All degrees	Only private
R^2 (full model)	0.6272	0.7695
R^2 (projection model)	0.0191	0.0768
F-statistic (full model)	165.77	272.28
F-statistic (projection model)	21.63	61.04
Num. obs.	8960	5944

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors in parantheses. Cost shifters are the wages of faculty, by department, and the wages of all college employees. Statistics reported for the full (projection) model do (not) include major dummies.

Table 20: Robustness for first stage (homogeneous demand)

	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
All	1.445^{***}							
	(0.125)							
Group 1		0.407^{***}						
		(0.054)						
Group 2			0.606^{***}					
			(0.056)					
Group 3				0.627^{***}				
				(0.059)				
Group 4					1.025^{***}			
					(0.055)			
Group 5						0.396^{***}		
						(0.051)		
Group 6							0.745^{***}	
							(0.056)	
Group 7								0.247^{***}
								(0.038)
Major FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2 (full)	0.6218	0.6217	0.6227	0.6221	0.6249	0.6216	0.6225	0.6210
R^2 (proj)	0.0050	0.0049	0.0073	0.0058	0.0131	0.0046	0.0069	0.0029
F (full)	175.81	175.78	176.47	176.04	178.13	175.69	176.37	175.22
F (proj)	44.39	43.55	65.23	51.69	117.76	40.69	62.05	25.82
Num. obs.	5944	5944	5944	5944	5944	5944	5944	5944

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Standard errors in parantheses. Regression explaining private prices from the cost shifter and major dummies. Cost shifters are the wages of faculty, by department, and the wages of all college employees. Statistics reported for the full (projection) model do (not) include major dummies.

Table 21: Correlation matrix: cost shifters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	All
Group 1	1.00	0.72	0.58	0.20	0.63	0.54	0.39	0.33
Group 2	0.72	1.00	0.71	0.15	0.81	0.59	0.52	0.30
Group 3	0.58	0.71	1.00	0.32	0.63	0.80	0.44	0.36
Group 4	0.20	0.15	0.32	1.00	0.07	0.39	-0.05	0.53
Group 5	0.63	0.81	0.63	0.07	1.00	0.54	0.61	0.31
Group 6	0.54	0.59	0.80	0.39	0.54	1.00	0.45	0.39
Group 7	0.39	0.52	0.44	-0.05	0.61	0.45	1.00	0.31
All	0.33	0.30	0.36	0.53	0.31	0.39	0.31	1.00

Appendix C Solution to the Problem of Public Schools

The objective of this appendix is to prove that (9) is a solution to the problem of public schools stated in (24).

To do that, we will first prove that $\mu(X_i;\lambda)$ and $N_d^j(\lambda_d^j)$ are non-increasing in λ_d^j. These are very useful results for the solution of our model.

Then, we discuss five hypotheses that we need to guarantee the existence and unicity of a solution for the problem of public schools. Then, we prove a group of propositions that allow us to fully characterize the solution to the problem of public schools.

Lemma 1. *$\mu(X_i;\lambda)$ is non-increasing in λ*

Proof. From equation (13):
$\mu(X_i;\lambda) = \mathbb{1}(x_i^{SCORE} \geq \lambda)$
For $\epsilon > 0$:
$\mu(X_i;\lambda) = \mathbb{1}(x_i^{SCORE} \geq \lambda + \epsilon) + \mathbb{1}(\lambda + \epsilon > x_i^{SCORE} \geq \lambda)$
Moreover, from equation (13):
$\mu(X_i;\lambda + \epsilon) = \mathbb{1}(x_i^{SCORE} \geq \lambda + \epsilon)$
So:
$\mu(X_i;\lambda) = \mu(X_i;\lambda + \epsilon) + \mathbb{1}(\lambda + \epsilon > x_i^{SCORE} \geq \lambda)$
$\mu(X_i;\lambda) - \mu(X_i;\lambda + \epsilon) = \mathbb{1}(\lambda + \epsilon > x_i^{SCORE} \geq \lambda)$
Because $\mathbb{1}(\lambda + \epsilon > x_i^{SCORE} \geq \lambda)$ is an indicator function, $\mathbb{1}(\lambda + \epsilon > x_i^{SCORE} \geq \lambda) \geq 0$. Therefore:
$\mu(X_i;\lambda) - \mu(X_i;\lambda + \epsilon) = \mathbb{1}(\lambda_d^0 + \epsilon > x_i^{SCORE} \geq \lambda_d^0) \geq 0$
$\mu(X_i;\lambda) - \mu(X_i;\lambda + \epsilon) \geq 0$
$\mu(X_i;\lambda) \geq \mu(X_i;\lambda + \epsilon)$ □

Lemma 2. *$N_d^j(\lambda_d^j)$ is non-increasing in λ_d^j.*

Proof. From (17), (18), and (19):
$N_d^j(\lambda_d^j) = \sum_{x_i^{AA}=j} \mu(X_i;\lambda_d^j)\mathbb{1}(u_{id} > u_{iw}\,\forall w \neq d \,|\, \mu(X_i;\lambda_w^j) = 1)$.
For $\epsilon > 0$:
$N_d^j(\lambda_d^j) - N_d^j(\lambda_d^j + \epsilon) = \sum_{x_i^{AA}=j} (\mu(X_i;\lambda_d^j) - \mu(X_i;\lambda_d^j + \epsilon))\mathbb{1}(u_{id} > u_{iw}\,\forall w \neq d \,|\, \mu(X_i;\lambda_w^j) = 1)$.
From lemma (1): $\mu(X_i;\lambda_d^j) \geq \mu(X_i;\lambda_d^j + \epsilon)$.
From the definition of indicator function: $\mathbb{1}(u_{id} > u_{iw}\,\forall w \neq d \,|\, \mu(X_i;\lambda_w^j) = 1) \geq 0$.
Therefore:
$\sum_{x_i^{AA}=j} (\mu(X_i;\lambda_d^j) - \mu(X_i;\lambda_d^j + \epsilon))\mathbb{1}(u_{id} > u_{iw}\,\forall w \neq d \,|\, \mu(X_i;\lambda_w^j) = 1) \geq 0$.
$N_d^j(\lambda_d^j) - N_d^j(\lambda_d^j + \epsilon) \geq 0$.
$N_d^j(\lambda_d^j) \geq N_d^j(\lambda_d^j + \epsilon)$.

□

We need six hypotheses to proceed. The first one, stated in (2), is necessary to guarantee that there is always a cutoff score that makes demand equal to or smaller than the number of seats. In other words, it assures that cutoff scores are effective in producing a matching that is feasible.

Assumption 2. *(Effectiveness)*
$N_d^0(1) \approx 0$
$N_d^1(1) \approx 0$

The second one, stated in (3), is necessary to guarantee that the solution to the problem of public schools is unique. Without assuming injectivity, we proved that $N_d^k(\lambda_d^k)$ is non-increasing in λ_d^k in lemma (2). Therefore, the assumption implies that the number of students choosing degree d in each possible test score above the cutoff for degree d is always positive.

Assumption 3. *(Injectivity)*
$N_d^j(\lambda_0) = N_d^j(\lambda_1)$ *iff* $\lambda_0 = \lambda_1$

Noteworthy, if degree d has a cutoff score equal to λ_d and $t \geq \lambda$, the probability that a student with test score t chooses d is always positive. Therefore, assumption (3) holds when markets are large enough.

The next four hypotheses are called quasi-continuity hypotheses. Their purpose is to avoid numerical problems that can arise from the fact that $N_d^j(\lambda_d^j)$ is not continuous on λ_d^j.

After all, the solution to the problem of public schools involves making $N_d^j(\lambda_d^j)$, or a function of $N_d^j(\lambda_d^j)$, close to a target, and we want to avoid the possibility that discontinuities in $N_d^j(\lambda_d^j)$ prevent us from getting a precise solution. The quasi-continuity hypotheses encompass assumptions (4), (5), (6), and (7).

Assumption 4. *(Quasi-continuity)*
$\forall \lambda_d^1 : N_d(0, \lambda_d^1) > K_d > N_d(1, \lambda_d^1),\ \exists \lambda_d^0 : N_d(\lambda_d^0, \lambda_d^1) \approx K_d$

Assumption 5. *(Quasi-continuity)*
$\forall \lambda_d^0 : N_d(\lambda_d^0, 0) > K_d > N_d(\lambda_d^0, 1),\ \exists \lambda_d^1 : N_d(\lambda_d^0, \lambda_d^1) \approx K_d$

Assumption 6. *(Quasi-continuity)*
If: $\psi_d(0) > \Psi_d$*:*
$\exists \lambda_d^R$*:* $\Psi_d \approx \psi_d(\lambda_d^R)$

Assumption 7. *(Quasi-continuity)*
If $N_d(0, 0) \geq K_d$ *and:*
(1) $\Psi_d = 0$ *or*
(2) $\psi_d(0) > \Psi_d$ *and* $N_d(\lambda_d^R, \lambda_d^R) < K_d$
Then:
$\exists \lambda$*:* $N_d(\lambda, \lambda) \approx K_d$

Noteworthy, if $N_d^0(\lambda_d^0)$ and $N_d^1(\lambda_d^1)$ were continuous, assumption (4) would follow from the intermediate value theorem and assumptions (6) and (7) would follow from assumption (2) and the intermediate value theorem.

In other words, to solve the problem of public schools we do not need any additional strong hypotheses.

Using assumption (3), we write a stronger version of lemma (2) in lemma (3).

Lemma 3. $N_d^j(\lambda)$ *is strictly decreasing in* λ.

Proof. From lemma (2), $\forall \epsilon > 0$, $N_d^j(\lambda) \geq N_d^j(\lambda + \epsilon)$
From assumption (3), $N_d^j(\lambda) = N_d^j(\lambda + \epsilon)$ iff $\epsilon = 0$.
Because $\epsilon > 0$:
$N_d^j(\lambda) > N_d^j(\lambda + \epsilon)$

□

Now, we define $n_d^j(\lambda) = N_d^j(\lambda) - N_d^j(\lambda + \Delta)$, where Δ is the smallest change in λ and $n_d^j(\lambda)$ is the enrollment in degree d of students from affirmative action group j whose test scores are equal to λ.

Having discussed the basic properties of the enrollment function, we discuss the solution to the problem of public schools. First, we split the objective function in (20) into two components, O_0 and O_1, as shown in equation (42).

$$
\begin{aligned}
O(\lambda_d^0, \lambda_d^1) &= O_0(\lambda_d^0) + O_1(\lambda_d^1) \\
O_j(\lambda_d^j) &= \sum_{i \in A_d^j} x_i^{SCORE} = \sum_{i=\frac{\lambda_d^j}{\Delta}}^{\frac{1}{\Delta}} i\Delta n_d^j(i\Delta)
\end{aligned}
\tag{42}
$$

The strict monotonicity of $N_d^j(\lambda)$ implies that $n_d^j(\lambda)$ is always positive, as stated in lemma (4).

Lemma 4. $n_d^j(\lambda) > 0$, $\forall \lambda$

Proof. From lemma (3), $N_d^j(\lambda + \epsilon) < N_d^j(\lambda)$, $\forall \epsilon > 0$.
From the definition of $n_d^j(\lambda)$: $n_d^j(\lambda) = N_d^j(\lambda) - N_d^j(\lambda + \Delta)$.
Because $\Delta > 0$: $N_d^j(\lambda) > N_d^j(\lambda + \Delta)$ and $N_d^j(\lambda) - N_d^j(\lambda + \Delta) > 0$.
Therefore: $n_d^j(\lambda) > 0$.

□

In lemma (5), we prove that $O_j(\lambda)$ is strictly decreasing in λ. The lemma will be very important to show that $O(0, 0)$ is an unconstrained maximum of the objective function.

Lemma 5. *$O_j(\lambda)$ is strictly decreasing in λ*

Proof. Using 42:
$O_j(\lambda) = \sum_{i=\frac{\lambda}{\Delta}}^{\frac{1}{\Delta}} i\Delta n_d^j(i\Delta)$
For $\epsilon > 0$:
$O_j(\lambda) - O_j(\lambda + \epsilon) = \sum_{i=\frac{\lambda}{\Delta}}^{\frac{1}{\Delta}} i\Delta n_d^j(i\Delta) - \sum_{i=\frac{\lambda+\epsilon}{\Delta}}^{\frac{1}{\Delta}} i\Delta n_d^j(i\Delta)$
$O_j(\lambda) - O_j(\lambda + \epsilon) = \sum_{i=\frac{\lambda}{\Delta}}^{\frac{\lambda+\epsilon}{\Delta}} i\Delta n_d^j(i\Delta)$
From lemma (4), $i\Delta n_d^j(i\Delta) > 0$.
Therefore: $O_j(\lambda) > O_j(\lambda + \epsilon)$

□

In lemma (6), we show that if a cutoff for reserved seats satisfies the affirmative action restriction, all lower cutoffs also satisfy it.

Lemma 6. *If λ satisfies the affirmative action constraint, $\forall \epsilon > 0$ such that $1 \geq \lambda - \epsilon \geq 0$, $\lambda - \epsilon$, satisfies the affirmative action constraint.*

Proof. From equation (22): $\psi_d(\lambda) = \frac{N_d^1(\lambda)}{K_d}$
From lemma (3), for $\epsilon > 0$ such that $1 \geq \lambda - \epsilon \geq 0$:
$N_d^1(\lambda - \epsilon) > N_d^1(\lambda)$
Because $K_d > 0$:
$\psi_d(\lambda - \epsilon) > \psi_d(\lambda)$
From equation (23), the affirmative action restriction is given by:
$\min(\Psi_d, \psi_d(0)) \leq \psi_d(\lambda)$
Because $\psi_d(\lambda - \epsilon) > \psi_d(\lambda)$:
$\min(\Psi_d, \psi_d(0)) \leq \psi_d(\lambda) < \psi_d(\lambda - \epsilon)$.
And $\lambda - \epsilon$ satisfy the affirmative action restriction.

□

Lemma (7) shows that $\lambda_d^1 = 0$ always satisfies the affirmative action restriction.

Lemma 7. *The $\lambda_d^1 = 0$ always satisfies the affirmative action restriction.*

Proof. If $\lambda_d^1 = 0$, the affirmative action constraint shown in equation (23) is $\min(\Psi_d, \psi_d(0)) \leq \psi_d(0)$.
If $\Psi_d < \psi_d(0)$, the affirmative action restriction can be rewritten as $\Psi_d \leq \psi_d(0)$ and it is satisfied.
If $\Psi_d \geq \psi_d(0)$, the affirmative action restriction can be rewritten as $\psi_d(0) \leq \psi_d(0)$ and it is satisfied.
Therefore, $\lambda_d^1 = 0$ satisfies the affirmative action restriction.

□

Now, we prove the first part of our solution. Noteworthy, we define as feasible the set of cutoff scores $(\lambda_d^0, \lambda_d^1)$ that satisfies the capacity constraint given by (21) and the affirmative action restriction given by (23).

Proposition 2. *If $(\lambda_d^0, \lambda_d^1) = (0, 0)$ satisfies the capacity constraint, it is the only solution to the problem stated in 24.*

Proof. From lemma (7), $(0, 0)$ satisfies the affirmative action restriction. Because $(0, 0)$ satisfies the affirmative action and the capacity restriction, it is a feasible solution.
Let us assume the existence of $(\lambda_d^0, \lambda_d^1) \neq (0, 0)$ that is also feasible.
Then, we have to show that $(\lambda_d^0, \lambda_d^1)$ gives lower utility to public schools than $(0, 0)$.
In other words, we have to show that $O(\lambda_d^0, \lambda_d^1) < O(0, 0)$.
From equation (42):
(1) $O(\lambda_d^0, \lambda_d^1) = O_0(\lambda_d^0) + O_1(\lambda_d^1)$
(2) $O(0, 0) = O_0(0) + O_1(0)$
From lemma (5):
(1) If $\lambda_d^0 > 0$ and $\lambda_d^1 \geq 0$: $O_0(\lambda_d^0) < O_0(0)$ and $O_1(\lambda_d^1) \leq O_0(1)$.
(2) If $\lambda_d^1 > 0$ and $\lambda_d^0 \geq 0$: $O_1(\lambda_d^1) < O_1(0)$ and $O_0(\lambda_d^0) \leq O_0(0)$.
Therefore, if $\lambda_d^0 > 0$ or $\lambda_d^1 > 0$: $O(\lambda_d^0, \lambda_d^1) < O(0, 0)$

□

Now, we prove that if $(0,0)$ is not feasible, the solution to the problem stated in (24) requires that the capacity constraint is a binding restriction.

Proposition 3. *If $\lambda_d = (0,0)$ does not satisfy the capacity constraint, the solution to the problem stated in (24) requires that $N_d(\lambda_d) = K_d$.*

Proof. Let us assume that $(\lambda_d^0, \lambda_d^1)$ solve (24), satisfy the affirmative action constraint, and make $N_d(\lambda_d) < K_d$.
Case 1: $\lambda_d^1 > 0$, $\lambda_d^0 > 0$, and $N_d(0, \lambda_d^1) > K_d$.
Because $N_d(\lambda_d^0, \lambda_d^1) < K_d$, it follows that $N_d(\lambda_d^0, 1) < K_d$.
Then, according to assumption (5):
$\exists \lambda^*$: $N_d(\lambda_d^0, \lambda^*) = K_d$.
Because $N_d(\lambda_d^0, \lambda^*) = K_d > N_d(\lambda_d^0, \lambda_d^1)$, lemma (3) implies that $\lambda^* < \lambda_d^1$.
Therefore, if λ_d^1 satisfies the affirmative action restriction, lemma (6) implies that λ^* also satisfies the affirmative action restriction.
Consequently, $(\lambda_d^0, \lambda_d^1)$ and (λ_d^0, λ^*) are feasible.
So, to prove that $(\lambda_d^0, \lambda_d^1)$ is not a solution, we have to show that $O(\lambda_d^0, \lambda_d^1) < O(\lambda_d^0, \lambda^*)$, which follows from lemma (5) and the fact that $\lambda^* < \lambda_d^1$.
Case 2: $\lambda_d^1 > 0$, $\lambda_d^0 > 0$, and $N_d(0, \lambda_d^1) \leq K_d$
In such case $(0, \lambda_d^1)$ satisfies the capacity constraint and the affirmative action constraint, so it is feasible.
Now, we have to show that $O(\lambda_d^0, \lambda_d^1) < O(0, \lambda_d^1)$, which follows from lemma (5) and the fact that $0 < \lambda_d^0$.
Case 3: $\lambda_d^1 > 0$ and $\lambda_d^0 = 0$
According to lemma (3):
$N_d(\lambda_d^0, 0) > K_d > N_d(\lambda_d^0, 1)$
Then, according to assumption (5):
$\exists \lambda^*$: $N_d(\lambda_d^0, \lambda^*) = K_d$.
Because $N_d(\lambda_d^0, \lambda^*) = K_d > N_d(\lambda_d^0, \lambda_d^1)$, lemma (3) implies that $\lambda^* < \lambda_d^1$.
Therefore, if λ_d^1 satisfies the affirmative action restriction, lemma (6) implies that λ^* also satisfies the affirmative action restriction.
Consequently, $(\lambda_d^0, \lambda_d^1)$ and (λ_d^0, λ^*) are feasible.
To prove that $(\lambda_d^0, \lambda_d^1)$ is not a solution, we have to show that $O(\lambda_d^0, \lambda_d^1) < O(\lambda_d^0, \lambda^*)$, which follows from lemma (5) and the fact that $\lambda_d^* < \lambda_d^1$.
Case 4: $\lambda_d^1 = 0$
According to lemma (3):
$N_d(0, \lambda_d^1) > K_d > N_d(1, \lambda_d^1)$
According to assumption (5):
$\exists \lambda^*$: $N_d(\lambda^*, \lambda_d^1) = K_d$.
Because $N_d(\lambda^*, \lambda_d^1) = K_d > N_d(\lambda_d^0, \lambda_d^1)$, lemma (3) implies that $\lambda^* < \lambda_d^0$.
Consequently, $(\lambda_d^0, \lambda_d^1)$ and (λ^*, λ_d^1) are feasible.
To prove that $(\lambda_d^0, \lambda_d^1)$ is not a solution, we have to show that $O(\lambda_d^0, \lambda_d^1) < O(\lambda^*, \lambda_d^1)$, which follows from lemma (5) and the fact that $\lambda_d^* < \lambda_d^0$.

□

Now, we prove that if $(0,0)$ is not feasible and $\psi_d(0) < \Psi_d$, the solution to the problem stated in (24) is given by $N_d(\lambda_d^0, 0) = K_d$.

Proposition 4. *If* $(0,0)$ *does not satisfy the capacity constraint and* $\psi_d(0) < \Psi_d$*, the* λ_d^0 *that solves (24) is the only* λ *that satisfies* $N_d(\lambda, 0) = K_d$*.*

Proof. If $\psi_d(0)$, the affirmative action constraint in (23) can be rewritten as $\psi_d(0) \le \psi_d(\lambda_d^1)$
From lemma (3), $N_d^1(\lambda)$ and the definition of $\psi_d(\lambda)$ in equation (22), $\psi_d(\lambda)$ is strictly monotonic and $\psi_d(0) < \psi_d(\lambda)\forall\lambda > 0$.
Therefore, the only solution to the affirmative action restriction is $\lambda_d^1 = 0$.
Assumption (2) implies that $N_d^0(1) = 0$. We know that $N_d^1(0) < \Psi_d K_d$, so $N_d(1,0) < \Psi_d K_d < K_d$. we also know that $N_d(0,0) > K_d$.
Therefore, from assumption (4):
$\exists\lambda_d^0$: $N_d(\lambda_d^0, 0) = K_d$, which is the solution to (24).

□

Now, we discuss the solution to (**??**) when $(0,0)$ is not feasible and $\psi_d(0) \ge \Psi_d$. We divide the solution into two propositions.

Proposition 5. *If* $(0,0)$ *does not satisfy the capacity constraint,* $\psi_d(0) \ge \Psi_d$*, and* $N_d(\lambda_d^R, \lambda_d^R) < K_d$*,* $\psi_d(\lambda_d) = \Psi_d$*, the solution to (24) is* λ*:* $N_d(\lambda, \lambda) = K_d$*,* $\lambda \le \lambda_d^R$*.*

Proof. From assumption (6), λ_d^R exists.
From assumption (3), N_d^1 is injective, so ψ_d is injective and λ_d^R is unique.
From assumption (7), λ exists.
Because $\lambda \le \lambda_d^R$, lemma (6) assures that λ satisfies the affirmative action constraint.
Because $N_d(\lambda, \lambda) = K_d$, (λ, λ) is feasible.
Let us assume that there is another feasible score combination: $(\lambda_d^0, \lambda_d^1)$.
From proposition (3), $N_d(\lambda_d^0, \lambda_d^1) = K_d$.
We have to prove that $O(\lambda, \lambda) > O(\lambda_d^0, \lambda_d^1)$.
Then:

$$O_j(\lambda) = \sum_{i=\frac{\lambda}{\Delta}}^{\frac{1}{\Delta}} i\Delta n_d^j(i\Delta)$$

$$O_j(\lambda_d^j) = \sum_{i=\frac{\lambda_d^j}{\Delta}}^{\frac{\lambda}{\Delta}-1} i\Delta n_d^j(i\Delta) + \sum_{i=\frac{\lambda}{\Delta}}^{\frac{1}{\Delta}} i\Delta n_d^j(i\Delta) = \sum_{i=\frac{\lambda_d^j}{\Delta}}^{\frac{\lambda}{\Delta}-1} i\Delta n_d^j(i\Delta) + O_j(\lambda)$$

Therefore:

$$O_j(\lambda_d^j) - O_j(\lambda) = \sum_{i=\frac{\lambda_d^j}{\Delta}}^{\frac{\lambda}{\Delta}} i\Delta n_d^j(i\Delta)$$

Because $N_d(\lambda_d^0, \lambda_d^1) = K_d = N_d(\lambda, \lambda)$:

$$\sum_{i=\frac{\lambda_d^0}{\Delta}}^{\frac{1}{\Delta}} n_d^0(i\Delta) + \sum_{i=\frac{\lambda_d^1}{\Delta}}^{\frac{1}{\Delta}} n_d^1(i\Delta) = \sum_{i=\frac{\lambda}{\Delta}}^{\frac{1}{\Delta}} n_d^0(i\Delta) + \sum_{i=\frac{\lambda}{\Delta}}^{\frac{1}{\Delta}} n_d^1(i\Delta).$$

Therefore:

$$W = \sum_{i=\frac{\lambda_d^0}{\Delta}}^{\frac{\lambda}{\Delta}-1} n_d^0(i\Delta) = \sum_{i=\frac{\lambda}{\Delta}}^{\frac{\lambda_d^1}{\Delta}-1} n_d^1(i\Delta)$$

From lemma (3):
(1) $\lambda_d^1 > \lambda$ iff $\lambda_d^0 < \lambda$
(2) $\lambda_d^1 < \lambda$ iff $\lambda_d^0 > \lambda$

Without loss of generality, let us assume: $\lambda_d^0 > \lambda$ and $\lambda_d^1 < \lambda$

In this case: $\frac{\lambda_d^1}{\Delta} < \frac{\lambda}{\Delta} - 1$ and $\frac{\lambda}{\Delta} - 1 = \max\{\frac{\lambda_d^1}{\Delta}, \frac{\lambda_d^1}{\Delta} + 1, ..., \frac{\lambda}{\Delta} - 1\}$. So:

$O_1(\lambda_d^1) - O_1(\lambda) = \sum_{i=\frac{\lambda_d^1}{\Delta}}^{\frac{\lambda}{\Delta}-1} i\Delta n_d^j(i\Delta) \leq \left[\frac{\lambda}{\Delta} - 1\right] \sum_{i=\frac{\lambda_d^1}{\Delta}}^{\frac{\lambda}{\Delta}-1} \Delta n_d^j(i\Delta)$

Similarly:

$O_0(\lambda_d^0) - O_0(\lambda) \leq \left[\frac{\lambda_d^0}{\Delta}\right] \sum_{i=\frac{\lambda}{\Delta}-1}^{\frac{\lambda_d^0}{\Delta}} \Delta n_d^j(i\Delta)$

Therefore:

$O(\lambda_d^0, \lambda_d^1) - O(\lambda, \lambda) \leq \left[\frac{\lambda}{\Delta} - 1\right] \sum_{i=\frac{\lambda_d^1}{\Delta}}^{\frac{\lambda}{\Delta}-1} \Delta n_d^j(i\Delta) + \left[\frac{\lambda_d^0}{\Delta}\right] \sum_{i=\frac{\lambda}{\Delta}-1}^{\frac{\lambda_d^0}{\Delta}} \Delta n_d^j(i\Delta)$

$O(\lambda_d^0, \lambda_d^1) - O(\lambda, \lambda) \leq= \left[\frac{\lambda}{\Delta} - 1\right] W + \left[\frac{\lambda_d^0}{\Delta}\right] (-W)$

$O(\lambda_d^0, \lambda_d^1) - O(\lambda, \lambda) \leq W \left[\frac{\lambda}{\Delta} - \frac{\lambda_d^0}{\Delta} - 1\right]$

$O(\lambda, \lambda) - O(\lambda_d^0, \lambda_d^1) \geq W \left[1 + \frac{\lambda_d^0 - \lambda}{\Delta}\right] > 0$

Therefore: $O(\lambda, \lambda) > O(\lambda_d^0, \lambda_d^1)$ □

Proposition 6. *If* $(0, 0)$ *does not satisfy the capacity constraint,* $\psi_d(0) \geq \Psi_d$*, and* $N_d(\lambda_d^R, \lambda_d^R) \geq K_d$*,* $\psi_d(\lambda_d) = \Psi_d$*, the solution to (24) is* $(\lambda, \lambda_d^R$ *in which* $N_d(\lambda, \lambda_d^R) = K_d$*.*

Proof. From assumption (4), $\exists\lambda_d^0$ that satisfies $N_d(\lambda_d^0, \lambda_d^R) = K_d$

From lemma (3), if $\lambda_d^1 > \lambda_d^R$ then $\psi_d(\lambda_d^1) < \Psi_d$, which does not satisfy the affirmative action constraint. Therefore, $\lambda_d^1 \leq \lambda_d^R$

From proposition (3), λ_d^0 and λ_d^1 only solve the problem stated in (24) if $N_d(\lambda_d^0, \lambda_d^1) = K_d$.

Let us assume that there is no $(\lambda_d^0, \lambda_d^1)$, $\lambda_d^1 < \lambda_d^R$, that is feasible and solves the problem stated in (24).

Because $N_d(\lambda, \lambda_d^R) = N_d(\lambda_d^0, \lambda_d^1)$: $\sum_{i=\frac{\lambda}{\Delta}}^{\frac{1}{\Delta}} n_d^0(i\Delta) + \sum_{i=\frac{\lambda_d^R}{\Delta}}^{\frac{1}{\Delta}} n_d^1(i\Delta) = \sum_{i=\frac{\lambda_d^0}{\Delta}}^{\frac{1}{\Delta}} n_d^0(i\Delta) + \sum_{i=\frac{\lambda_d^1}{\Delta}}^{\frac{1}{\Delta}} n_d^1(i\Delta)$.

Therefore: $W = \sum_{i=\frac{\lambda}{\Delta}-1}^{\frac{\lambda_d^0}{\Delta}} n_d^0(i\Delta) = \sum_{i=\frac{\lambda_d^1}{\Delta}-1}^{\frac{\lambda_d^R}{\Delta}} n_d^1(i\Delta)$

Now:

$O_1(\lambda_d^R) - O_1(\lambda_d^1) = \sum_{i=\frac{\lambda_d^R}{\Delta}}^{\frac{1}{\Delta}} i\Delta n_d^1(i\Delta) - \sum_{i=\frac{\lambda_d^1}{\Delta}}^{\frac{1}{\Delta}} i\Delta n_d^1(i\Delta)$

Because $\lambda_d^1 < \lambda_d^R$:

$O_1(\lambda_d^R) - O_1(\lambda_d^1) = -\sum_{i=\frac{\lambda_d^1}{\Delta}-1}^{\frac{\lambda_d^R}{\Delta}} i\Delta n_d^1(i\Delta)$

Because $\frac{\lambda_d^R}{\Delta} = \max\{\frac{\lambda_d^1}{\Delta} - 1, \frac{\lambda_d^1}{\Delta}, \frac{\lambda_d^1}{\Delta} + 1, ..., \frac{\lambda_d^R}{\Delta}\}$:

$O_1(\lambda_d^R) - O_1(\lambda_d^1) \geq -\lambda_d^R \sum_{i=\frac{\lambda_d^1}{\Delta}-1}^{\frac{\lambda_d^R}{\Delta}} n_d^1(i\Delta)$

Similarly:

$O_0(\lambda) - O_0(\lambda_d^0) = \sum_{i=\frac{\lambda}{\Delta}}^{\frac{1}{\Delta}} i\Delta n_d^0(i\Delta) - \sum_{i=\frac{\lambda_d^0}{\Delta}}^{\frac{1}{\Delta}} i\Delta n_d^0(i\Delta)$

According to lemma (3): $\lambda_d^1 < \lambda_d^R$ and $N_d(\lambda, \lambda_d^R) = N_d(\lambda_d^0, \lambda_d^1)$ imply $\lambda_d^0 > \lambda$:

$O_0(\lambda) - O_0(\lambda_d^0) = \sum_{i=\frac{\lambda}{\Delta}}^{\frac{\lambda_d^0}{\Delta}-1} i\Delta n_d^0(i\Delta)$

Because $\frac{\lambda}{\Delta} = \min\{\frac{\lambda}{\Delta}, \frac{\lambda}{\Delta}+1, ..., \frac{\lambda_d^0}{\Delta}-1\}$:

$O_0(\lambda) - O_0(\lambda_d^0) \geq \lambda \sum_{i=\frac{\lambda}{\Delta}}^{\frac{\lambda_d^0}{\Delta}-1} n_d^0(i\Delta)$

Therefore:

$O(\lambda, \lambda_d^R) - O(\lambda_d^0, \lambda_d^1) \geq -\lambda_d^R \sum_{i=\frac{\lambda_d^1}{\Delta}-1}^{\frac{\lambda_d^R}{\Delta}} n_d^1(i\Delta) + \lambda \sum_{i=\frac{\lambda}{\Delta}}^{\frac{\lambda_d^0}{\Delta}-1} n_d^0(i\Delta) = W(\lambda - \lambda_d^R)$

From assumption (3), $N(\lambda_d^R, \lambda_d^R) > K_d = N(\lambda, \lambda_d^R)$ implies $\lambda > \lambda_d^R$. Therefore:

$O(\lambda, \lambda_d^R) - O(\lambda_d^0, \lambda_d^1) \geq 0$

□

Now, we combine the solutions from proposition (5) and (6) into a single formula.

Proposition 7. *If* $(0,0)$ *does not satisfy the capacity constraint and* $\psi_d(0) \geq \Psi_d$*, the solution to (24) is* (λ_d^R, λ)*, in which* $N_d(\lambda, \min(\lambda_d^R, \lambda)) = K_d$*.*

Proof. **Case 1**: $N_d(\lambda_d^R, \lambda_d^R) \geq K_d$, $\lambda_d^1 = \lambda_d^R$.

According to proposition (6), the solution of the problem of public schools is $(\lambda, \lambda_d^R$ in which $N_d(\lambda, \lambda_d^R) = K_d$.

Because $N_d(\lambda d^R, \lambda_d^R) \geq K_d$, assumption (3) makes $N_d(\lambda, \lambda_d^R) = K_d$ imply $\lambda > \lambda_d^R$.

Therefore: $\lambda_d^1 = \lambda_d^R = \min(\lambda_d^R, \lambda)$

Case 2: $N_d(\lambda_d^R, \lambda_d^R) < K_d$, $\lambda_d^0 = \lambda_d^1 = \lambda$.

According to proposition (5), the solution of the problem of public schools is λ: $N_d(\lambda, \lambda) = K_d$, $\lambda \leq \lambda_d^R$.

In such case: $\lambda_d^1 = \min(\lambda_d^R, \lambda)$.

□

Proposition (2) states that (9) solves (24) when $N_d(0,0) < K_d$.

Proposition (4) states that (9) solves (24) when $N_d(0,0) \geq K_d$ and $\psi_d(0) < \Psi_d$.

Proposition (7) states that (9) solves (24) when $N_d(0,0) \geq K_d$ and $\psi_d(0) \geq \Psi_d$.

Appendix D Proof for the main theorems

An important result to prove Proposition (1) is stated in Proposition (8).

Proposition 8. $0 \leq \frac{s^R_{pri}(\lambda, p_{pri}) - s^U_{pri}(\lambda, p_{pub}, p_{pri})}{s_{pub}(\lambda, p_{pub}, p_{pri})} \leq 1$

Proof. $\frac{s^R_{pri}(\lambda, p_{pri}) - s^U_{pri}(\lambda, p_{pub}, p_{pri})}{s_{pub}(\lambda, p_{pub}, p_{pri})} = \frac{\frac{exp(V_{pri}(\lambda) - \alpha(\lambda)p_{pri})}{1 + exp(V_{pri}(\lambda) - \alpha(\lambda)p_{pri})} - \frac{exp(V_{pri}(\lambda) - \alpha(\lambda)p_{pri})}{1 + exp(V_{pri}(\lambda) - \alpha(\lambda)p_{pri}) + exp(V_{pub}(\lambda) - \alpha(\lambda)p_{pub})}}{\frac{exp(V_{pub}(\lambda) - \alpha(\lambda)p_{pub})}{1 + exp(V_{pri}(\lambda) - \alpha(\lambda)p_{pri}) + exp(V_{pub}(\lambda) - \alpha(\lambda)p_{pub})}}$
$\frac{s^R_{pri}(\lambda, p_{pri}) - s^U_{pri}(\lambda, p_{pub}, p_{pri})}{s_{pub}(\lambda, p_{pub}, p_{pri})} = s^R_{pri}(\lambda, p_{pub}, p_{pri})$ Because $0 \leq s^R_{pri}(\lambda, p_{pub}, p_{pri}) \leq 1$, the proof is complete. □

Proof for Proposition (1):

Proof. We have to prove that $\frac{\partial S_{pri}}{\partial p_{pri}} + \frac{\partial S_{pri}}{\partial \lambda} \frac{\partial \lambda}{\partial p_{pri}} < 0$.
Applying the Leibniz rule to equation (6):
$-s_{pub}(\lambda, p_{pub}, p_{pri}) \frac{d\lambda}{dp_{pri}} + \int\limits_{\lambda}^{1} \frac{\partial s_{pub}(a, p_{pub}, p_{pri})}{\partial p_{pri}} = 0$
We know that $\lambda = \lambda(p_{pri}, p_{pub})$, so: $\frac{d\lambda}{dp_{pri}} = \frac{\partial \lambda}{\partial dp_{pub}} \frac{dp_{pub}}{dp_{pri}} + \frac{\partial \lambda}{\partial p_{pri}} \frac{dp_{pri}}{dp_{pri}}$.
Because public prices are exogenous ($\frac{dp_{pub}}{dp_{pri}} = 0$): $\frac{d\lambda}{dp_{pri}} = \frac{\partial \lambda}{\partial p_{pri}}$
Therefore: $\frac{\partial \lambda}{\partial p_{pri}} = \frac{\frac{\partial S_{pub}}{\partial p_{pri}}}{s_{pub}(\lambda, p_{pub}, p_{pri})}$ and we can write:
$\frac{\partial S_{pri}}{\partial p_{pri}} + \frac{\partial S_{pri}}{\partial \lambda} \frac{\partial \lambda}{\partial p_{pri}} = \frac{\partial S_{pri}}{\partial p_{pri}} + \frac{s^R_{pri}(\lambda, p_{pri}) - s^U_{pri}(\lambda, p_{pub}, p_{pri})}{s_{pub}(\lambda, p_{pub}, p_{pri})} \frac{\partial S_{pub}}{p_{pri}}$
From proposition (8), we can rewrite the previous expression as:
$\frac{\partial S_{pri}}{\partial p_{pri}} + \frac{\partial S_{pri}}{\partial \lambda} \frac{\partial \lambda}{\partial p_{pri}} = \frac{\partial S_{pri}}{\partial p_{pri}} + s^R_{pri}(\lambda, p_{pub}, p_{pri}) \frac{\partial S_{pub}}{p_{pri}}$
Using the definition of S_{pri} and S_{pub}, we write:
$\frac{\partial S_{pri}}{\partial p_{pri}} + \frac{\partial S_{pri}}{\partial \lambda} \frac{\partial \lambda}{\partial p_{pri}} = \int\limits_{0}^{\lambda} \frac{\partial s^R_{pri}(a, p_{pub}, p_{pri})}{\partial p_{pri}} da + \int\limits_{\lambda}^{1} \left[\frac{\partial s^R_{pri}(a, p_{pub}, p_{pri})}{\partial p_{pri}} + s^R_{pri}(\lambda, p_{pub}, p_{pri}) \frac{\partial s_{pub}(a, p_{pub}, p_{pri})}{\partial p_{pri}} \right] da$
We know that $\int\limits_{0}^{\lambda} \frac{\partial s^R_{pri}(a, p_{pri})}{\partial p_{pri}} da < 0$. Therefore, $\int\limits_{\lambda}^{1} \left[\frac{\partial s^U_{pri}(a, p_{pub}, p_{pri})}{\partial p_{pri}} + s^R_{pri}(\lambda, p_{pub}, p_{pri}) \frac{\partial s_{pub}(a, p_{pub}, p_{pri})}{\partial p_{pri}} \right] da < 0$ is a sufficient condition for $\frac{\partial S_{pri}}{\partial p_{pri}} + \frac{\partial S_{pri}}{\partial \lambda} \frac{\partial \lambda}{\partial p_{pri}} < 0$
In fact: $\frac{\partial s^U_{pri}(a, p_{pub}, p_{pri})}{\partial p_{pri}} + s^R_{pri}(\lambda, p_{pri}) \frac{\partial s_{pub}(a, p_{pub}, p_{pri})}{\partial p_{pri}} = -\alpha(a) s^U_{pri}(a, p_{pub}, p_{pri})(1 - s^U_{pri}(a, p_{pub}, p_{pri})) + s^R_{pri}(\lambda, p_{pri}) \alpha(a) s^U_{pri}(a, p_{pub}, p_{pri}) s_{pub}(a, p_{pub}, p_{pri}) = -\alpha(a) s^U_{pri}(a, p_{pub}, p_{pri})(1 - s^U_{pri}(a, p_{pub}, p_{pri}) - s^R_{pri}(\lambda, p_{pri}) s_{pub}(a, p_{pub}, p_{pr}$
Because $s^R_{pri}(\lambda, p_{pri}) < 1$: $1 - s^U_{pri}(a, p_{pub}, p_{pri}) - s^R_{pri}(\lambda, p_{pri}) s_{pub}(a, p_{pub}, p_{pri}) > 1 - s^U_{pri}(a, p_{pub}, p_{pri}) - s_{pub}(a, p_{pub}, p_{pri}) = s_{out}(a, p_{pub}, p_{pri}) > 0$
Therefore: $\frac{\partial s^U_{pri}(a, p_{pub}, p_{pri})}{\partial p_{pri}} + s^R_{pri}(\lambda, p_{pri}) \frac{\partial s_{pub}(a, p_{pub}, p_{pri})}{\partial p_{pri}} < 0$
Therefore: $\int\limits_{\lambda}^{1} \left[\frac{\partial s^U_{pri}(a, p_{pub}, p_{pri})}{\partial p_{pri}} + s^R_{pri}(\lambda, p_{pub}, p_{pri}) \frac{\partial s_{pub}(a, p_{pub}, p_{pri})}{\partial p_{pri}} \right] da < 0$
Therefore: $\frac{\partial S_{pri}}{\partial p_{pri}} + \frac{\partial S_{pri}}{\partial \lambda} \frac{\partial \lambda}{\partial p_{pri}} < 0$. □