[PRELIMINARY – DO NOT DISTRIBUTE]
College Expansion and Unequal Access to Education in Peru

José Luis Flor Toro
Northwestern U.

Matteo Magnaricotte
Northwestern U. *

September 2021

Abstract

Enrollment gaps are pervasive within developing countries, despite public investment and legislation aimed at democratizing access to college. We study the effect of opening new college campuses in previously unattended areas, a commonly proposed policy to reduce such gaps. Using Peruvian census data to estimate a difference in differences model, we find that enrollment increased by about 18% in the long term. However, effects are much larger for more advantaged individuals, widening preexisting gaps. We build a model of demand for education with heterogeneous preferences and probability of admission to study mechanisms, and simulate counterfactuals. We assemble a new administrative dataset on college applications and admissions for estimation. The results show that initial advantage and meritocratic criteria interact to reinforce educational inequality: the equalizing effect of increased proximity is entirely offset by the meritocratic admission process. Finally, we show that unequal access to quality high schools is a major driver of the ethnicity enrollment gap.

*We thank Matt Notowidigdo, Chris Udry, Nicola Bianchi, and Gaston Illanes for guidance and support. We thank Fabiola Alba Vivar, Matteo Camboni, Leander Heldring, Tai Lam, Joris Mueller, Michael Porcellacchia, Robert Porter, Mar Reguant, Lorenzo Stanca, Miguel Talamas, Francesca Truffa, for useful discussions and suggestions. Ricardo Sanchez and Mario Gonzales provided help with data and legislative context. Lucia Gomez Llactahuamani provided excellent research assistance.

1 Introduction

Increasing college attainment rates is important from a development perspective for regional development (Barro (1991), and Gennaioli et al. (2013)) and for the creation of upper tail human capital (Mokyr (2005), and Squicciarini, Voigtländer (2015)). This raises the problem of who should go to college. On the one hand, misallocation of talent affects negatively aggregate productivity (Hsieh et al. (2019)), and suggests the use of meritocratic allocation criteria even when talent is measured imperfectly. On the other, as an important factor for intergenerational income persistence (Chetty et al. (2020)), a preference for equality leads governments to intervene and increase equality of access for all citizens.[1] While some policies, like affirmative action, target specific groups that are considered to have structural disadvantages (Arcidiacono, Lovenheim (2016)), others either do not target specific population groups (like offering free community college enrollment) or target individuals who are considered meritorious like top-percent policies (Bleemer et al. (2020)).

The effects of meritocratic criteria for the allocation of college seats when initial resources are unequally distributed is an underexplored topic.[2] This interaction is particularly salient for policies that expand the higher education system and allocate seats based on standardized tests or admission exams. Many developing countries, despite a robust expansion of higher education capacity, are still facing large gaps in educational achievement along ethnic and socioeconomic dimensions, and between men and women (e.g. see Ferreyra et al. (2017) for a comparative assessment of Latin American countries). Because of its social mobility implications, some governments have set the reduction of gaps in access as a first-order policy goal (Ministerio de Educación (2020)).

We focus on the case of Peru, a country that experienced a rapid expansion of the higher education sector in recent decades, similarly to most Latin American countries[3], and is currently facing large gaps related to socio-economic status (SES), race, and gender.[4] In the first

[1] Another tightly connected argument regards the fact that students with more initial resources (e.g. family wealth, more educated, or connected parents) have an advantage over similarly talented students without those resources.

[2] Arcidiacono et al. (2021) studies how a non-meritocratic selection rule affects the composition of the student body at Harvard. Akbarpour et al. (2020) represents a recent attempt to model market design under non-market allocation rules for scarce public goods.

[3] Ferreyra et al. (2017) provides a description of the sector for Peru and other LATAM countries.

[4] See the "National plan for higher, technical, and productive education" (or PNESTP by the Spanish acronym) for stated government objectives (Ministerio de Educación (2020)); see SUNEDU (2020) and Sánchez et al. (2021)

part of the paper, we quantify the effects that expansions of capacity through the creation of new college campuses since 1960 had on enrollment of the local population and of several subgroups. In the second part, we build and estimate a discrete choice model of demand for education to study the relative importance of meritocratic selection and preference heterogeneity of different groups in determining gaps in enrollment. Finally, we simulate and discuss alternative policies aimed at reducing gaps and increasing achievement.

In the first part, we use a difference in differences approach to learn how exposure to the higher education sector's expansion affects college enrollment and completion. We see that, when a new college campus opens up in a location that did not have any previously, enrollment increases among cohorts 18 or younger[5] by approximately 1.3 percentage points (p.p.) or 10.3%; completion increases by 1p.p. or 9.4%. Private college openings are associated with larger effects than public ones, while enrollment in other forms of higher education (corresponding to community college in the US) is not significantly affected. This increase is not evenly distributed among different groups of the population: women and the ethnic majority have higher increases in enrollment and completion. In particular, we find a much smaller increase in enrollment among ethnic minorities following entry of selective public colleges. Because we cannot observe application behavior from the census, it is not possible to understand whether these results are due to heterogeneous reactions in application or to differential probability of admission conditional on applying. Understanding the mechanism at the root of this heterogeneity is fundamental to design policies that mitigate preexisting educational gaps.

To tease apart the importance of preferences (affecting application behavior) and selectivity conditional on application in determining the heterogeneous effects of college opening on different groups, we model the decision process of students and the selection process of public colleges. We leverage a separate dataset, tracking individuals in the cohort graduating from high school in 2018 through their educational career, including college application, admission results, and enrollment, to estimate the parameters of this model. We observe that proximity to campus is particularly valued by individuals with less educated parents, but less advantaged students also have a lower probability of being admitted conditional

for reports on stratification of access.

[5] High school graduation is at 16 in Peru. More details about the educational system can be found in Section 2

on applying. This is an example of how the two channels can affect in different ways the achievement gap.

For the simulation of counterfactuals, we estimate a flexible specification of utility that allows both distance and high school quality to affect admission probabilities and preferences. As a first counterfactual, we simulate the opening of new college campuses in areas that do not have any. This simulation serves two purposes: first, it provides a way to externally validate our model by comparing the predicted effects to those estimated from the difference in differences model ; second, it allows us to learn more about the importance of selection in shaping achievement gaps. We find that the effects predicted by the demand model are comparable to those found in the previous section, also showing the same heterogeneity along the gender and ethnicity dimensions. As a direct consequence of this heterogeneity, we find that opening college campuses *increases* ethnicity and gender achievement gaps by 8% and 6% respectively, while it leaves unaffected the SES gap (proxied by parental education).

We also simulate the effects of an affirmative action policy where the probability of admission of ethnic minority students is increased by $10p.p.$[6], and a large educational investment policy aimed at secondary schooling that increases quality at below median quality high schools to the level of those above median. The first simulation finds a reduction in the ethnicity gap of 24% but no effect on gender and SES gaps: this is consistent with the highly targeted nature of the intervention. The second policy is predicted to almost close (-87%) the ethnicity gap while increasing the gender and SES gaps. Because ethnic minority students disproportionately attend low quality high schools, they also benefit disproportionately by interventions that raise its quality. On the other hand, high school quality appears to increase utility from college more for females, high SES, and majority students, hinting at the presence of complementarities.[7]

Finally, we study the importance of selection in determining these outcomes, by keeping preferences fixed after policy changes. We find that the selection process increases gaps along all measured dimensions after college campuses are opened. This is explained by

[6] At the same time we do not decrease the probability of admission of other students, which corresponds to an increase in public college capacity to accommodate the increase in admissions from minority students.

[7] Heckman, Mosso (2014) considers complementarities to be a hypothesis in need of further validation, reporting Pop-Eleches, Urquiola (2013) and Gelber, Isen (2013) as examples of studies contributing evidence of its plausibility.

complementarities between proximity, and demographics of the advantaged groups. The same conclusions are drawn for the counterfactual of high school quality improvement.

Our analysis contributes to several strands of literature. First, we argue that college campus openings should be seen as an additional policy tool to increase enrollment rates and, potentially, as a way to reduce attainment gaps of URM and lower SES students.[8,9] In this context, we contribute to the literature estimating effects of policies for equitable access to higher education. In the US context, the two main policies implemented are affirmative action (see Arcidiacono, Lovenheim (2016) for a recent review) and "top percent" programs (Black et al. (2020) and Bleemer et al. (2020) are recent examples), with the main concern regarding the potential mismatch between the students who benefit of the policies and the institutions that they gain access to. In Brazil, Mello (????) studies the effects and interaction of affirmative action and centralization of admission. Other papers studying interventions to reduce attainment gaps have shown heterogeneous effects (e.g. Carrell, Sacerdote (2017)) and the potential for the emergence of private responses that offset the equity gains from the policy (e.g. Chatterjee et al. (2020)).

We contribute to this literature by documenting the heterogeneous effects on higher education achievement of an expansion of supply in higher education, a commonly used policy tool that few studies have analyzed, and focusing on the factors that favor a more equitable distribution of access.[10] Russell et al. (2021) compares sites where a college campus was open between 1839 and 1954 with runner-up locations, finding that cumulative exposure to colleges significantly affects college attainment today. We use a different empirical strategy from these papers to show that effects of campus opening are heterogeneous depending on the type of college (private or public) and on students' characteristics. Our focus also differs in that we look at short and medium run effects of the opening of the first campus in each

[8]See Bailey, Dynarski (2011) and Hanushek et al. (2020) for descriptive evidence of these gaps and their evolution over time in the US. Both papers document large gaps for lower SES and URM students, and agree that the gaps did not decrease over time.

[9]Heckman, Mosso (2014) notes that universal provision of policies need not promote equality of outcomes when complementarities are present and more advantaged households are better able to take advantage of them.

[10]Kyui (2016) and Belskaya et al. (2020) leverage the opening of new college campuses in Russia during the 90s as IV to estimate returns to college and show that returns are higher in areas where no local option was previously available. These papers rely on some version of distance instruments, like the seminal Card (1993). We focus on the "first stage" of these papers, looking at how increasing local college availability affects students' educational decisions: our findings provide insights about the robustness of IV assumptions in this context and improve the understanding of the complier population.

area. Finally, through the estimation of a model for educational demand, we delve into the drivers of our findings, and build counterfactual scenarios for different policies.

Second, we build on the literature estimating students' demand for education[11], recently employing tools developed in Industrial Organization.[12] Carneiro et al. (2016) is a recent example of this literature, estimating demand for public and private schooling: they find that near, low-fee, private schools are important for households' welfare. We find that private colleges differ importantly from public ones, both because they have a much higher cost, and because they are typically non-selective in the Peruvian context. While we find that they expand access to higher education the most, we also find concerning results regarding their benefits for students. Jacob et al. (2018) estimates a discrete choice model where students care about college characteristics, and finds heterogeneity in students' preferences to be important in determining the supply of amenities at different colleges. Our estimation procedure is closest to the two-steps Maximum Likelihood procedure used in Hastings et al. (2017) to determine the effects of marketing efforts of insurance providers in Mexico.

Third, we contribute to the literature on intertemporal decisions of students. The effect of campus openings is theoretically ambiguous: human capital models (e.g. Eisenhauer et al. (2015)) say that the availability of more college options should increase returns to high school graduation and decrease drop out rates; on the other hand, signalling models, as in Bedard (2001), predict that dropout rates should increase as lower ability students lose the incentive to mimic higher ability ones. Our model makes clear that selectivity mediates such effect, clarifying our difference in differences findings of positive effects on high school completion being proportional to the probability of admission.

Our findings also contribute to the vast literature on effects of increased education. Our results align qualitatively with previous findings on reductions in fertility (Osili, Long (2008), Duflo et al. (2015)), and on increases in employment (Beuermann et al. (2018) and Bautista et al. (2020)). Fertility results are proportional to enrollment effects. Interestingly, our results for employment are large and significant only for public colleges, while private ones have non-significant, economically small estimates: we also find a reduction in migration within the country with the same pattern. This latter pattern might be due to differences in quality

[11] For early examples, see Alderman et al. (2001) for school and Fuller et al. (1982) for post-secondary education.

[12] For more examples of papers applying IO tools to education markets see: Neilson (2013); Hastings et al. (2016); Kapor et al. (2017); and Dinerstein, Smith (2014).

of public and private institutions.

The rest of the paper is organized as follows: in Section 2 we describe the characteristics of the Peruvian educational system and context that are relevant to our analysis, and Section 3 describes the main datasets that are used in our empirical analysis; in Section 4 we estimate a difference-in-differences model and discuss its results. Section 5 delves into the mechanisms explaining the heterogeneous effects of college expansion and simulates counterfactual policies to increase college enrollment and reduce attainment gaps in the population; Section 6 concludes the paper.

2 Institutional Context

2.1 Higher education system in Peru

Peru's primary education starts at the age of 6 and lasts 6 years, while secondary schooling starts at 12 and lasts 5 more years.[13] Within two years from high school graduation, most students apply and enroll in majors whose formal length is usually 5 years (see Figure **??** in the Appendix for the full distribution of time to college). In practice, many students take 6-7 years to graduate.

Public universities only charge small administrative fees, while private universities do not face restrictions on tuition setting: for this reason, and thanks to a perception of higher average quality, public colleges, whose admission procedures are decentralized, have much higher ratios of applicants to admitted students than private ones. While no centralized admission system exists, all public colleges admit students mainly through admission tests that are generally held twice a year. Students can take admission tests at multiple colleges and can re-take the test without limits: each attempt requires the payment of a small fee that varies across universities. About 20% of applications at public colleges are successful. Private universities have a greater variance, with only a few of them display a significant degree of selectivity and the rest admitting almost any applicant. Admission procedures at private

[13]Peru has an area approximately 3 times the one of California (see Figure **??**) and population of roughly 31 millions: it is organized into 25 regions and 196 provinces (approximately corresponding to US states and counties, respectively). Peru is quite representative of several other Latin American countries, with a large share of its population concentrated around the capital and long distances between other populated areas. For a comparison of the Peruvian higher education sector with other Latin American countries, see Ferreyra et al. (2017).

universities are not standardized. Scholarship availability and financial aid instruments are extremely limited.

The main alternative to colleges are the so-called technical institutes, which correspond loosely to community colleges in the US context: they are in majority private (69% of the total according to the 2020 census of educational institutions) and provide vocational education without the possibility to transfer and obtain a university degree.[14]

Congress is the sole institution endowed with the power to create new public universities.[15,16] Recognizing the need for a more diffused supply of higher education, Peru has opened new public universities and university branches in 35 provinces out of the country's 196 since 1990. In addition, with the *law 882 on the promotion of private investment in education* of 1996, the Peruvian Congress allowed for profit institutions in the higher education market.[17] One of the main purposes for the new regulation included the democratization of access (Cuenca (2015)). The same period saw the creation of many new private universities, for a total of 152 new branches. Figure 1 shows the distribution of university campuses in 2017 in each province.[18]

In 2012 Congress passed a moratorium on the creation of new universities to limit the uncontrolled entry of universities of dubious quality and allow for the institution of minimum quality standards.[19] The moratorium limited the creation of public and private universities alike and prohibited the opening of new branches for already established institutions. Starting in 2015 a new public body, named *SUNEDU*[20], was given supervisory duties over the quality of Peruvian universities. This monitoring activity culminated between 2018 and

[14] The law N° 30512 of Peru regulates these institutions.

[15] The last 30 years have seen a tight relationship between Peruvian politics and higher education, with Congress expanding the role of the state at first and implementing the entry moratorium later on, and with private college owners successfully entering politics as Congress members.

[16] Congress is prohibited from making changes to the budget law or increasing public expenditure: *Artículo 79* **Restricciones en el Gasto Público** Los representantes ante el Congreso no tienen iniciativa para crear ni aumentar gastos públicos, salvo en lo que se refiere a su presupuesto. El Congreso no puede aprobar tributos con fines predeterminados, salvo por solicitud del Poder Ejecutivo. 5 Artículo modificado por el Artículo Único de la Ley Nº 26472, publicada el 13 junio 1995 en el diario oficial El Peruano.

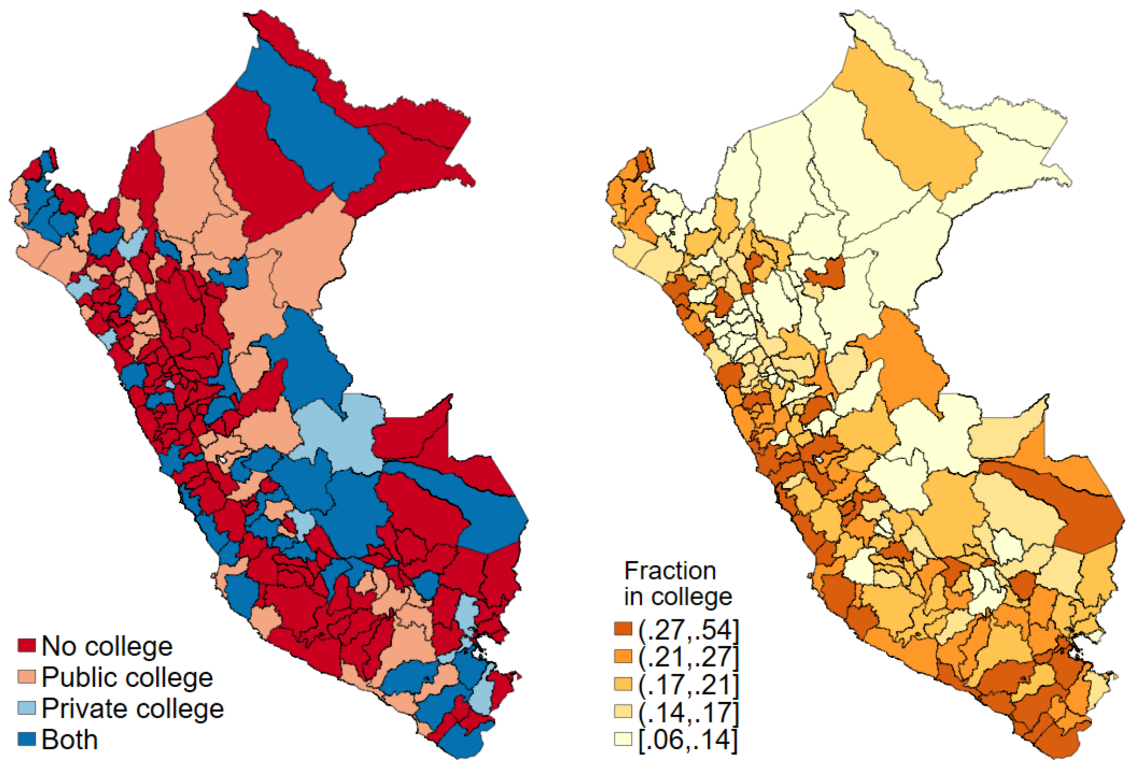
[17] Non profit institutions could be established even before said law, as regulated by article 26 of the same *Ley Universitaria*.

[18] In the empirical analysis of this paper, we will define college campuses as all the buildings belonging to a university in a given province. Consequently, if a university has buildings in two different provinces, we will say that it has two campuses.

[19] The 2012 moratorium has since been renewed several times. The latest renewal is contained in the Law 31193 of Peru.

[20] Acronym for National Superintendence of Higher Education in Spanish.

Figure 1: Geographical College Availability and Access in Peru

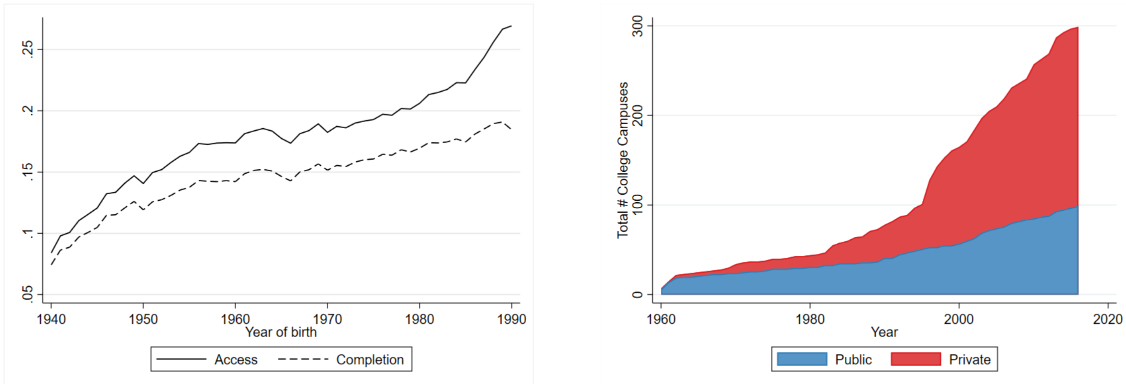


Panel A: Geographical Distribution of College Campuses

Panel B: Geographical Distribution of College Enrollment

Notes. Panel A shows the geographical distribution of college campuses in the 196 provinces of Peru in 2017. Section 3 reports how the location of each campus of each university is obtained using an administrative census of education institutions. Panel B shows the fraction of individuals born in 1996 declaring college enrollment in the 2017 Census, by province of birth.

Figure 2: College Availability and Access over Time in Peru



Panel A: Access and Completion of College by Cohort of Birth

Panel B: College Openings over Time, by College Type

Notes. Panel A plots the rates of college enrollment and completion for the cohorts born in the period 1940-1990 in Peru, according to the 2017 Census. Both rates increase over time, with a sharp increase for access for cohorts born after 1980. Panel B reports the number of college campuses open in each year, by management type (public or private). The process used to obtain the opening year of each campus is described in Section 3. In 1996, Congress authorized the creation of for-profit colleges, leading to a faster increase in college openings.

2020 with the effective closure of about a third of Peruvian universities, all of them private except for one public college.

In the moment of maximum expansion, in 2017, Peru had about 300 different college campuses, with two thirds of them being private (see Figure 2 for the evolution over time).

2.2 Access to university

College enrollment rates have grown over time, reflecting the general economic development of Peru and increasing from less than 10% for individuals born in 1940 to almost 30% for the 1990 cohort, as shown in panel A of Figure 2. As can be seen in the Figure 2, this growth has been paralleled by the opening of new college campuses. This expansion has allowed for more students to attend colleges close to their places of birth: indeed, student mobility in the higher education setting is fairly limited. Analyses of the 2017 Census of population and the 2010 Census of university students, reported in the Appendix, show that approximately 90% of university students enroll in a college located in their province of birth.

One important political and policy concern has regarded the participation to postsecondary education of ethnic minorities. Peru has a very diverse population, with a quarter of individuals born in the period 1960-2000 self-identifying as belonging to different groups of indigenous populations according to the 2017 Census[21] and indigenous populations have mobilized to increase their access to higher education. This has led to the opening of several "intercultural universities" that provide bilingual education.

As a matter of fact, the racial gap in enrollment is one of the largest, as highlighted in Figure **??** and in other reports (SUNEDU (2020)). Socioeconomic status, represented by parental education, and rurality-urbanicity also emerge as important factors. Females have a smaller but relevant advantage over male students. In 2016 a report of the OECD (OECD (2016)) highlighted gaps in access among the challenges that Peru still faces.[22] Disparities are generally smaller at the primary level and increasing at higher levels, with regional gaps being among the widest. Figure 1 reports college enrollment rates at the province level for the cohort born in 1996, as recorded in the 2017 Census. In response to these gaps, in 2020 the Ministry of Education issued the National Plan for Higher and Technical Education (PNESTP by the Spanish acronym, Ministerio de Educación (2020)) highlighting as the first policy goal the reduction of gaps in access to higher education.

3 Data

The empirical analysis in this paper relies on several sources of administrative and survey microdata.

In Section 4 we estimate a difference in differences model using the 2017 Peruvian Census of the Population. The Census includes questions on basic demographics, educational achievement, and employment.[23] We produce aggregates by cohort for each province of birth for the variables of interest. The same variables are also produced separately by gender and by native language.

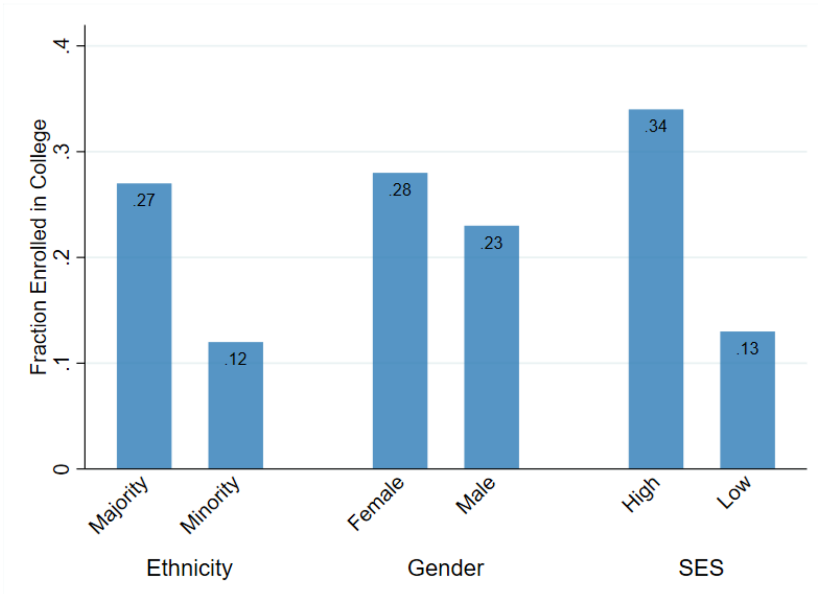
We use the variable reporting the first language learned as our definition of ethnicity, creating a dummy variable for all individuals that first learned a language other than Span-

[21] Figure **??** in the Appendix shows the percentage of people in each province that learned a language different from Spanish as their first language: we can observe a large level of segregation, with concentration of minorities in the Andean area and clear Spanish-speaking majorities along the coast.

[22] For more descriptive evidence on the state of Peruvian higher education in recent years, see SUNEDU (2020).

[23] The complete questionnaire can be found at the following link.

Figure 3: Percent of Enrollees by Sociodemographic Groups



Notes. Fraction of college enrollees by sociodemographic groups, according to ENAHO.

ish.[24] The Census provides 8 suggested options, but, considering responses other than the suggested ones, documents the presence of more than 50 surviving Indigenous languages. Figure **??** shows the percentage of the population that learned an Indigenous language as their first at the province level in 2017.

We report summary statistics for the variables used from this dataset in Table 1. Statistics are reported separately depending on whether the province of birth had a college campus in 2017. We can see that individuals from provinces with one or more college campuses are more likely to learn Spanish as their first language, have higher educational achievement, are more likely to be employed, and to give birth at least once (for women). It can be observed that age in 2017 ranges between 26 and 75: this is because, as discussed in the next section, we exclude individuals born before 1942 or after 1991 from our analysis.

We build our treatment variables using information on the time of opening of new campuses. However, administrative records generally report the year of creation of universities, but not the year in which additional campuses admitted their first cohort. First, we use

[24] An alternative definition of ethnicity could be constructed using the answer to the question "Given your customs and ancestors, how do you feel or consider yourself?". However, the answer to this question is inherently subjective and potentially affected by the treatment variable. For this reason, we prefer as measure of ethnicity the more objective answer to the question "What is the language that you first learned to speak as a child?".

Table 1: Summary Statistics of 2017 Census

	N	Mean	SD	Min	Max
Has College					
Age	11729322	44.5	13.0	26	75
Female	11729322	0.52	0.50	0	1
Nat. Spanish Speaker	11729322	0.86	0.35	0	1
Completed High School	11729322	0.66	0.47	0	1
Attended College	11729322	0.23	0.42	0	1
Completed College	11729322	0.18	0.38	0	1
Employed	11729322	0.60	0.49	0	1
Gave Birth	5978651	0.85	0.36	0	1
No College					
Age	3386324	46.6	13.5	26	75
Female	3386324	0.51	0.50	0	1
Nat. Spanish Speaker	3386324	0.63	0.48	0	1
Completed High School	3386324	0.44	0.50	0	1
Attended College	3386324	0.10	0.30	0	1
Completed College	3386324	0.085	0.28	0	1
Employed	3386324	0.52	0.50	0	1
Gave Birth	1729881	0.90	0.29	0	1

Notes. Summary statistics for the relevant variables of the 2017 Peruvian Census. The top panel reports the statistics calculated using individuals born in provinces that are treated at some point in time with the opening of a college campus; the bottom panel is calculated using individuals born in "pure control" provinces, that have never had a college campus. Age is restricted between 26 and 75. Further restrictions, leading to the exclusion of some provinces (including the capital Lima), are discussed in Section 4.1. Presence of a college campus is obtained through administrative records, as described in Section 3. *Age* is declared at the time of the Census (2017); *Female* represents gender as reported; *Nat. Spanish Speaker* is a dummy for whether the first language learned as a kid was Spanish; *Completed High School* is a dummy for secondary schooling completion; *Attended College* is a dummy for college enrollment (but not necessarily completion); *Completed College* is a dummy for college completion; *Employed* is a dummy for self reported employment in the week previous to the Census measurement; *Gave Birth* is a dummy for having delivered a living baby (only applies for women 12 or older).

administrative records on the registration of majors to identify the location of university branches.[25] Then, we combine several administrative sources and institutional information from the universities to infer the opening of each additional campus.[26] The geographical location of colleges, and time distribution of the reconstructed year of entry by type of college ownership are shown in Figures 1 and 2, respectively.

In Section 5, we introduce a novel dataset combining information from high school records, college applications and enrollment. This allows us to track students in the first year of high school through college enrollment. High school records are included in the SIAGIE dataset maintained by the Ministry of Education; similarly, application (successful and unsuccessful) and enrollment data come from the SIRIES dataset. More information about dropout, application, and enrollment rates is reported in Section 5.

We complement this data with three main variables. First, we define the ethnicity of each individual based on the district (approximately correspondent to US census tracts) of birth. An individual is considered from an ethnic minority if born in a district where at least 80% of the population learned a language other than Spanish as their first based on the 2017 Census. Second, we use information on the high school attended by each individual and the dataset on campus locations previously described to identify the closest educational options for each individual. Finally, we collect data on tuition levels: we combine several sources to build a novel dataset with the most detailed tuition information to our knowledge. While cost measures are aggregated at the university level, disregarding differences in costs across campuses, our tuition measure is defined at the campus level by taking the median cost of all the majors offered there. This allows us to have more precise cost information without relying on self-reporting of students in surveys. Details on how the data is collected are available in the Appendix.

Other sources of data used to produce auxiliary evidence, like the National Survey of Households (*ENAHO*), are briefly described as needed along with their analysis.

[25] Similar information can be obtained through the ESCALE system of the Ministry of Education.

[26] A detailed description of the process is included in the Appendix.

4 Difference-in-Differences for Campus Openings

4.1 Empirical specification

In order to answer the question of how much the opening of new college campuses increases local enrollment and which groups are most affected by these openings, we use a difference in differences approach. This means comparing cohorts that have access to a local college when graduating from high school to earlier cohorts that did not have such access, while controlling for trends in enrollment. This set up suggests a model of the form:

$$y_{t,p} = \sum_{\pi} \sum_{\tau} \beta_{\pi,\tau} \mathbb{1}\{p = \pi\} \mathbb{1}\{\tau = t - g(p)\} + \phi_t + \psi_p + e_{t,p}$$

where $g(p)$ is the period in which a college opened in province p, and t indexes different cohorts. $\beta_{\pi,\tau}$ is the parameter of interest: the effect of exposure to the campus opening on the outcome $y_{t,p}$ (e.g. enrollment rates). The term $\beta_{\pi,\tau}$ highlights the possibility of effects being heterogeneous for different provinces (π) and depending on exposure length (τ). Allowing for heterogeneity and dynamics requires us to avoid standard "two-way fixed effects" (TWFE) regressions, as they have been shown to be problematic in such setups (Goodman-Bacon (2021), Baker et al. (2021)).

The frailties of TWFE come from the inclusion of already-treated groups within the comparison group: if treatment effects are heterogeneous, the TWFE estimator will be biased.[27] Several recent papers have introduced solutions to address these problems, by making sure that only never-treated or not-yet-treated units are included in the comparison group. Similar to the basic difference in differences, these papers rely on parallel trends assumptions to build consistent estimators that do not suffer from the same problems.[28]

We use the estimator proposed in Callaway, Sant'Anna (2020) and provide estimates using only never-treated units, and including not-yet-treated units as well. This estimator is robust to dynamic effects (e.g. increasing with length of exposure to treatment) and effects being heterogeneous across provinces (e.g. because of colleges having different sizes or ma-

[27] Dynamic effects, e.g. when treatment effect is growing as time from the event passes, also lead to misspecification and inconsistency when a constant effect is assumed. This can be addressed by estimating treatment effects relative to even time (Borusyak et al. (2021)).

[28] See as examples of such new estimators Callaway, Sant'Anna (2020), Borusyak et al. (2021), and Sun, Abraham (2020).

jors). We benchmark these estimates against the TWFE estimator.[29] The difference in difference estimator in Callaway, Sant'Anna (2020) also allows for a double robust procedure that is consistent if either a propensity score or outcome regression working models are correctly specified (see details in Sant'Anna, Zhao (2020)): we report the results in Appendix **??**. Standard errors are calculated using the bootstrap procedure suggested in the paper and implemented in the statistical software provided by the authors.[30]

Starting from the full-count individual level data of the 2017 Peruvian Census, we obtain enrollment rates (and other measures) at the age cohort by province level. We use the *province of birth* to match individuals to the creation of new campuses. Cohorts will be considered treated if their birth province had a university campus opened by the time they were 18.[31] We only include in the analysis those individuals born in the period 1942 to 1991.[32]

We estimate the difference in differences model for everyone and separately for different demographic groups to study the heterogeneity of treatment effects. For each specification, the Appendix reports "event study figures" where we plot treatment effects estimates for each cohort relative to the opening of the campus: these figures allow for a visual inspection of pre-trends and an assessment of the effect's dynamics. In addition, we also divide treatment events in two groups depending on whether the opened campus is public or private. "Event study figures" for these specifications (using the whole population or separate demographic groups) are also reported in the Appendix.

Contrary to TWFE, the estimator in Callaway, Sant'Anna (2020) separates estimation from aggregation of effects. This requires the researcher to choose the appropriate aggregation and summary statistic. We report dynamic effects relative to treatment time with provinces in graphical form; average treatment effects reported in table format are unweighted averages of the effects for all cohorts and for cohorts in school at the time of opening (aged 6-16) at the time of first treatment.

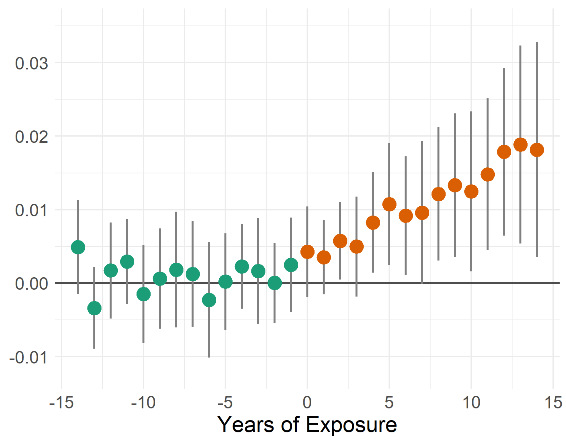
[29]The appendix describes the exact specification used for the TWFE estimator.

[30]Standard errors are clustered at the treatment level, and critical values in event study graphs are robust to multiple hypothesis testing.

[31]Section 3 and the Appendix describe the process used to construct the dataset with the opening of campuses.

[32]Limiting our analysis to cohorts born after 1942 helps to reduce bias from differential mortality rates (in 2017 an individual born in 1942 would be 75 years old). Limiting our analysis to cohorts born before 1992 allows time for the outcomes of the youngest individuals to fully realize (in 2017 an individual born in 1991 would be 26 years old).

Figure 4: Diff-in-Diff Estimates for Educational Variables



Notes. Event study graphs for the effect of the opening of any kind of college campus on enrollment (left panel) and completion (right panel). Source: 2017 Peruvian Census of the Population.

4.2 Results

Table 3 reports the average treatment effect for models estimated with the difference in differences estimator described in Section 4.1. We report results where the outcome is whether an individual ever enrolled in a college, whether she completed it, and whether she enrolled in a community college.

First, we can see from Figure 4 that we cannot reject the absence of pre-trends, suggesting the validity of the parallel trends' assumption. In the same Figure, we also observe that effects are increasing over time, and estimates are statistically significant for individuals 16 and younger, who are still in high school at the time of first admission. The first row of Table 3 shows that college openings significantly increased enrollment and completion at universities, with a larger effect for private institutions than for public ones; however, they did not affect enrollment at community colleges. Larger effects for private openings might be due to a preference for private institutions over public ones, or, most likely, to differences in capacity constraints. Public colleges are generally very selective, with only one application out of five being successful on average, while 78% of applications to private colleges are successful. This is because private universities can expand their capacity faster than public ones that are financed only through public funds[33] and are unable to expand in response to

[33]The amount of funds for the year is agreed between each public university and the Ministry of Economics and Finance. Historically, these funds have been adjusted based on previous years' funding and enrollment

Table 2: Diff-in-Diff: ATE Estimates by College Type

	College Enrollment			College Completion		
	All	Private	Public	All	Private	Public
TWFE						
	.0207***	.0390***	.0162**	.0160***	.0297***	.0126**
	(.00481)	(.0105)	(.00512)	(.00384)	(.00831)	(.00409)
All Cohorts						
	.0245***	.0427***	.0224***	.0176**	.0309***	.016**
	(.0081)	(.0073)	(.0084)	(.0069)	(.0058)	(.0072)
School Age						
	.0116***	.0156**	.0104***	.0091***	.0133*	.008***
	(.0024)	(.0073)	(.0023)	(.0022)	(.0074)	(.002)

Notes. This table shows the estimated ATE for the difference in differences model using the estimator proposed in Callaway, Sant'Anna (2020), as described in Section 4. The dependent variables *College Enrollment*, *College Completion*, and *Community College Enrollment* are dummy variables. The ATE is obtained by averaging the dynamic effects estimated on all post-treatment coefficients. Estimation is carried out using only Not-Yet-Treated provinces as comparison group. Different rows report the estimated ATE for different subgroups of the population. Different columns report estimates for three different treatment definitions for each dependent variable: in the first model (indicated by "All") we define as treated a cohort aged 18 or younger at the time of first opening of *any* college campus; in the second (indicated by "Private") only openings of *private* college campuses are considered; in the third (indicated by "Public") only openings of *public* college campuses are considered. Standard errors (in parentheses) clustered at the treatment level are calculated using the bootstrap procedure developed in Callaway, Sant'Anna (2020) and implemented in the provided software. Dependent variables' means for 20-year-old individuals at the time of opening of a new campus are 0.130 for college enrollment, 0.109 for completion, 0.135 for community college enrollment. Stars represent statistical significance of the single hypothesis test. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Diff-in-Diff: Heterogeneous Effects

	College Enrollment			College Completion		
	All	Private	Public	All	Private	Public
Gender						
Female	.0126***	.0161**	.0116***	.0104***	.0151*	.0091***
	(.0025)	(.0075)	(.0028)	(.0024)	(.0089)	(.0023)
Male	.0104***	.0147**	.0092***	.0079***	.0113*	.0069**
	(.0028)	(.0073)	(.0029)	(.0028)	(.0068)	(.0029)
Gender Gap	.0022	.0058*	.0018	.0025	.0081***	.0018
	(.003)	(.0035)	(.0038)	(.0026)	(.0031)	(.0031)
Ethnicity						
Majority	.014***	.0178**	.013***	.0117***	.0136	.0111***
	(.0032)	(.009)	(.0031)	(.0028)	(.0102)	(.0027)
Minority	.0082**	.0125	.0061	.0048*	.006	.0041
	(.0035)	(.0078)	(.0038)	(.0029)	(.0059)	(.0035)
Ethnic Gap	.0087*	.0221*	.0103**	.0098**	.0201**	.0103**
	(.0052)	(.0115)	(.0048)	(.0046)	(.0092)	(.0044)

Notes. This table shows the estimated ATE for the difference in differences model using the estimator proposed in Callaway, Sant'Anna (2020), as described in Section 4. The dependent variables *College Enrollment*, *College Completion*, and *Community College Enrollment* are dummy variables. The ATE is obtained by averaging the dynamic effects estimated on all post-treatment coefficients. Estimation is carried out using only Not-Yet-Treated provinces as comparison group. Different rows report the estimated ATE for different subgroups of the population. Different columns report estimates for three different treatment definitions for each dependent variable: in the first model (indicated by "All") we define as treated a cohort aged 18 or younger at the time of first opening of *any* college campus; in the second (indicated by "Private") only openings of *private* college campuses are considered; in the third (indicated by "Public") only openings of *public* college campuses are considered. Standard errors (in parentheses) clustered at the treatment level are calculated using the bootstrap procedure developed in Callaway, Sant'Anna (2020) and implemented in the provided software. Average effects are calculated for cohorts aged 6-16. We exclude provinces with less than 2500 individuals for the subgroup. Dependent variables' means for 20-year-old individuals at the time of opening of a new campus are 0.130 for college enrollment, 0.109 for completion, 0.135 for community college enrollment. Stars represent statistical significance of the single hypothesis test. *** p<0.01, ** p<0.05, * p<0.1.

an increase in applications.

In the Appendix, we show that enrollment at community colleges does not appear to be affected in a statistically significant way. Given that most estimated effects are positive and below $0.2 p.p.$ in absolute value, we consider the impact on community college enrollment to be negligible.[34]

Dividing the population by gender and ethnicity, we can see that the treatment effects are driven by women and individuals belonging to the ethnic majority. In particular, we can see that individuals in the ethnic minority see almost no increase following the opening of a public campus and less than half of the majority increase for private ones. This implies a widening of the gap for minority students who don't get to benefit from the subsidized public option. The heterogeneity by gender is not as strong, but is consistent with women being more responsive on the enrollment margin, as found previously in related situations (e.g. see Carrell, Sacerdote (2017)). Different responses to the policy can be explained by two main mechanisms. The first one works through preferences: either individuals from one group are closer to the choice margin, so that small increases in the utility of enrollment can induce more of them to apply/enroll, or they obtain a larger benefit from attending college close to home (again inducing them to apply and enroll). The second possibility is that different groups have different chances of admission conditional on application, which may be caused, for example, by differences in their previous education and resources. In Section 5, we will leverage detailed information on individuals' application choices to disentangle the two mechanisms.

The census also allows us to observe a number of other outcomes that are potentially affected by college openings and enrollment. Table 4 reports estimates of the average treatment effects on high school completion, fertility, employment and migration.

numbers; only recently performance indicators have been introduced in the process.

[34]The confidence interval for all openings and all individuals has a lower bound of $-0.36 p.p.$, or about one fourth of the point estimate for college enrollment.

Table 4: Other Outcomes: Diff-in-Diff ATE Estimates by College Type

Dependent Variable:	HS Completion			Has Kids (only Women)			Employment			Migration		
	All	Private	Public	All	Private	Public	All	Private	Public	All	Private	Public
All	.0018	.04*	-.002	-.0151**	-.0265**	-.0142**	.0263**	.0084	.0275**	-.0327**	.0019	-.0362**
	(.0146)	(.021)	(.0152)	(.0071)	(.0114)	(.0071)	(.011)	(.0067)	(.0121)	(.0151)	(.0125)	(.0156)
Female	.0105	.0457*	.0065	-.0151**	-.0265**	-.0142**	.0214*	.0118	.0216*	-.0466***	-.0002	-.0512***
	(.0149)	(.0258)	(.0156)	(.0069)	(.0103)	(.0072)	(.0118)	(.0097)	(.0121)	(.0168)	(.0148)	(.0175)
Male	-.0081	.0333*	-.0118	.	.	.	.0295**	.0011	.0317**	-.0187	.0042	-.0212
	(.0157)	(.0192)	(.0168)	(.)	(.)	(.)	(.0148)	(.0082)	(.0161)	(.0157)	(.0123)	(.0158)
Other Lang.	-.0136	.0238	-.0173	-.0279	.005	-.0308	-.0069	-.0058	-.0065	-.0521	-.0013	-.0583
	(.0296)	(.0217)	(.0326)	(.0185)	(.0151)	(.0205)	(.0182)	(.0095)	(.0198)	(.0352)	(.0133)	(.0425)
Spanish	.0008	-.0142**	.0025	-.005	-.0376***	-.0029	.0108	-.0025	.012	-.0051	-.0329**	-.0037
	(.0105)	(.0061)	(.0105)	(.0082)	(.0127)	(.0088)	(.0128)	(.0144)	(.0141)	(.0173)	(.0153)	(.0178)

Notes. This table shows the estimated ATE for the difference in differences model using the estimator proposed in Callaway, Sant'Anna (2020), as described in Section 4. The dependent variables *HS Completion*, *Has Kids*, *Employment*, and *Migration* are dummy variables. The ATE is obtained by averaging the dynamic effects estimated on all post-treatment coefficients. Estimation is carried out using only Not-Yet-Treated provinces as comparison group. Different rows report the estimated ATE for different subgroups of the population. Different columns report estimates for three different treatment definitions for each dependent variable: in the first model (indicated by "All") we define as treated a cohort aged 18 or younger at the time of first opening of *any* college campus; in the second (indicated by "Private") only openings of *private* college campuses are considered; in the third (indicated by "Public") only openings of *public* college campuses are considered. Standard errors (in parentheses) clustered at the treatment level are calculated using the bootstrap procedure developed in Callaway, Sant'Anna (2020) and implemented in the provided software. In square brackets we report p-values adjusted for False Discovery Rate, or sharpened q-values, as proposed by Benjamini et al. (2006) and implemented by Anderson (2012). Stars represent statistical significance of the single hypothesis test. *** p<0.01, ** p<0.05, * p<0.1.

First, we notice that high school completion appears to be positively affected by private openings, even though the estimates are only significant at 10%. Fertility appears to be reduced importantly, with 1.5 percent less women reporting having kids. Once again private colleges seem to have the largest effect: the results for high school completion and fertility are consistent with the effects being mediated by (expectation of) college enrollment as the latter is also affected more by private institutions. Finally, we find positive effects on employment and negative ones on migration, driven by public openings. This is not surprising as Peru has recently undergone a licensing process that led to the closure of one third of universities due to low quality standards, almost all private institutions.

In the next section, we will focus on how preferences and selectivity explain the results we just saw. We will use recent data following students through their high school, application decisions, and enrollment outcomes. This data includes information about important demographics that are unavailable in the census, and also reports instances where students unsuccessfully applied to college. Tooled with estimates about the importance of each channel, we will be able to simulate the impact on enrollment gaps of different policies.

5 Estimating a Dynamic Model of Demand for Education

The analysis in Section 4 shows that (i) enrollment and completion increased as a consequence of the opening of new universities in the proximity of students, and (ii) the increase was concentrated among women and individuals belonging to the ethnic majority. However, it is not possible to identify how much of these findings is due to differences in preferences or differences in the probability of being admitted conditional on applying. Additionally, the creation of new colleges should be evaluated against other policies that affect only one of the two factors, thereby requiring the use of more sophisticated tools to evaluate and compare them.[35,36]

To learn about preferences and their interaction with selective admissions, we build a

[35]Take, for example, the case of affirmative action that influences the probability of admission but not necessarily preferences.

[36]Finally, policymakers might be uncomfortable relying on historical data and would prefer to have estimates based on more recent data. However, due to the 2012 Moratorium and the time needed for outcomes to realize, credibly estimating the model of section **??** is not possible. This provides further motivation for the development of a model that does not rely on variation in the availability of colleges over time.

model of demand for education and college admissions in the Peruvian context. We leverage a new dataset that combines detailed individual information on demographics, high school attainment, college applications and enrollment to estimate separately admission probabilities and preference parameters.[37] This dataset follows the cohort expected to graduate high school in 2018 for the period 2013-2020, as described in Section 3. We then simulate counterfactual policies and describe their effects on enrollment levels and gaps.

Individuals in our model can obtain four types of educational achievement: public university degree (giving utility u_{public}); private university degree ($u_{private}$); high school diploma (u_{hs}); or drop out of high school ($u_{dropout}$). Two characteristics of the process make a standard multinomial logit model inadequate. First, a student whose desired outcome is to attend a (selective) public university ($u_{public} \geq u_j$, $\forall j \in \{private, hs\}$) might not be admitted and have to choose another option. Thanks to information about both applications (successful and not) and enrollment, we know the favorite option of all individuals[38] and the "second-choice" option of those who were not admitted to a public university. This gives us more information than in a setting where only the realized outcome or only the application decision is known. Second, the decision is dynamic, in that a student whose application to a public university has been rejected cannot choose to drop out of high school as he must have completed high school to apply in the first place.[39]

Figure 5 shows a graphical depiction of the model. Individuals make two sequential choices. Students first choose whether to drop out of high school or not, based on the expected utility from graduating (given by the maximum utility of the subsequent choices) and on individual preference shocks for these two options – preference shocks for the second choice are not observed at this time.[40] Then, the students who did not drop out observe

[37]Census data only reports few variables to characterize individuals: the dataset that we use in this section provides us with information on parental educational achievement, place of birth, high school attended and GPA, among other variables that have been shown to be important for educational achievement in other studies. Additionally, not observing application behavior and admission results would require imposing additional, strong assumption on the data.

[38]Here, we assume that a student who applies to a public university would receive the most utility from attending it. If applications are costless, all individuals whose favorite choice is enrollment in a public university would apply, even with very small probabilities of admission.

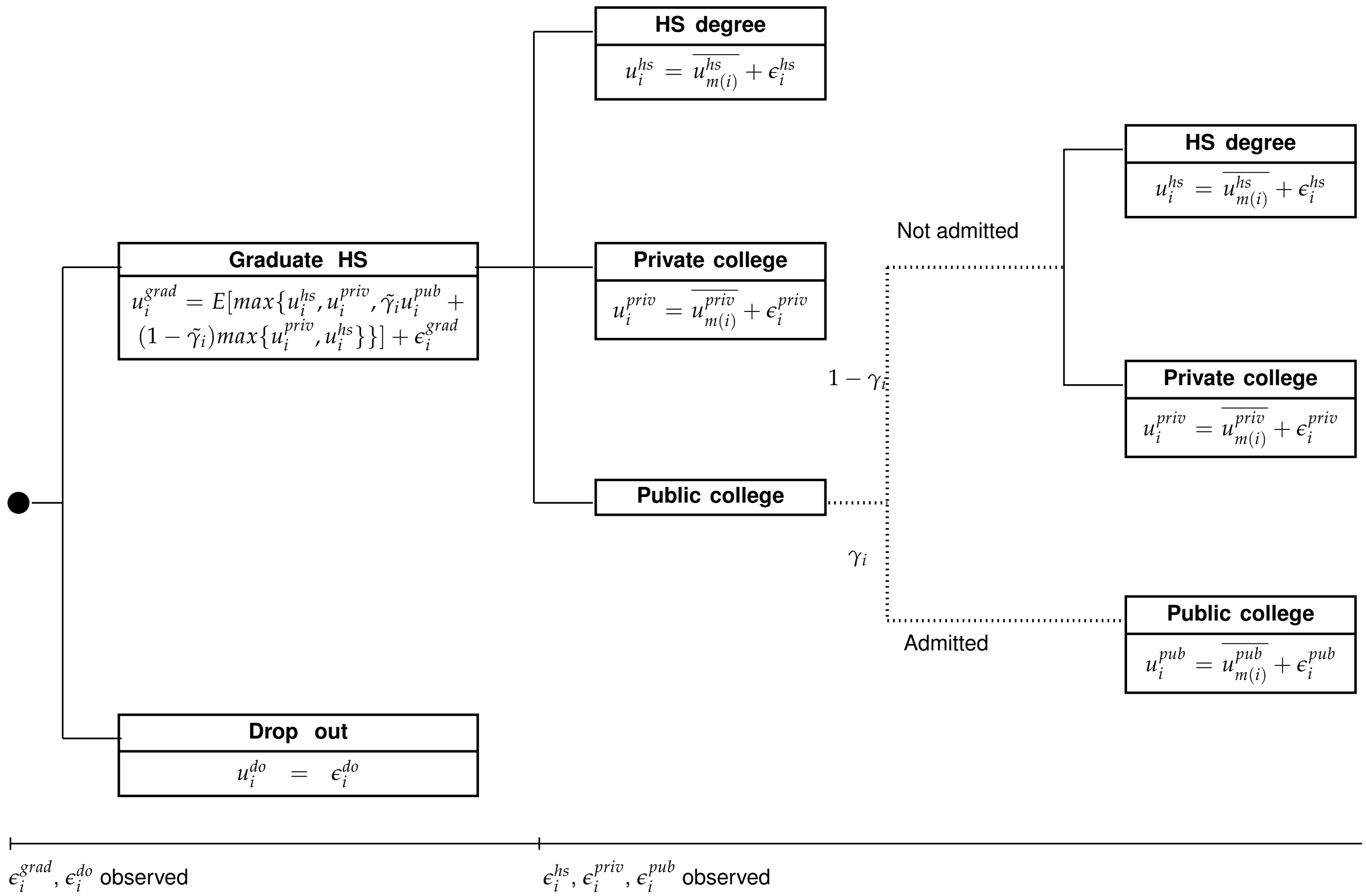
[39]It is important to include high school dropout decisions, as they appear to be affected by college openings for some subgroups of the population. Further, one of our counterfactual policies will involve an improvement in secondary schooling quality, which is likely to affect dropout decisions as well, leading to important welfare consequences that we want to be able to account for.

[40]Calculating expected utility requires that students use their beliefs over the probability of admission to public

preference shocks for all options and decide whether to apply for a public college, enroll in a private one, or not pursue any tertiary education option. If enrollment at a public college represents the preferred option, an application is made that results into admission with probability γ_i. If the applicant is not admitted, she has to choose between the other two remaining options.

colleges, which we indicate as $\tilde{\gamma}_i$.

Figure 5: Model for education demand



It is worth highlighting some important simplifying assumptions inherent to this model. Private universities are assumed to be non-selective. This is correct for the vast majority of private colleges, with few exceptions that are primarily located in Lima. Indeed, according to the administrative dataset SIRIES, 78% of applications to private colleges in 2017 were successful, against 20% for public colleges.[41] We further assume that application to universities is free, but can only happen once: applications fees for the median public college were 228 Soles in 2017 or USD 68 (1% of GDP per capita), according to the 2017 ENUE survey, while the corresponding amounts for private colleges were 120 Soles and USD 37.[42] While applying multiple times is possible, most students only apply once[43], and we calculate probability of admission allowing for multiple attempts. This simplifies the model by reducing the choice space[44] while assigning appropriate probabilities of admission to students.

Finally, notice that individuals do not know their preferences regarding tertiary education until after high school graduation, and the expected utility of graduating from high school is correct on average but noisy. These assumptions appear reasonable given the extensive literature documenting the potential of informational interventions in education and the dispersion of beliefs over returns from college (see, e.g., Hastings et al. (2016)).

5.1 Estimation

Before estimating preferences, we predict the individual probability of admission γ_i. Because we observe the outcome for each individual who applied, we can use it as an outcome in a linear model including individual characteristics. [Eventually, will implement Lasso.] In this case, we can use OLS to estimate the model:

$$\mathbb{1}\{admitted_i\} = X_i\beta + \varepsilon_i \tag{1}$$

We use dummies for the following demographic characteristics as regressors: minority status, gender, parental education, below median and top quartile high school GPA, and above

[41] Also consider that non-admitted students can re-apply as many times as desired.

[42] This information is calculated using the Ministry of Education's ENEU survey of college students and is not necessarily representative of the median cost of all applicants.

[43] According to the census of college students of 2010 (CENAUN), 43% of students applied to only one university and 28% to two; 73% of college students applied just once to the college they attend (15% applied twice).

[44] Modeling the possibility of applying after being rejected would add another choice to the "Not admitted" node.

median high school quality. We also include dummies for the department where the student attended high school, for the presence of a public college in the province where the high school is located, and whether that high school was in the province of Lima. Finally, we include interactions between high school quality and presence of a public college, and the demographic dimensions where relevant gaps have been shown (ethnicity, gender, and parental education).[45,46] The predicted probability of admission will be called $\hat{\gamma}_i$.

To estimate utility parameters, we use a two-steps approach in the vein of **?**. We group individuals into cells defined by their demographic characteristics.[47] This produces 9408 potential cells, of which 2332 are populated by individuals and have non-zero shares for all options.[48] Then, in the first step, we use Maximum Likelihood Estimation to obtain mean utilities for each option for all demographic cells, $m(i)$. In the second step, we project regressors and instruments on the estimated mean utilities to recover the parameters of interest – tuition and distance. This approach allows us to include information about first and second choice for individuals who apply to public college and are rejected, and allows for parallel estimation of the first step.[49] As in any usual conditional logit model, we normalize the utility of one choice, dropping out of high school, to have zero mean – coefficients will be identified relative to that option. Relatedly, we also set the variance of all errors (scale parameter) to be equal to 1.

Define pairs $\omega \equiv (j, a)$ where $j \in J = \{pub, priv, hs, do\}$ is the realized outcome and $a \in A = \{0, 1\}$ representing whether the student applied to a public university or not.[50] The

[45] As this is a prediction exercise, we keep the specification as simple as possible to avoid overfitting. Results from this specification and several alternatives are reported in the Appendix.

[46] One caveat is that students that apply to college might be selected on some unobserved characteristics – e.g. students with low test-taking abilities might not apply at all. While possible, this is unlikely to affect the estimates given the very high R^2 of the regression and the inclusion of several ability measures.

[47] These are the same as in Model 1, with the difference that we now use the province where the high school was located rather than its department. This also implies that all variables at the province level are redundant (being in Lima and presence of a public college in the province).

[48] We only need non-zero shares for dropout, private college enrollment, no enrollment, and public college application (not necessarily admission).

[49] Since we do not use a Generalized Method of Moments approach, we will not ask the model to match any data moment. As a validation exercise, we will compare the changes in enrollment caused by the opening of new campuses predicted by the model and compare them to the estimates of Section 4. At no point in the model estimation we require that the difference in differences estimates are matched.

[50] Because students can't be admitted to public university if they don't apply, we can omit the indicator for the pair $(pub, 0)$. Similarly, students that drop out of high school cannot apply at any public college, so $\Pr(\omega = (do, 0)|m(i)) = Pr(\omega = (do, 1)|m(i))$.

likelihood function will also depend on admission probability γ_i.[51]

The probability of observing the pair ω for individual i in cell $m(i)$ is:

$$P_i = Pr(\omega = (j_i, a_i)|m(i)) = \begin{cases} (1-\Gamma) & if \quad \omega = (do, .) \\ \Gamma * \gamma * Pr(u_{pub} > u_{priv} \quad \& \quad u_{pub} > u_{hs}) & if \quad \omega = (pub, 1) \\ \Gamma * (1-\gamma) * Pr(u_{pub} > u_{priv} > u_{work}) & if \quad \omega = (priv, 1) \\ \Gamma * (1-\gamma) * Pr(u_{pub} > u_{work} > u_{priv}) & if \quad \omega = (hs, 1) \\ \Gamma * Pr(u_{priv} > u_{pub} \quad \& \quad u_{priv} > u_{hs}) & if \quad \omega = (priv, 0) \\ \Gamma * Pr(u_{hs} > u_{pub} \quad \& \quad u_{hs} > u_{priv}) & if \quad \omega = (hs, 0) \end{cases}$$

where $(1-\Gamma) \equiv Pr(E[max\{u_{hs}, u_{priv}, \tilde{\gamma} u_{pub} + (1-\tilde{\gamma}) max\{u_{hs}, u_{priv}\}\}] < 0)$ and we omit the index i on the right-hand side to reduce clutter.

The likelihood function will then be equal to $\mathcal{L}(X, \theta) = \prod_i P_i$. In the first step, we estimate mean utilities for each option for each group $m(i)$: this is equivalent to estimating the model separately for each group with an intercept for each option (except for the normalization of u_{do}). During this step, beliefs about the probability of admission, $\tilde{\gamma}_{m(i)}$ are also backed out. We provide results using beliefs backed out within the MLE procedure and set equal to the predictions from the LPM described above, $\hat{\gamma}_i$. $\tilde{\gamma}_{m(i)}$ is identified by the parametric assumptions on the expected utility from graduation and by trying to rationalize dropout behavior.[52] Table **??** shows how the recovered beliefs relate with demographic characteristics, comparing them to the observed probabilities of admission. The results show that beliefs relate similarly to demographics as observed probability, except for gender: women appear to be overconfident regarding their chances of admission.

In the second step, we use an IV approach to recover estimates for tuition and distance parameters. This requires estimating a linear regression using the mean utilities estimated in step one as outcome. Because tuition is set endogenously by private universities, we need to use an exogenous instrument in order to recover a consistent estimate. We propose the interaction between distance from the closest public college, the difference in wage between

[51] In the estimation, we will replace γ_i with $\hat{\gamma}_i$ estimated using the model described above.

[52] Intuitively, there are two decisions being made by everyone. The second one is based on the utility from each option. Taking these utilities as fixed, the first choice depends on the beliefs about the probability of admission.

college and high school graduates, and an indicator for private colleges as an instrument. Intuitively, the higher the college premium, the more private colleges might be able to charge; distance from the closest public college proxies competition. When returns are high and competition low private colleges will charge higher tuition. The exogeneity assumption needs to be holding conditional on province fixed effects and local returns to college, among others.

We obtain the following model:

$$\overline{u^j_{m(i)}} = \beta_0 + \beta_1 tui\hat{t}ion_{m(i),j} + \beta_2 returns_{m(i),j} + \beta_3 proximity_{m(i),j}$$
$$+ X_{1,m(i)} + \theta\mathbb{1}\{\text{j = HS degree}\} X_{2,m(i)} + \zeta_{m(i),j}$$

where is the tuition predicted in the first stage regression; proximity is measured either as a continuous variable (distance in km from the closest campus[53]), as a dummy indicating whether a campus is present in the province, or interacted with demographics (gender, ethnicity, SES) to allow for heterogeneous preferences; and returns are calculated relative to dropout wages in the province using the household survey ENAHO. $X_{1,m(i)}$ includes all the available demographics, and $X_{2,m(i)}$ includes an SES dummy, a minority dummy, and a gender dummy: this allows for relative preferences for work versus college to vary with demographics.

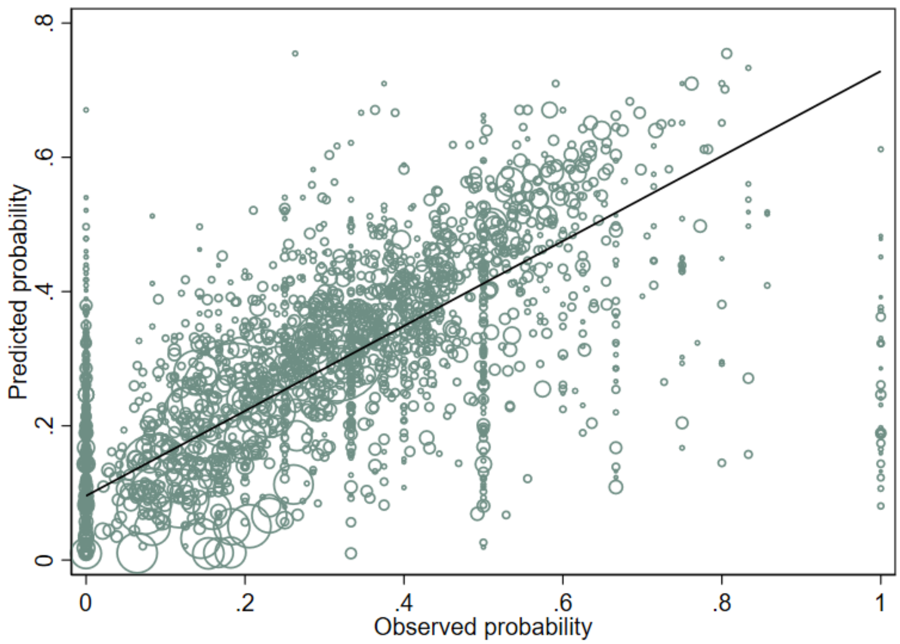
5.2 Estimation Results

Figure 6 shows how the distribution of the predicted probability of admission for each demographic cell, $\hat{\gamma}_i$ aligns with the observed ones, as reflected by the R^2 of 0.73 for the chosen model of column 4 in Table **??** in the Appendix. We can see that some groups appear to have probabilities equal to 0 or to 1 in the data, due to their small size; this issue is solved by the linear probability model. We impose the constraint that predicted probabilities are strictly between 0 and 1, which is binding for 1.1% of the individuals in the sample (the minimum predicted probability is -5.6%). Further discussion about different specifications for the LPM are included in the Appendix.

In Table 5 we report the results from the second step of the estimation. Columns 1 to 3

[53]Distance is assumed to be equal to 0 for the drop out and high school diploma options.

Figure 6: Probability of Admission, Observed and Predicted



Notes. The x-axis of this figure reports the fraction of students that was admitted to a public college among those who applied, for each demographic group. The y-axis represents the predicted probability of admission predicted by Model 1. The size of each marker represents the number of individuals in the demographic group. The reported line represents the linear fit of predicted on observed probability of admission. Information on applications and admissions is contained in the SIRIES dataset.

report results using different definitions of distance/proximity. Tuition is negatively related to utility as predicted by standard demand theory, suggesting that the instrument is effective in removing the endogenous relation between prices and quality or demand shocks.[54] We find that reducing the distance from college by $10km$ is valued on average as much as a 1000 Soles (USD250) discount on yearly tuition. When estimating the effect separately for high and low SES students, we can see that the latter value proximity about twice as much. Wage premiums do not appear to be significantly related to utility. The instrument appears to be relevant, as shown by the Kleibergen-Paap F-Stats ranging from 38 to 40 depending on the specific model. Standard errors are clustered at the province level, similar to Hastings et al. (2017) and do not adjust for the estimated outcome variable.

5.2.1 Model Fit and Comparison with Difference in Differences

In order to assess the fit and validity of the estimated model, we first compare the shares of each choice[55] in the data and the predicted shares from the model. The results for all individuals and for different subsamples are reported in Table 6. We can observe that the model does a good job at predicting choice shares overall, and is able to sensibly reflect heterogeneity of different groups.

As a stronger validation exercise, we build a counterfactual simulation where public college campuses are opened in provinces that do not have any. This is the closest counterfactual to the quasi-experiment in Section 4.[56] Two crucial differences are the time frames (1960-2009 for the difference in differences, and 2017 for the current exercise), and the geographical areas, because provinces currently without any college are necessarily not treated in the quasi-experiment. Figure 7 shows the comparison of the effects estimated in the difference in differences model and the predicted effect estimated under the counterfactual policy. To reflect the characteristics of the data used in the current section, we report the estimates for 12 years old students.

[54] Tuition is measured in thousands of Soles, and ranges between 7 and 9 for private colleges.

[55] We report dropout, enrollment at public or private colleges, or high school completion as outcomes. Not everyone who applies is allowed to subsequently enroll in a public university, as discussed previously.

[56] Removing universities from the provinces involved provides an intuitive alternative to the proposed counterfactual. We do not simulate this counterfactual, as multiple entry of private and public colleges might have occurred after the first treatment, making the subtraction of one college from those provinces a very different treatment than the creation of the first one.

Table 5: Estimated Utility Parameters

	(1) Utility	(2) Utility	(3) Utility	(4) Tuition
Tuition	-0.0657**	-0.0648**	-0.0697**	
	(0.0265)	(0.0267)	(0.0275)	
Wage premium	0.00303	0.00322	0.00283	-0.168***
	(0.00550)	(0.00544)	(0.00543)	(0.0468)
Distance	-0.00721***			
	(0.00160)			
Distance (Low Ed. Parents)		-0.00869***		0.00251
		(0.00164)		(0.0103)
Distance (High Ed. Parents)		-0.00342		0.000853
		(0.00244)		(0.0115)
Proximity (Low Ed. Parents)			0.456***	
			(0.101)	
Proximity (High Ed. Parents)			0.218*	
			(0.129)	
Instrument				0.120***
				(0.0193)
Observations	1377858	1377858	1377858	1377858
Kleibergen-Paap F-stat	38.49	38.44	39.75	
Within R-squared	0.0170	0.0206	0.0138	

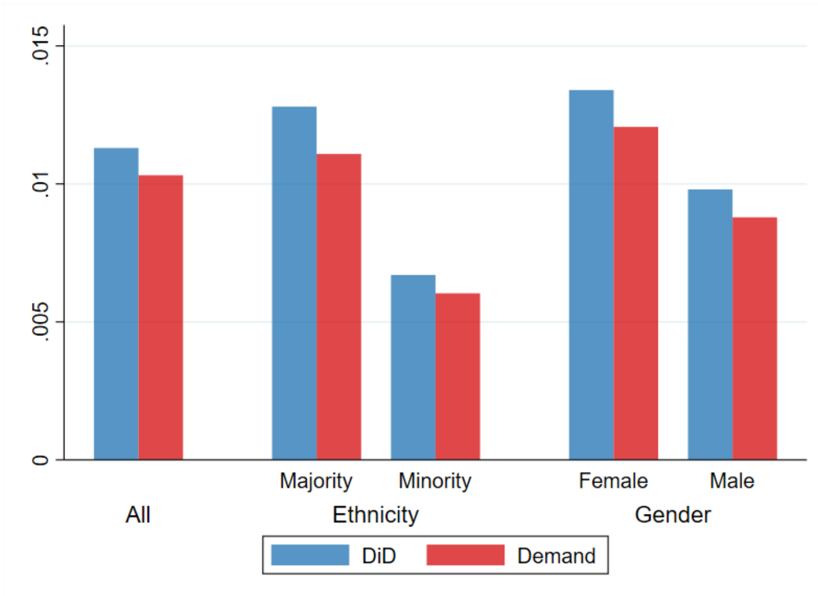
Notes. This table shows the estimated coefficients for model 5.1 using an instrumental variable estimator, as described in Section 5. The dependent variable is the utility for each available option (graduating high school with no further education, attending a public college, attending a private college) estimated in the first step through Maximum Likelihood. All models also include dummies for demographic characteristics (gender, parental education, minority status, above median GPA, lowest quartile GPA, high school quality, province of attended high school). Finally, we include the interaction of parental education, minority status, and gender with high school quality and a dummy for the college options. The instrument used for the estimation is the interaction of a dummy for the private college option, a measure of local college premium, and distance from the closest public college campus. The number of different demographic combinations for which utility estimates (our dependent variable) are available is 2332; 191 different provinces are represented in those groups. Standard errors (in parentheses) are clustered at the province level. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Model Fit of Outcome Shares

		Data	Model
All			
	Dropout	0.173	0.193
	Public Coll. (Adm.)	0.086	0.079
	Private Coll.	0.253	0.205
	High School	0.489	0.523
	N	459286	459286
Female			
	Dropout	0.144	0.158
	Public Coll. (Adm.)	0.088	0.086
	Private Coll.	0.291	0.241
	High School	0.476	0.515
	N	219926	219926
Low GPA			
	Dropout	0.603	0.701
	Public Coll. (Adm.)	0.01	0.015
	Private Coll.	0.078	0.076
	High School	0.309	0.209
	N	88329	88329
Parental Ed.			
	Dropout	0.128	0.151
	Public Coll. (Adm.)	0.111	0.12
	Private Coll.	0.386	0.309
	High School	0.375	0.421
	N	224735	224735

Notes. The first column of this table reports the shares of individuals choosing each available option: dropping out of high school, graduating from high school with no further education, and attending a private or public college. We then compare these shares to those predicted by the estimated model in the second column. We report shares from the data and the model for the whole population studied and for three subgroups – women, students with low GPA, and with parents who completed high school.

Figure 7: Model Validation: Diff in Diff Comparison



Notes. Comparison of the estimates from the difference in difference model estimated in Section 4 and from the counterfactual where currently unattended (i.e. without any college campus) provinces are treated with the opening of a public college campus. Difference in differences estimates are ATE obtained by aggregating the dynamic effects for relative periods 0-18, where treatment is the opening of any campus. The estimates are comparable, and predicted effects from the demand model are also characterized by heterogeneity similar to that found in the difference in differences estimates.

We can see that the predicted effects for the overall population are very similar to the estimates of Section 4. Remarkably, the model predicts heterogeneous effects along the gender and ethnic dimensions that resemble the heterogeneity found previously: as our main focus regards the effects on different groups in the population, being able to capture their different reactions to policy is particularly important. This gives us confidence in (i) using this counterfactual exercise to study the interaction of preferences and selectivity, and (2) comparing the effects that different simulated policies might have on enrollment gaps.

Finally, as will be discussed in the next Section, patterns are overall consistent with economic intuition: as public universities are opened, some individuals shift away from private ones. This pattern was not visible in the previous analysis, as census data does not report separately enrollment and completion at private and public universities.

5.3 Counterfactual exercises

The estimated model allows us to (i) simulate the effects of alternative policies on enrollment, and (2) quantify the role of heterogeneous preferences and selection rules in determining outcomes. In this section, we will study two alternative policies: affirmative action, and investment in basic education.

The first is one of the most "contentious policies that exist in US labor and education markets" (Arcidiacono, Lovenheim (2016)) which is highly targeted but criticized for the potential harm to those it should benefit and the exclusionary effects on non-beneficiaries. We simulate an affirmative action policy where the probability of admission of each individual is $10p.p.$ higher among those identified as belonging to an ethnic minority. We adjust capacity constraints so that non-beneficiaries have the same probability of admission as before the policy (in the Appendix we also provide a simulation where capacity constraints are kept constant at the national level).

The second policy addresses the central point in the meritocracy-equity tension: if all students had access to the same resources before the selection process, we would expect more equal outcomes. We focus on access to high quality secondary education, a problem that has been highlighted by several Peruvian scholars. We simulate an improvement in the quality of instruction for all below median schools, so that their average quality is the same as the average of those above median. This represents a massive investment on secondary education that provides all students with access to a similar quality of secondary education.

In Figure 8 we report the results of the policy simulations, benchmarked against the simulated policy of opening new campuses in unattended provinces.[57] We can see that, as seen in the difference in differences analysis and in Section 5.2.1, opening new college campuses increases the gaps along the ethnicity and gender dimensions. In the second panel of Figure 8 we can see that the strong affirmative action policy we simulate reduces by about one quarter the gap along the targeted dimension. As expected, other dimensions are not affected.

In the last panel, we observe the effects of a large educational investment in secondary schooling: this policy decreases by 87% the ethnicity gap but increases the gender gap by 42% and the SES gap by 8%. This is consistent with ethnic segregation in secondary schooling,

[57]We maintain a consistent sample, focusing on the unattended provinces, as discussed in the notes of Figure 7.

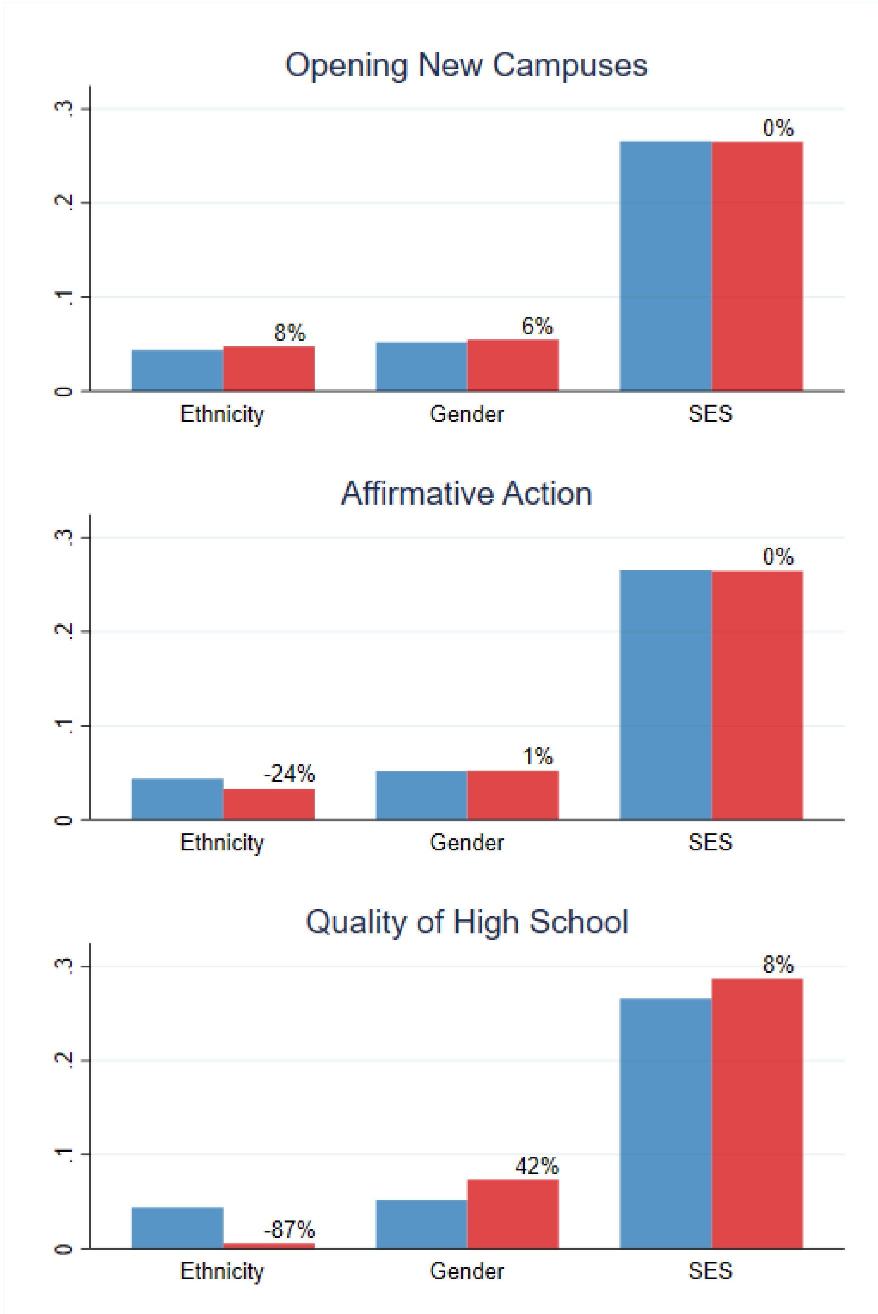
with the data showing that in the relevant provinces only 5% of minority students attend an above median quality high school, compared to 38% of non-minority students. This means that the policy will be fairly well targeted along the ethnic dimension. On the contrary, while children of more educated parents are also more likely to attend high quality high schools, the difference is not as large, 40% versus 30%. It is also worth noting that the model predicts a decrease in dropout rates greater than $1p.p.$ for any group and as large as $3.8p.p.$ for low GPA students.

In Figure 9, we turn on either the selection or the preference channels.[58] We can see that the preference channel tends to reduce gaps, thanks to the fact that disadvantaged students typically benefit more from proximity to college (see discussion in Section 5.2) and attend worse high schools. On the other hand, the selection channel increases all gaps in both shown scenarios, as more advantaged students are more likely to be able to take advantage of opportunities: for example, majority status and high school quality are complementary in increasing the chances of admission. Finally, it is worth highlighting that while high school quality improvements benefit minorities the most due to them attending on average worse schools, the majority group is the one that mostly benefits from proximity to college, especially students with less educated parents.

Our simulations have some important limitations. First, we do not model the tuition setting decision of private colleges: this means that private colleges are not allowed to change tuition in reaction to the implemented policies and changes in demand. The focus on provinces without private college campuses makes it less likely that private colleges in other provinces would adjust tuition sizably. If they did, we would expect them to reduce tuition when public supply is increased, and to increase it when demand increases. From our simulations, we can see that opening new campuses and increasing capacity to accommodate more minority students through affirmative action leads to reductions in private enrollment which should trigger reductions in tuition; conversely, we expect increased tuition following a secondary schooling quality improvement that raises overall demand for education. The direction of the bias coming from holding tuition fixed depends crucially on the sensitivity of various groups to monetary costs.

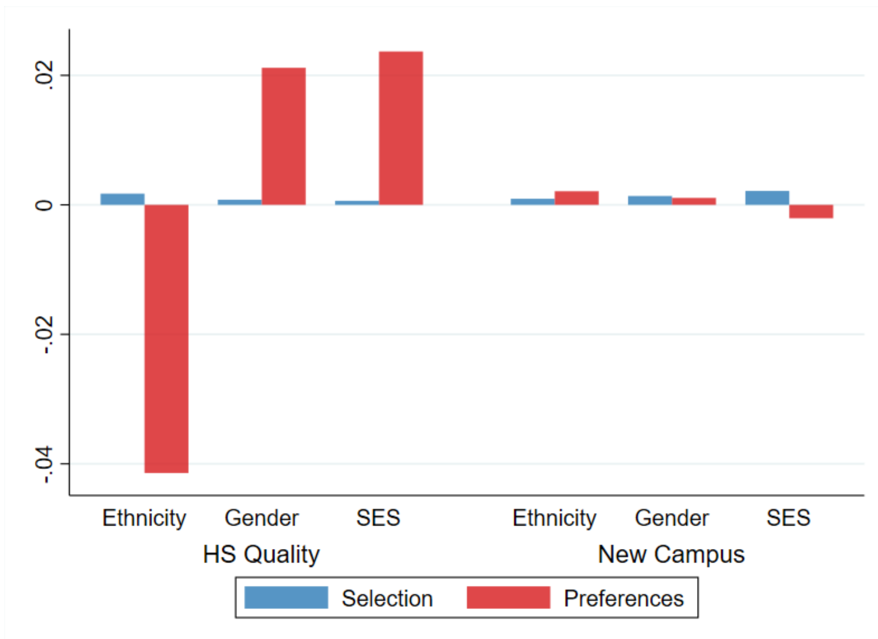
[58] We do not show or discuss the affirmative action counterfactual, because in our model it is assumed to only affect selection.

Figure 8: Predicted Gaps in Enrollment under Counterfactual Policies



Notes. The figure shows predicted gaps in college enrollment by ethnicity, gender, and SES (proxied by parental education), before (blue) and after (red) the implementation of several policies. In the top panel, new college campuses are opened in provinces that do not have any at the moment. In the middle panel, the probability of admission of all minority students is increased by $10p.p.$. In the bottom panel, the average quality of below median quality high schools is increased to match that of above median ones. In all three panels, the predictions are made for the same provinces – those that do not have any college campus at the moment.

Figure 9: Decomposition of Predicted Changes in Enrollment Gaps



Notes. The figure shows the percentage points change in predicted gaps in college enrollment by ethnicity, gender, and SES (proxied by parental education), following the implementation of two policies: on the left, the average quality of below median quality high schools is increased to match that of above median ones; on the right, new college campuses are opened in provinces that do not have any at the moment. In both cases, the predictions are made for the same provinces – those that do not have any college campus at the moment. In blue, we report the change in gap due to changes in the probability in admission, and, in red, the change in gap due to changes in preferences.

Second, as previously discussed, we do not model credit constraints which might influence some conclusions. Table **??** in the Appendix compares how demographics relate to beliefs and real probability of admission: if beliefs were exactly correct and the model perfectly specified, we would expect them to relate in the same way. One factor that might explain part of the difference is the lack of credit constraints in the model. As most coefficients have the same sign and comparable magnitudes, we think that the scope for credit constraints should not be major. Finally, we assume that students' behavior along different dimensions, such as the use of private tutoring (e.g. see Chatterjee et al. (2020) for primary education), does not react to the changing environment. This possibility cannot be easily ruled out.

6 Conclusions

How scarce college seats are allocated has been a debated topic for a while. Similar to how the University of California system grew from 2 to 9 campuses offering college curricula between 1943 and 1965 to accommodate demand (Stadtman (1970)), developing countries have been expanding the supply of higher education at a geographical level and increasing the capacity of existing campuses in the last few decades (e.g., see Ferreyra et al. (2017)). Increasing supply, however, has not made the allocation problem any less salient.

In fact, gaps in access to post-secondary education are still a recognized problem in the US (see, for example, Bailey, Dynarski (2011), Hanushek et al. (2020), and Chetty et al. (2020)) and governments of developing countries like Peru have raised concerns of stratification and made equitable access a primary goal (Ministerio de Educación (2020)).

In this paper, we have studied the effects of meritocratic admission criteria for the allocation of college seats in a context where initial resources, such as high quality secondary education, are unequally distributed. We have shown that in the case of Peru, the expansion of the higher education system through the creation of new campuses has led to an increase in college enrollment and completion but benefited more advantaged groups of the population the most, thereby increasing achievement gaps. Building and estimating a discrete choice model of demand for education, we have studied the relative importance of meritocratic selection and preference heterogeneity of different groups in determining this outcome: the selection process increases gaps along all measured dimensions after college campuses are opened due to complementarities between proximity of educational institutions and demographics of the advantaged groups.

It is important to note that meritocratic selection criteria do not necessarily produce unequal outcomes, just like non-meritocratic ones do not necessarily worsen allocation. In a world where initial resources (like family wealth, networks, educational opportunities) are equally distributed and talent is not prerogative of one group, meritocratic rules might lead to the best allocation *and* equal access to education; because talent is more equally distributed than access in poorer countries (Agarwal, Gaule (2020)) affirmative action policies or other non purely meritocratic selection rules might cause improvements in the allocation of talent. We think that exploring the interaction of allocation rules for scarce goods like public college seats and prior inequality will yield a deeper understanding into the mechanisms that cause income persistence, and potentially help the design of better institutions that foster social mobility and a true equality of opportunities.

References

Agarwal Ruchir, Gaule Patrick. Invisible Geniuses: Could the Knowledge Frontier Advance Faster? // American Economic Review: Insights. 12 2020. 2, 4. 409–24.

Redistributive Allocation Mechanisms *. // . 2020.

Alderman H., Orazem P. F., Paterno E. M. School quality, school cost, and the public/private school of choices of low-income households in Pakistan // Journal of Human Resources. 2001. 36, 2. 304–326.

Anderson Michael L. Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects // https://doi.org/10.1198/016214508000000841. 12 2012. 103, 484. 1481–1495.

Arcidiacono Peter, Kinsler Josh, Ransom Tyler. Legacy and Athlete Preferences at Harvard // https://doi.org/10.1086/713744. 1 2021.

Arcidiacono Peter, Lovenheim Michael. Affirmative action and the quality-fit trade-off. 3 2016. 3–51.

Bailey Martha, Dynarski Susan. Gains and Gaps: Changing Inequality in U.S. College Entry and Completion // National Bureau of Economic Research. 12 2011.

Baker Andrew, Larcker David F., Wang Charles C. Y. How Much Should We Trust Staggered Difference-In-Differences Estimates? // SSRN Electronic Journal. 3 2021.

Barro Robert J. Economic Growth in a Cross Section of Countries // The Quarterly Journal of Economics. 5 1991. 106, 2. 407–443.

Bautista Maria Angelica, González Felipe, Martinez Luis R., Munoz Pablo, Prem Mounu. Dictatorship, Higher Education and Social Mobility // SSRN Electronic Journal. 8 2020.

Bedard K. Human capital versus signaling models: University access and high school dropouts // Journal of Political Economy. 7 2001. 109, 4. 749–775.

Belskaya Volha, Peter Klara Sabirianova, Posso Christian M. Heterogeneity in the effect of college expansion policy on wages: Evidence from the Russian labor market // Journal of Human Capital. 3 2020. 14, 1. 84–121.

Benjamini Yoav, Krieger Abba M., Yekutieli Daniel. Adaptive linear step-up procedures that control the false discovery rate // Biometrika. 9 2006. 93, 3. 491–507.

Beuermann Diether, Jackson C. Kirabo, Navarro-Sola Laia, Pardo Francisco. What is a Good School, and Can Parents Tell? Evidence on the Multidimensionality of School Output // National Bureau of Economic Research Working Paper Series. 12 2018.

Black Sandra, Denning Jeffrey, Rothstein Jesse. Winners and Losers? The Effect of Gaining and Losing Access to Selective Colleges on Education and Labor Market Outcomes. Cambridge, MA, 3 2020.

Bleemer Zachary, Brady Henry, Card David, Cummins Joseph, Delong Brad, Goodman Josh, Kline Pat, Moretti Enrico, Olney Martha, Rothstein Jesse, Steier Zoë, Walters Chris. Top Percent Policies and the Return to Postsecondary Selectivity * // SSRN. 2020. December.

Borusyak Kirill, Jaravel Xavier, Spiess Jann. Revisiting Event Study Designs: Robust and Efficienct Estimation // Work in Progress. 2021. 1–48.

Callaway Brantly, Sant'Anna Pedro H.C. Difference-in-Differences with multiple time periods // Journal of Econometrics. 12 2020.

Card David. Using Geographic Variation in College Proximity to Estimate the Return to Schooling. Cambridge, MA, 10 1993.

Carneiro Pedro, Das Jishnu, Reis Hugo. The Value of Private Schools: Evidence from Pakistan // Iza 9960. 2016. 9960, 9960.

Carrell Scott, Sacerdote Bruce. Why do college-going interventions work? // American Economic Journal: Applied Economics. 2017. 9, 3. 124–151.

Chatterjee Chirantan, Hanushek Eric, Mahendiran Shreekanth. Can Greater Access to Education Be Inequitable? New Evidence from India's Right to Education Act. Cambridge, MA, 6 2020.

Chetty Raj, Friedman John N., Saez Emmanuel, Turner Nicholas, Yagan Danny. Income Segregation and Intergenerational Mobility across Colleges in the United States // Quarterly Journal of Economics. 8 2020. 135, 3. 1567–1633.

Cuenca Ricardo. La educación universitaria en el Perú : democracia, expansión y desigualdades // Instituto de Estudios Peruanos. 2015. 4, 1850. 53–74.

Dinerstein Michael, Smith Troy. Quantifying the Supply Response of Private Schools to Public Policies. 2014.

Duflo Esther, Dupas Pascaline, Kremer Michael. Education, HIV, and Early Fertility: Experimental Evidence from Kenya // American Economic Review. 9 2015. 105, 9. 2757–97.

Eisenhauer Philipp, Heckman James J., Mosso Stefano. Estimation of dynamic discrete choice models by maximum likelihood and the simulated method of moments // International Economic Review. 5 2015. 56, 2. 331–357.

Ferreyra Maria Marta, Avitabile Ciro, Botero Álvarez Javier, Haimovich Paz Francisco, Urzúa Sergio. At a Crossroads: Higher Education in Latin America and the Caribbean. 5 2017.

Fuller Winship C., Manski Charles F., Wise David A. New Evidence on the Economic Determinants of Postsecondary Schooling Choices // The Journal of Human Resources. 23 1982. 17, 4. 477.

Gelber Alexander, Isen Adam. Children's schooling and parents' behavior: Evidence from the Head Start Impact Study // Journal of Public Economics. 5 2013. 101, 1. 25–38.

Gennaioli Nicola, La Porta Rafael, Silanes Florencio Lopez-de, Shleifer Andrei. Human Capital and Regional Development // The Quarterly Journal of Economics. 2 2013. 128, 1. 105–164.

Goodman-Bacon Andrew. Difference-in-differences with variation in treatment timing // Journal of Econometrics. 6 2021.

Hanushek Eric, Peterson Paul, Talpey Laura, Woessmann Ludger. Long-run Trends in the U.S. SES-Achievement Gap. Cambridge, MA, 2 2020.

Hastings Justine, Hortaçsu Ali, Syverson Chad. Sales Force and Competition in Financial Product Markets: The Case of Mexico's Social Security Privatization // Econometrica. 11 2017. 85, 6. 1723–1761.

Hastings Justine S., Neilson Christopher A., Ramirez Anely, Zimmerman Seth D. (Un)informed college and major choice: Evidence from linked survey and administrative data // Economics of Education Review. 4 2016. 51. 136–151.

Heckman James J., Mosso Stefano. The Economics of Human Development and Social Mobility // http://dx.doi.org/10.1146/annurev-economics-080213-040753. 8 2014. 6. 689–733.

Hsieh Chang-Tai, Hurst Erik, Jones Charles I., Klenow Peter J. The Allocation of Talent and U.S. Economic Growth // Econometrica. 9 2019. 87, 5. 1439–1474.

Jacob Brian, McCall Brian, Stange Kevin. College as Country Club: Do Colleges Cater to Students' Preferences for Consumption? // https://doi.org/10.1086/694654. 2 2018. 36, 2. 309–348.

Heterogeneous Beliefs and School Choice Mechanisms. // . 2017.

Kyui Natalia. Expansion of higher education, employment and wages: Evidence from the Russian Transition // Labour Economics. 4 2016. 39. 68–87.

Mello Ursula. Centralized Admissions, Affirmative Action and Access of Low-income Students to Higher Education // American Economic Journal: Economic Policy. ????

Ministerio de Educación . Política Nacional de Educación Superior y Técnico-Productiva. 2020. 92.

Mokyr Joel. Long-Term Economic Growth and the History of Technology // Handbook of Economic Growth. 1 2005. 1, SUPPL. PART B. 1113–1180.

Neilson Christopher. Targeted Vouchers, Competition Among Schools, and the Academic Achievement of Poor Students // Job Market Paper. 2013.

OECD . Avanzando hacia una mejor educacion para Perú. 3. 2016. 36.

Osili Una Okonkwo, Long Bridget Terry. Does female schooling reduce fertility? Evidence from Nigeria // Journal of Development Economics. 8 2008. 87, 1. 57–75.

Pop-Eleches Cristian, Urquiola Miguel. Going to a Better School: Effects and Behavioral Responses // American Economic Review. 6 2013. 103, 4. 1289–1324.

Russell Lauren, Yu Lei, Andrews Michael J. Historical Happenstance and Local Educational Attainment: Evidence from the Establishment of U.S. Colleges. 2021.

SUNEDU . II Informe Bienal sobre la Realidad Universitaria en el Perú | Gobierno del Perú. 2020.

Sánchez Alan, Favara Marta, Porter Catherine. Stratification of returns to higher education in Peru: the role of education quality and major choices Stratification of returns to higher education in Peru: the role of education quality and major choices 1. 2021.

Sant'Anna Pedro H.C., Zhao Jun. Doubly robust difference-in-differences estimators // Journal of Econometrics. 11 2020. 219, 1. 101–122.

Squicciarini Mara P., Voigtländer Nico. Human Capital and Industrialization: Evidence from the Age of Enlightenment // The Quarterly Journal of Economics. 11 2015. 130, 4. 1825–1883.

Stadtman Verne A. The University of California, 1868–1968. New York: McGraw-Hill, 1970. 355–358.

Sun Liyang, Abraham Sarah. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects // Journal of Econometrics. 12 2020.

[APPENDIX AVAILABLE UPON REQUEST]