Outline of Empirical Designs

October 2025

Hastings, Neilson, Zimmerman (2015, unpublished) document a market failure in higher education in Chile: Low-income students overestimate the returns to a degree from nonselective universities. When a less selective school offers vacancies, low-income students will attend them, but they won't get a significant earnings boost to compensate for the foregone years of labor market experience and tuition cost.

There is a second market failure that seems to have been underexplored, and this is where we aim to contribute. Schools do not internalize the public benefit to opening up a vacancy. This probably explains why high quality schools remain small and selective. In this case, even without the belief biases documented by Hastings, Neilson, and Zimmerman, expansions at nonselective colleges don't create as much of a benefit to the economy as expansions at Harvard (or Catolica). Harvard underprovides vacancies, perhaps because they are concerned about their reputation, while nonselective colleges overprovide vacancies. Students attend Boca Raton community college even if they would benefit more by attending Harvard.

1 Empirical Designs

To assess whether this is in fact a market failure, we need to compare how the effect of increasing access to a high quality school for the marginal student compares with the effect of that expansion on inframarginal students at the same school. A regression discontinuity design could identify the effects for marginal students if the marginal student attends a higher-quality school as a result of crossing the threshold. To estimate effects for inframarginal students we would need a separate design. Several possible candidates are listed below in section 1.2.

1.1 Effects of university quality for marginal students:

Let i denote the individual and p denote the program attended by individual i. Assuming a measure of quality, $q_{i,p}$, we would estimate the following RDD

$$
\begin{aligned}
y_{i,p} = \beta_0 &+ \beta_1 \,\mathrm{Admit}_{i,p} + \beta_2 X_{i,p} + \beta_3 \,\Delta q_{i,p} \\
&+ \beta_4 \, X_{i,p} \times \Delta q_{i,p} + \beta_5 \,\mathrm{Admit}_{i,p} \times \Delta q_{i,p} + \varepsilon_{i,p}\,.
\end{aligned} \tag{1}
$$

where $X_{i,p}$ is the weighted PSU score (running variable) determining admission to program p and $\Delta q_{i,p}$ is a measure of the quality difference between program p and the next-best feasible option for individual i.[1] β_1 captures the effect of threshold-crossing when the quality difference around the threshold is 0. We could expect β_1 to equal zero or to be positive. A positive value for β_1 would suggest that applicants possess some knowledge about their idiosyncratic fit at each institution at the time they submit their ROLs. This would imply that, even if there isn't a quality difference at the threshold, individuals rank programs higher where they have an idiosyncratically better fit. AD: what would be $y_{i,p}$? i.e, starting salary?

β_3 and β_4 capture the level and slope (with respect to X) of selection bias: Do people who apply to schools with larger quality differences in their ROLs differ from students who choose smaller quality differences? β_5 would capture how the effect of threshold-crossing differs depending on the quality-difference between the institutions. This is the parameter of interest: if $\beta_5 > 0$, then the marginal student benefits more from attending a high quality institution than a low quality institution.

We discussed several possible measures of quality, including (i) expenditure per student, (ii) avg PSU score of enrolled students on the mandatory portions of the PSU, (iii) a binary indicator for whether program p is in the top 33 (traditional) universities. It would also be possible to estimate measures of program value-added (on wages or employment) and use these as measures of quality.

There was some question of whether we should run these regressions only for individuals whose top K options are in the same field, to avoid mixing quality differences with cross-field comparisons. According to Andres, there are two types of applicants: those who list many programs within a school first before listing another school, and those who fix the program and list many schools consecutively before listing a second school. A first attempt would be to estimate effects of threshold-crossing only for the second group, fixing the field of study, as the effect of attending a higher quality institution might be different for engineering that for political science. This might, of course, run into the limitation of not having enough power, but it's worth trying at least.

1.2 Estimating effects for inframarginal students:

Marginal students are those who change program, and thus program quality, as a result of an expansion of program capacity, which we measure by an increase

[1] Note that this is only the simplest possible RD that could be estimated. It is possible to allow for separate polynomials in the running variable on the right and left hand side of the cutoff.

in vacancies. Inframarginal students are instead those whose program does not change because of the increase in program capacity. They, however, may still be affected by the increase in program capacity because of congestion effects or because of changes in peer effects arising due to the inclusion of the marginal students.

The following several designs seek to estimate effects of program size on inframarginal students and how these effects vary by the quality of the institution. The key aspect in all of these designs is the choice of the subsample on which to run the regressions. We want to identify students who do not change institution because of a change in vacancies and estimate average effects for them only. We list several empirical designs below, with an understanding that some are better-suited to the Chilean higher education context.

OLS Regressions to estimate effects for inframarginal students:

We would like to estimate the following regression at the cohort-program level:

$$y_{c,p} = \alpha_0 + \alpha_1 Z_{c,p} + X'_{c,p}\alpha_3 + \phi_{c,p} + \varepsilon_{c,p} \ , \tag{2}$$

where $Z_{c,p}$ is the number of vacancies for cohort c and program p and $y_{c,p}$ is the average starting salary for graduates of cohort c and program p. This regression suffers from a significant problem: There is perfect multicollinearity between the cohort-program fixed effects $\phi_{c,p}$ and $(Z_{c,p}, X_{c,p})$. However, it is important to include these fixed effects, because programs that are large and have high $Z_{c,p}$ may also have high earnings. If we assume that $\phi_{c,p}$ evolves linearly over time for each program, so that $\phi_{c,p} - \phi_{c-1,p} = \phi_p$, then we can estimate the model in first differences, while still controlling for the potential threat between program size and graduates' earnings. This specification is

$$\Delta y_{c,p} = \phi_p + \alpha_1 \Delta Z_{c,p} + \Delta X'_{c,p}\alpha_3 + \Delta\varepsilon_{c,p} \ . \tag{3}$$

The identifying assumption for 3 to provide a consistent estimate of α_1 is that programs do not increase vacancies in a way that is correlated with potential graduates' earnings six years later.

There are two ways to come up with a sample of inframarginal students. Option 1 uses all students who are above the cutoff for program p for cohorts c and $c-1$. This produces a sample of students who would be admitted to the program given either the amount of vacancies $Z_{c,p}$ or $Z_{c-1,p}$. The problem with this approach is that the range for admitted PSU scores changes across cohorts, so if there are heterogeneous effects on inframarginal students, we may combine some of the causal effect of $Z_{c,p}$ on $y_{c,p}$ with compositional differences in $y_{c,p}$ across cohorts. Option 2 uses students who are above the maximum cutoff for program p across all cohorts in our sample. This will reduce the sample size but ensure that all cohorts of inframarginal students have PSU scores in the same range.

We want as much variation as we can get in $Z_{c,p}$, so we may want to use finer groupings of majors than the 6 we have used so far (STEM, Law, Education, Health, Business, Social Science).

Like the RDD design, which estimates how threshold-crossing effects vary by the quality of the institution, we may also be interested in how the effects of increased capacity on inframarginal wages varies by the quality of the institution. To do this, we can modify equation (3) as follows:

$$\Delta y_{c,p} = \phi_p + \alpha_1 \Delta Z_{c,p} + \alpha_2 \Delta Q_{c,p} + \alpha_3 \Delta Z_{c,p} Q_{c,p} + \Delta X'_{c,p}\beta + \Delta\varepsilon_{c,p} \ . \tag{4}$$

In (4), α_1 captures a level effect of vacancies on the wages of inframarginal students, while α_3 captures the slope of how this effect varies with the quality of the program. Note that effects of vacancy increases and reductions need not be symmetric. A vacancy increase may run into physical constraints (crowded classrooms), while a vacancy decrease will not. It may be therefore useful to add a separate regressor to (4) that indicates whether the vacancy increase is positive, $1(\Delta Z_{c,p} > 0)$.

Two Stage Least Squares Regressions:

Equation (3) estimates reduced form effects of vacancies on wages for inframarginal students. We may also want to estimate the effect of *program size* on wages for these students. This could be define through estimating the following two-equation model in first differences:

$$\begin{aligned} \Delta y_{c,p} &= \phi_p + \alpha_1 \Delta Size_{c,p} + \Delta X'_{c,p}\beta + \Delta\varepsilon_{c,p} \ , \\ \Delta Size_{c,p} &= \lambda_p + \delta_1 \Delta Z_{c,p} + \Delta X'_{c,p}\delta_2 + \Delta\nu_{c,p} \ , \end{aligned} \tag{5}$$

where $Size_{c,p}$ is the number of students in program p for cohort c. For programs that are capacity-constrained, we would expect that $\delta_1 = 1$. Some programs are not capacity-constrained, so we wouldn't expect a change in vacancies to affect enrollment for them, hence we'd expect to estimate $\widehat{\delta}_1 < 1$. α_1 has an interpretation as a Local Average Treatment Effect (LATE). It is the effect of program size on the wages of inframarginal students for the set of programs that expand size because of a change in vacancies. Unlike some LATEs, this is a policy-relevant parameter. A policymaker is more interested in the effects of expanding vacancies at capacity-constrained programs (the most selective ones) than at programs that are not constrained.

If we have enough power, we could estimate the two-equation model in (5) separately by major to allow for the possibility that the effect of program size on inframarginal students differs by major.

AD: This is the same as using vacancies to instrument size, or is there something different to use as an instrument?

Individual-level regression:

It would be possible to estimate the following regression by including lots of dummy variables:

$$y_{i,c,p} = \alpha_0 + \alpha_1 Z_{c,p} + X'_{c,p}\alpha_2 + \phi_c + \phi_p + \varepsilon_{i,c,p} \ , \tag{6}$$

where $y_{i,c,p}$ is the starting wage of individual i who entered school as part of cohort c with major p. The dataset for this is a repeated cross-section, not a panel. That's why we couldn't do a first-difference estimation. Identification of α_1 would come off of longitudinal variation in $Z_{c,p}$ within each particular major, p, and how that correlates with changes in the wages of individuals who enrolled with that cohort in that major.

DID regressions to estimate effects for inframarginal students:

In a difference in differences design, the ideal control group would have no change in peer composition and no change in vacancies. This is not entirely realistic, but it's possible that some programs are never at capacity, so increases in vacancies should not have any effect on enrollment. However, if the competitors of these programs increase capacity, they may still face a change in peer composition. The way to implement a DID would be to compare programs that have a change in vacancies with programs that do not experience any change in vacancies or peer composition. The next paragraph outlines how to implement this.

There may be a way to use student application data to identify *competing programs*. To find the competitors for program p, identify all the programs that are ranked 1 above or 1 below p on each students rank order list. The top K (say 5) programs constitute p's competitors. An alternative would be to say that program m is a competitor of program p if over 5% of students that rank program p list program m directly above or directly below it. Then, once we have a list of competitors for every program, we can define three groups:

1. Control: Program p is a control if
   - It is never at capacity during the sample period
   - p's competitors are never at capacity during the sample period
2. Hybrid: Program p is a hybrid if
   - It is never at capacity during the sample period
   - At least one of p's competitors are at capacity during the sample period
3. Treated: Program p is treated if
   - It is at capacity at some point during the sample period and experiences an increase in vacancies while it is at capacity

We then estimate the following difference-in-differences specification:

$$y_{c,p} = \beta_0 + \beta_1 Hybrid_p + \beta_2 Treat_p + \beta_3 Z_{c.p} + \beta_4 Treat_p \times Z_{c.p} + \epsilon_{c,p} \ , \quad (7)$$

where $Z_{c,p}$ equals 1 if program p is at capacity in cohort $c-1$ and there is an increase in vacancies between cohorts $c-1$ and c. $Z_{c.p}$ is therefore a cohort-varying version of the traditional *Post* variable you'd find in a DiD.

We may want to define treatment as being at capacity at some point during the sample period and experiencing an increase of $> 10\%$ while it is at capacity. Then any program at capacity which doesn't experience a large enough change in vacancies would fall into the hybrid or control groups, depending on its competitors.

If we can come up with the three groups with the data we currently have available, then it would be possible compute means of various variables by control-hybrid-treatment status (balance table) before we get access to individual data. As a first pass, we should compute the fraction of bachelors programs that are never at capacity during the sample period. If this fraction is very low, then we have a tiny control group, and this approach is not feasible.

2 Data (AI mostly)

The applications database is a harmonized extract of the Demre PSU centralized admissions system (Formulario C), containing each applicant's ranked set of program choices and the corresponding admission outcomes for a given application process. The unit of observation is an application preference: an applicant identifier observed in an admissions cycle and a specific ranked choice linked to a program code . In addition to program identifiers, the raw records provide descriptive strings for the option applied to—career/program name, campus/sede, and institution shorthand.

A key feature of this database is that it records applicants outcomes within the centralized assignment mechanism. Each preference includes an admission status, which allows constructing indicators for being admitted, placed on a waiting list, or not admitted at each rank, as well as summary measures across the preference portfolio (e.g., number of submitted preferences, whether the applicant was admitted to any option, the rank of the best admitted option, and whether the applicant is observed on the waiting list for a given choice). The database also contains the preference-level application score reported by Demre. From the admitted records, it is possible to compute program-by-year thresholds, defined as the minimum admitted score within each cell; this cutoff can be used to create a score distance/running variable relative to the threshold and an indicator for being above the cutoff.

Finally, the applications database supports constructing a consistent analysis score for each applicant–preference even when the reported puntaje is missing or zero. By combining (i) applicant-level PSU component scores (e.g., NEM, ranking, language, math, and history/science) and (ii) program-by-year weighting rules, one can reconstruct a weighted application score.

The enrollment data come from administrative matriculation records from MINEDUC's Higher Education enrollment system, complemented (when needed) with Demre enrollment information. These records allow us to observe whether and where an applicant enrolls in post-secondary education in the year of the admissions process. The enrollment file also includes additional descriptors such as broad field classifications and campus information, as well as tuition/fees.

Graduation outcomes are incorporated by merging applicants to administrative degree-completion records at the individual level. These records provide whether the individual graduates and a degree completion year, along with identifiers describing the graduating institution and program and a broad field classification for the completed degree. Because the applications data identify a "target" option (the program at which the applicant is placed/admitted) and a "next-best" alternative, the graduation records can also be used to construct measures of "alignment" between the completed degree and the assigned option—e.g., whether the completed degree matches the target program/institution string identifiers.

The PSU data come from the standardized college-entry exam system used in Chile's centralized university admissions process. These records provide the underlying test scores and pre-admissions academic measures that are used to construct applicants' eligibility and weighted application scores for specific programs. The unit of observation is an individual–process year. The dataset includes subject test scores for the current exam sitting, as well as high-school-based components that enter admissions formulas such as the NEM score and ranking.

We recover earnings from two databases. The first one corresponds to UI records, capturing earnings of formal jobs with permanent contracts outside government positions. The second source corresponds to social security records, which, in addition to the former earnings sources it recovers non-permanent jobs earnings, including government positions. We combine the two earnings sources at the individual–year level to maximize coverage while avoiding double counting. First, we build an annual earnings series from UI (AFC) records and a parallel annual earnings series from social security (AFP) records (the latter is rescaled to approximate total wages, since AFP contributions are reported as a fraction of earnings). Because the two sources overlap for standard formal jobs but each can miss relevant employment spells (UI misses some non-permanent and government jobs, while AFP captures those in addition to the standard formal sector), we do not sum them. Instead, for each person and year we define "total earnings" as the maximum of the UI-based and AFP-based measures for that year: if both are present, we keep the larger value (interpreting the smaller one as under-coverage or partial reporting); if only one source is available, we use that one. This produces a single earnings measure that preserves comparability over time and expands coverage once AFP data become available.

3 RDD estimates: effects for marginal students

Sample selection. Our baseline RD sample is built from centralized PSU admissions applications, which record each student's ranked order list (ROL) of programs (major–institution combinations) in a given admissions cycle. Using the full ROL, we construct an applicant–program–year dataset that includes all listed programs, regardless of their position in the ROL, and we focus on program-year cutoffs generated by the regular centralized assignment. For each program and year, we define the admissions cutoff as the lowest application score among applicants admitted to that program in that year (the "regular cutoff"). We then assign each applicant a comparable application score for each listed program: when the Demre-reported application score is available we use it, and when it is missing/zero we reconstruct it using PSU component scores and the program-specific weights used by the centralized system. The RD running variable is the applicant's score relative to the program-year cutoff, and treatment is crossing the cutoff (an indicator for scoring weakly above the cutoff).

To ensure the design is implemented around binding capacity constraints, we restrict attention to program-year cells that are oversubscribed. Operationally, a program-year is classified as oversubscribed if it has a positive waiting list (i.e., at least one applicant assigned waiting-list status for that program in that year). Finally, we limit the estimation sample to observations with non-missing scores and cutoffs and within a symmetric neighborhood of the cutoff. For our preliminary results we use a bandwidth of 25 points around the cutoff, which is close to the bandwidths typically selected by standard data-driven procedures in related setting

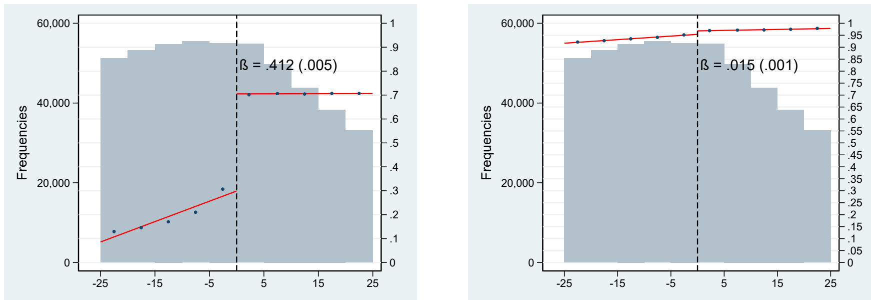
Estimation details. We estimate the following linear regression:

$$y_{i,m,c,t} = \beta D_{i,m,c,t} + \alpha_1 r_{i,m,c,t} + \alpha_2 r_{i,m,c,t} \times D_{i,m,c,t} + \mu_{m,c,t} + \varepsilon_{i,m,c,t},$$

where $y_{i,m,c,t}$ denotes the outcome for individual i who applied to major m at college c in year t. Outcomes include immediate enrollment (in any program and in the target program (m, c)), graduation (from any program and from the target program), and labor-market earnings measured at various horizons after college application. $D_{i,m,c,t}$ is an indicator equal to one if the applicant's weighted admission score for program (m, c) exceeds the year-specific cutoff. The running variable, $r_{i,m,c,t}$, is defined as the difference between the weighted score and the cutoff. The term $\mu_{m,c,t}$ captures major-college-year fixed effects. We estimate this equation by OLS, restricting the sample to observations with $|r_{i,m,c,t}| < 25$. Standard errors are clustered at the (m, c, t) level.

Overall threshold-crossing effects. Figure 1 presents the effects of program eligibility on enrollment outcomes. Panel (a) reports the effect on enrollment in the target program. The outcome variable is an indicator equal to one if the applicant appears in the enrollment database as enrolled in the same program to which they applied, within the same admission round. Eligibility for the target program increases the probability of enrollment in that program by 41 percentage points. Panel (b) shows the effect on enrollment in any bachelor's degree program. The estimated effect is 1.5 percentage points, indicating that being marginally ineligible for a given program does not meaningfully reduce the probability of enrolling in alternative programs.

Figure 1: Effects of program eligibility on higher education enrollment

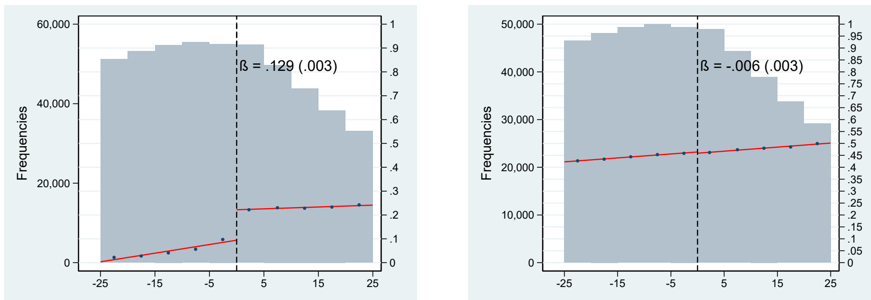


(a) Enrollment in target program (b) Enrollment in any bachelor program

Notes: The figure shows the effect of program eligibility on college enrollment. Panel (a) documents effects on enrollment in the target degree. Panel (b) depicts impact on enrollment in any bachelor degree. For both panels, we restrict the sample to applications within a 25 PSU-score window away from the cutoff. We control for year-program fixed effects. We show the corresponding RD point estimate and its standard error clustered at the year-program level.

Figure 2 presents the effects of program eligibility on graduation outcomes. Graduation is defined as showing up in the administrative graduation records at any point from the college entry examination up to eight years thereafter. Panel (a) reports the effect on graduation from the target program, while panel (b) examines graduation from any bachelor's degree program. We find large effects on graduation from the target program (13 percentage points) and statistically insignificant effects on graduation from any bachelor's program.

Figure 2: Effects of higher education program eligibility on graduation

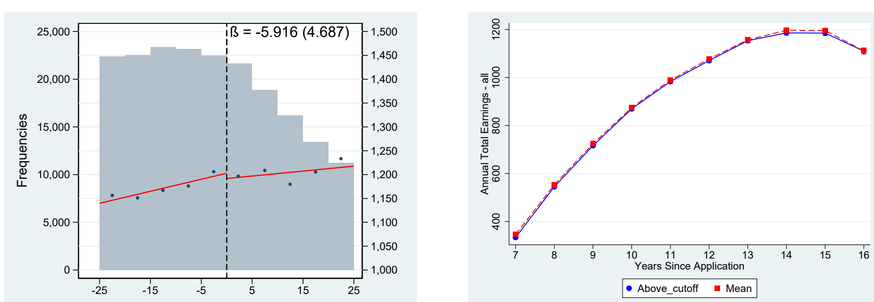


(a) Graduation in target program (b) Graduation in any bachelor program

Notes: The figure shows the effect of program eligibility on college graduation. The outcome variable equals 1 if the individual shows up in the graduation administrative records for up to eight years after college application in the centralized admission system. Panel (a) documents effects on graduation in the target degree. Panel (b) depicts impact on graduation in any bachelor degree. For both panels, we restrict the sample to applications within a 25 PSU-score window away from the cutoff. We control for year-program fixed effects. We show the corresponding RD point estimate and its standard error clustered at the year-program level.

We now examine effects on earnings. Figure 3 shows these results. Panel (a) presents the regression discontinuity estimate of average monthly earnings over the 15 years following college application. Panel (b) shows baseline mean earnings and estimated treatment effects for the period spanning 7 to 15 years after application. Across both panels, we find no evidence of earnings effects associated with marginal eligibility for a higher-education program.

Figure 3: Effects of higher education program eligibility on earnings



(a) Effect on earnings 15 years after application (b) Dynamic threshold-crossing effects

Notes: this figure documents threshold-crossing effects on earnings. The outcome variable is annual earnings. Panel (a) depicts average earnings, 15 years after application, as a function of PSU scores for marginally eligible and not eligible to a program. Panel (b) depicts baseline means and treatment effects from seven to fifteen years after college application.

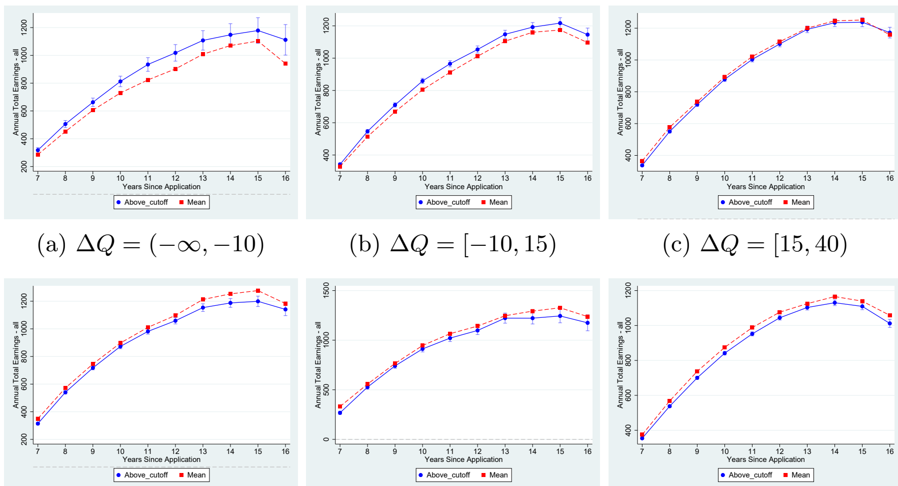
Threshold-crossing effects across quality fallbacks. Overall earnings effects reflect adjustments along multiple margins of college choice. As shown in Figures 1 and 2, being marginally ineligible for a given program in a particular year does not meaningfully alter the likelihood of enrolling in or graduating from an alternative major. As a result, crossing the eligibility threshold primarily affects the intensive margin of college and major choice. These reallocations across programs, however, do not appear to translate into differences in overall earnings.

We next analyze cases where eligibility effectively shifts students' potential choices between programs of different quality. These comparisons speak more directly to our research objective, which is to estimate the impact of expanding capacity in higher-quality or more selective programs relative to lower-quality and less selective majors.

The next set of figures examines earnings effects as a function of the quality difference between the target program and the next-best feasible alternative. For each application, we identify the next-best feasible option, where feasibility is defined by eligibility given the applicant's PSU scores and the program-specific test weights used by each major–institution pair. Once the next-best alternative is identified, we define ΔQ as the difference in average PSU scores of admitted students between the target program and the feasible next-best alternative, measured at the program–year level. We interpret ΔQ as a measure of the selectivity gap between the target program and the next-best option.

Figure 4 presents earnings effects across six intervals of selectivity differences ΔQ. Panels (a) through (e) span the range from negative values—where the next-best alternative is more selective than the target program—to increasingly positive values, corresponding to larger selectivity advantages of the target program. Panel (f) considers cases in which no next-best feasible alternative exists within the centralized admission system, meaning that the applicant is not eligible for any other program listed in their ROL. We find that when the next-best alternative is of higher quality ($\Delta Q < 0$), crossing the eligibility threshold increases earnings. As ΔQ becomes more positive (the next best is a less selective program), the earnings effect attenuates and eventually turns negative. Similarly, when no feasible alternatives exist within the system, threshold crossing is associated with lower earnings.

Figure 4: Effects of higher education program eligibility on earnings across selectivity differences



(a) $\Delta Q = (-\infty, -10)$ (b) $\Delta Q = [-10, 15)$ (c) $\Delta Q = [15, 40)$

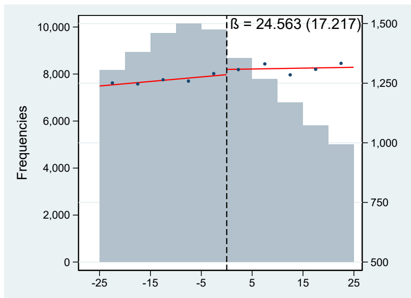
(d) $\Delta Q = [40, 70)$ (e) $\Delta Q = [70, +\infty)$ (f) No feasible fallback choice

Notes: This figure shows the effects of major eligibility on earnings across differences in selectivity between the target program and the next-best alternative. The outcome variable is average monthly earnings measured by years since college application. Program selectivity is defined as the average PSU score of admitted students in each program–year. For each application, we compute the selectivity gap between the target program and the next-best feasible alternative, denoted by ΔQ. Panels (a) through (e) report regression discontinuity estimates of the effect of crossing the eligibility cutoff on earnings for subsamples defined by different intervals of ΔQ. Panel (f) presents analogous estimates for the subsample of applications with no next-best feasible alternative within the centralized admission system.

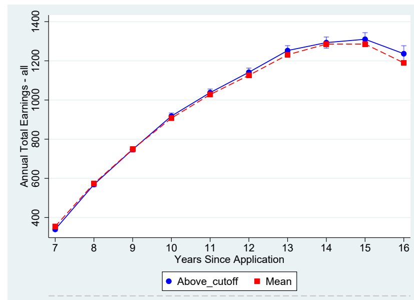
Next figure conditions the fallback choices such that we isolate horizontal cross-field differences in majors. To this end, we restrict our sample to applications such that its next-best feasible fallback option is a program belonging to the same major type of that of the target program. We estimate first an overall threshold-crossing effect and then we separate across categories of ΔQ.

Figure 5 presents effects for the sample mentioned above. Panel (a) illustrates effect 15 years after college applications while panel (b) depict the full set of point estimates across years after applications.

Figure 5: Effects of higher education program eligibility on earnings: target and fallback are in the same area of study



(a) Effect on earnings 15 years after application

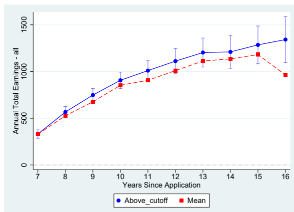


(b) Dynamic threshold-crossing effects

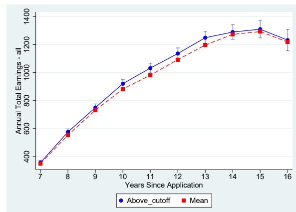
Notes: this figure documents threshold-crossing effects on earnings. The outcome variable is annual earnings. Panel (a) depicts average earnings, 15 years after application, as a function of PSU scores for marginally eligible and not eligible to a program. Panel (b) depicts baseline means and treatment effects from seven to fifteen years after college application. For both panels, we restrict the sample to applications where the target and next best share the same area of study.

Figure 6 estimates heterogeneous effects across ΔQ for the sample of applications such that the target and fallback options belong to the same field of study.

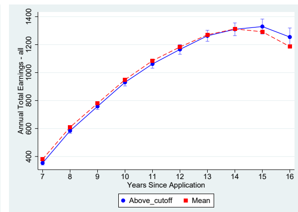
Figure 6: Effects of higher education program eligibility on earnings across selectivity differences: target and fallback are in the same area of study



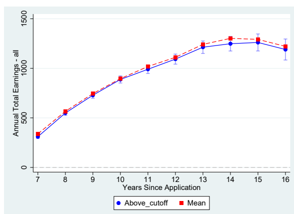
(a) $\Delta Q = (-\infty, -10)$



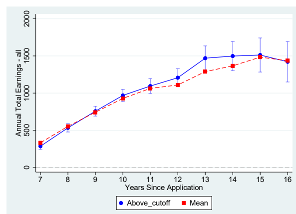
(b) $\Delta Q = [-10, 15)$



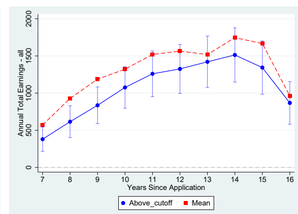
(c) $\Delta Q = [15, 40)$



(d) $\Delta Q = [40, 70)$



(e) $\Delta Q = [70, +\infty)$

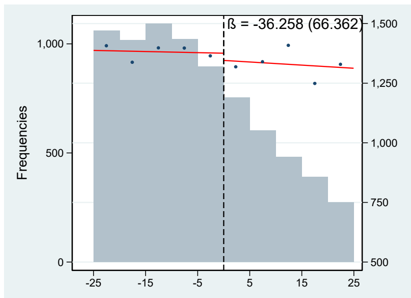


(f) No feasible fallback choice

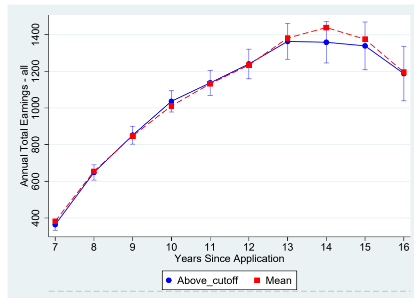
Notes: This figure shows the effects of major eligibility on earnings across differences in selectivity between the target program and the next-best alternative. The outcome variable is average monthly earnings measured by years since college application. Program selectivity is defined as the average PSU score of admitted students in each program–year. For each application, we compute the selectivity gap between the target program and the next-best feasible alternative, denoted by ΔQ. Panels (a) through (e) report regression discontinuity estimates of the effect of crossing the eligibility cutoff on earnings for subsamples defined by different intervals of ΔQ. Panel (f) presents analogous estimates for the subsample of applications with no next-best feasible alternative within the centralized admission system. All panels restrict the sample to applications in which the target and fallback programs belong to the same field of study.

Threshold-crossing effects to highly selective majors. In these exercises, we estimate eligibility effects to highly selective majors. First, we restrict our attention to applications to Universidad de Chile and Pontificia Universidad Católica. Furthermore, we condition on applications such that the fallback program is in the same field of study. Figure 7, panel (a), illustrates effects 15 years after application and panel (b) documents the full set of dynamic RDDs. Panels (c) and (d) further restrict our sample to applications to the following programs: business engineering (ingeniería comercial), law, engineering, and medicine, either at Universidad de Chile or Pontificia Universidad Católica.

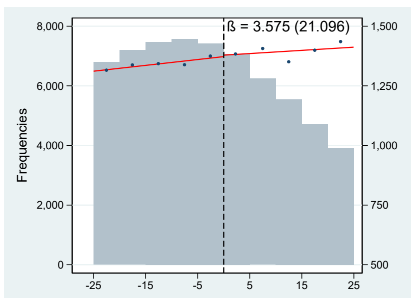
Figure 7: Effects of higher education program eligibility on earnings: target degree are programs from Universidad de Chile or Pontificia Universidad Católica



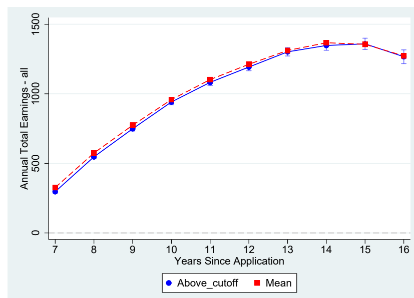
(a) Effect on earnings 15 years after application



(b) Dynamic threshold-crossing effects



(c) Effect on earnings 15 years after application: elite programs



(d) Dynamic threshold-crossing effects: elite programs

Notes: this figure documents threshold-crossing effects on earnings. The outcome variable is annual earnings. Panels (a) and (b) restrict the sample to applications to either at Universidad de Chile or Pontificia Universidad Católica such that the fallback program is in the same area. Panels (c) and (d) restrict the sample to applications to business engineering, law, engineering, and medicine, either at Universidad de Chile or Pontificia Universidad Católica. Panels (a) and (c) depicts average earnings, 15 years after application, as a function of PSU scores for marginally eligible and not eligible to a program. Panels (b) and (d) depicts baseline means and treatment effects from seven to fifteen years after college application.

4 Effects of expansions on inframarginal students

4.1 OLS Regressions Vacancies on Wages (Program-Level Analysis, CRUCH universities)

These regressions use only students who attended CRUCH universities, institutions that participated in the admission system pre 2011 when many private universities enter the admission system. I use CRUCH membership as a proxy for selectivity and participate in the centralized admissions process. However, this classification is imperfect: some private universities are highly selective, and selectivity varies substantially within CRUCH. We need to implement a more direct measure.

This analysis examines the relationship between program-level vacancy conditions at the time of program entry and wages at labor market entry for college graduates.

The sample consists of college graduates observed in their first year after completing their degrees. The key identifying assumption is that vacancies measured when students enter their program (6 years before graduation for college students) don't reflect current labor market conditions.

4.1.1 Levels Specification

The basic regression equation in levels is:

$$\ln(w_{ipt}) = \beta \ln(V_{pt-6}) + \alpha_{FE} + \varepsilon_{ipt} \tag{8}$$

where w_{ipt} is the wage for worker i in program p at time t (first year after graduation), V_{pt-6} represents vacancies in program p six years prior (when the cohort entered the program), and α_{FE} represents different sets of fixed effects.

Five specifications are estimated with progressively more saturated fixed effects:

- **Model 0 (No FE)**: No fixed effects, with standard errors clustered at the career level.
- **Model 1 (Cohort FE)**: Includes cohort fixed effects (α_c) to control for aggregate time trends affecting all graduates in a given year. Standard errors are clustered at the career level.
- **Model 2 (Career FE)**: Includes career fixed effects (α_p) to absorb all time-invariant characteristics of each career. Standard errors are clustered at the career level.
- **Model 3 (Cohort + Career FE)**: Includes both cohort and career fixed effects simultaneously ($\alpha_c + \alpha_p$). This specification absorbs both aggregate time trends and career-specific characteristics, identifying the effect from within-career, within-cohort variation.
- **Model 4 (Cohort-Program FE)**: Includes cohort-program fixed effects (α_{cp}). This absorbs all fixed characteristics of each cohort-program cell.

Table 1: Effect of Program-Level Vacancies on Entry Wages (Levels)

	(1)	(2)	(3)	(4)	(5)
ln(vacancies t-6)	0.231***	0.231***	0.172***	0.171***	0.036
	(0.035)	(0.035)	(0.034)	(0.034)	(0.060)
Observations	111824	111824	111817	111817	111407
Cohort FE		✓		✓	
Career FE			✓	✓	
Cohort-Program FE					✓
Clustering	Career	Career	Career	Career	Career

Notes: Dependent variable is log wage in the first year after degree completion. The independent variable is the log of program-level vacancies lagged 6 years (measured when the cohort was entering the program). Model 0 has no fixed effects; Model 1 includes cohort fixed effects; Model 2 includes career fixed effects; Model 3 includes both cohort and career fixed effects; Model 4 includes cohort-program fixed effects (most saturated). Standard errors are clustered at the career level to allow for arbitrary correlation within career.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.1.2 First Differences Specification

To address concerns about level effects and time-invariant unobservables, we also estimate the model in first differences:

$$\Delta \ln(w_{ipt}) = \beta \Delta \ln(V_{pt-6}) + \alpha_{FE} + \varepsilon_{ipt} \tag{9}$$

where Δ denotes the first difference operator. This specification examines whether changes in vacancy conditions are associated with changes in entry wages, controlling for fixed differences across cohorts or careers. I think the FE are not necessary, but I added just to see what happens

Table 2: Effect of Program-Level Vacancies on Entry Wages (First Differences)

	(1) No FE	(2) Cohort FE	(3) Career FE
Δln(vacancies)	0.165***	0.168***	0.107***
	(0.032)	(0.033)	(0.031)
Observations	59243	59243	59238
Clustering	Career	Career	Career

Notes: Dependent variable is the change in log wage in the first year after degree completion. The independent variable is the change in log of program-level vacancies. Model 0 has no fixed effects; Model 1 includes cohort fixed effects; Model 2 includes career fixed effects. Standard errors are clustered at the career level to allow for arbitrary correlation within career.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.2 Two-Stage Least Squares (IV) Estimation

We use a two-stage least squares (2SLS) approach to instrument first-year enrollment with program-level vacancies.

4.2.1 IV Specification in Levels

The two-stage system is:

First Stage:

$$\ln(M_{pt-6}) = \pi \ln(V_{pt-6}) + \alpha_{FE} + \nu_{ipt} \tag{10}$$

Second Stage:

$$\ln(w_{ipt}) = \beta \widehat{\ln(M_{pt-6})} + \alpha_{FE} + \varepsilon_{ipt} \tag{11}$$

where M_{pt-6} is first-year enrollment (matricula_primer_ano) in program p six years prior, V_{pt-6} represents vacancies, and $\widehat{\ln(M_{pt-6})}$ denotes the predicted value from the first stage.

The identifying assumption is that vacancies affect wages only through their impact on enrollment, conditional on the fixed effects.

- **Model 0 (No FE)**: IV with no fixed effects. Shows the unconditional IV relationship.
- **Model 1 (Cohort FE)**: IV with cohort fixed effects. Identifies from cross-program variation in how vacancies affect enrollment and subsequent wages.
- **Model 2 (Career FE)**: IV with career fixed effects. Identifies from within-career variation across cohorts in the vacancy-enrollment-wage relationship.

Table 3: IV Estimates: Effect of First-Year Enrollment on Entry Wages (Levels)

Panel A: Second Stage			
	(1)	(2)	(3)
	No FE	Cohort FE	Career FE
ln(enrollment t-6)	0.232***	0.233***	0.175***
	(0.033)	(0.032)	(0.034)
Constant	14.670***		
	(0.134)		
Observations	111208	111208	111201
Panel B: First Stage			
ln(vacancies t-6)	1.034***	1.034***	1.014***
	(0.017)	(0.017)	(0.018)
Cohort FE		✓	
Career FE			✓
Clustering	Career	Career	Career

Notes: Two-stage least squares estimates. Dependent variable in second stage is log wage in the first year after degree completion. The endogenous variable is log first-year enrollment (matricula primer ano), instrumented by log program-level vacancies lagged 6 years. Panel B reports the first-stage relationship between vacancies and enrollment. Model 0 has no fixed effects; Model 1 includes cohort fixed effects; Model 2 includes career fixed effects. Standard errors are clustered at the career level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.2.2 IV Specification in First Differences

For robustness, we also estimate the IV model in first differences:

First Stage:

$$\Delta \ln(M_{pt-6}) = \pi \Delta \ln(V_{pt-6}) + \alpha_{FE} + \nu_{ipt} \tag{12}$$

Second Stage:

$$\Delta \ln(w_{ipt}) = \beta \Delta \widehat{\ln(M_{pt-6})} + \alpha_{FE} + \varepsilon_{ipt} \tag{13}$$

The first differences specification removes time-invariant unobservables at the program level and focuses on how changes in enrollment (induced by changes in vacancies) relate to changes in wages.

- **Model 0 (No FE)**: IV in first differences with no fixed effects.
- **Model 1 (Cohort FE)**: IV in first differences with cohort fixed effects.
- **Model 2 (Career FE)**: IV in first differences with career fixed effects.

Table 4: IV Estimates: Effect of First-Year Enrollment on Entry Wages (First Differences)

	Panel A: Second Stage		
	(1)	(2)	(3)
	No FE	Cohort FE	Career FE
Δln(enrollment)	0.304***	0.305***	0.199**
	(0.092)	(0.091)	(0.082)
Observations	58735	58735	58731
	Panel B: First Stage		
Δln(vacancies)	0.506***	0.517***	0.519***
	(0.085)	(0.087)	(0.092)
Cohort FE		✓	
Career FE			✓
Clustering	Career	Career	Career

Notes: Two-stage least squares estimates in first differences. Dependent variable in second stage is the change in log wage. The endogenous variable is the change in log first-year enrollment, instrumented by the change in log program-level vacancies. Panel B reports the first-stage relationship. Model 0 has no fixed effects; Model 1 includes cohort fixed effects; Model 2 includes career fixed effects. Standard errors are clustered at the career level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.3 Program-Cohort Level Analysis

We collapse the data to the program-cohort level by computing average wages for each program-cohort combination.

4.3.1 OLS Specification at Program-Cohort Level

The regression equation at the program-cohort level is:

$$\overline{\ln(w)}_{pc} = \beta \ln(V_{pc-6}) + \alpha_{FE} + \varepsilon_{pc} \tag{14}$$

where $\overline{\ln(w)}_{pc}$ is the average log wage for program p in cohort c, and V_{pc-6} represents vacancies when that cohort entered the program.

Four specifications are estimated:

- **Model 0 (No FE)**: No fixed effects, showing the unconditional correlation at program-cohort level.
- **Model 1 (Cohort FE)**: Controls for cohort-level trends affecting all programs.
- **Model 2 (Career FE)**: Controls for time-invariant career characteristics.
- **Model 3 (Cohort + Career FE)**: Controls for both dimensions simultaneously.

Table 5: Effect of Vacancies on Average Wages at Program-Cohort Level (Levels)

	(1) No FE	(2) Cohort FE	(3) Career FE	(4) Cohort+Career FE
ln(vacancies t-6)	0.111***	0.111***	0.084**	0.084**
	(0.032)	(0.032)	(0.033)	(0.032)
Observations	6920	6920	6899	6899
Cohort FE		✓		✓
Career FE			✓	✓
Clustering	Career	Career	Career	Career

Notes: Dependent variable is average log wage by program-cohort in the first year after degree completion. The independent variable is the log of program-level vacancies lagged 6 years. Model 0 has no fixed effects; Model 1 includes cohort fixed effects; Model 2 includes career fixed effects; Model 3 includes both. Standard errors are clustered at the career level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.3.2 IV Specification at Program-Cohort Level

The IV system at the program-cohort level is:

First Stage:

$$\ln(M_{pc-6}) = \pi \ln(V_{pc-6}) + \alpha_{FE} + \nu_{pc} \tag{15}$$

Second Stage:

$$\overline{\ln(w)}_{pc} = \beta \widehat{\ln(M_{pc-6})} + \alpha_{FE} + \varepsilon_{pc} \tag{16}$$

where M_{pc-6} is first-year enrollment for program p and cohort c.
Four specifications mirror the OLS analysis:

- **Model 0 (No FE)**: IV with no fixed effects.
- **Model 1 (Cohort FE)**: IV with cohort fixed effects.
- **Model 2 (Career FE)**: IV with career fixed effects.
- **Model 3 (Cohort + Career FE)**: IV with both fixed effects.

Table 6: IV Estimates: Effect of Enrollment on Average Wages at Program-Cohort Level (Levels)

	(1)	(2)	(3)	(4)
Panel A: Second Stage				
ln(enrollment t-6)	0.110***	0.109***	0.086***	0.086***
	(0.030)	(0.030)	(0.032)	(0.032)
Constant	15.042***			
	(0.116)			
Observations	6879	6879	6857	6857
Panel B: First Stage				
ln(vacancies t-6)	1.029***	1.032***	1.000***	1.001***
	(0.031)	(0.031)	(0.037)	(0.037)
Cohort FE		✓		✓
Career FE			✓	✓
Clustering	Career	Career	Career	Career

Notes: Two-stage least squares estimates at program-cohort level. Dependent variable in second stage is average log wage by program-cohort. The endogenous variable is log first-year enrollment, instrumented by log program-level vacancies lagged 6 years. Model 0 has no fixed effects; Model 1 includes cohort fixed effects; Model 2 includes career fixed effects; Model 3 includes both. Standard errors are clustered at the career level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.3.3 First Differences at Program-Cohort Level

The first differences specification at the program-cohort level is:

$$\Delta\overline{\ln(w)}_{pc} = \beta\Delta\ln(V_{pc-6}) + \alpha_{FE} + \varepsilon_{pc} \tag{17}$$

Two specifications are estimated:

- **Model 0 (No FE)**: Pure first differences, which automatically removes all time-invariant effects.
- **Model 1 (Career FE)**: First differences with career fixed effects to control for career-specific trends.

Table 7: Effect of Vacancies on Average Wages at Program-Cohort Level (First Differences)

	(1)	(2)
Δln(vacancies)	0.011	0.013
	(0.038)	(0.038)
Observations	5750	5736
Career FE		✓
Clustering	Career	Career

Notes: Dependent variable is the change in average log wage by program-cohort. The independent variable is the change in log of program-level vacancies. Model 1 has no fixed effects (differencing removes time-invariant effects); Model 2 includes career fixed effects. Standard errors are clustered at the career level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.3.4 IV in First Differences at Program-Cohort Level

The IV specification in first differences at the program-cohort level is:

First Stage:

$$\Delta \ln(M_{pc-6}) = \pi \Delta \ln(V_{pc-6}) + \alpha_{FE} + \nu_{pc} \tag{18}$$

Second Stage:

$$\Delta \overline{\ln(w)}_{pc} = \beta \Delta \widehat{\ln(M_{pc-6})} + \alpha_{FE} + \varepsilon_{pc} \tag{19}$$

Two specifications are estimated:

- **Model 0 (No FE)**: IV in pure first differences.
- **Model 1 (Career FE)**: IV in first differences with career fixed effects.

Table 8: IV Estimates: Effect of Enrollment on Average Wages at Program-Cohort Level (First Differences)

Panel A: Second Stage		
	(1)	(2)
	No FE	Career FE
Δln(enrollment)	0.043	0.049
	(0.118)	(0.116)
Constant	0.031***	
	(0.004)	
Observations	5715	5700
Panel B: First Stage		
Δln(vacancies)	0.330***	0.338***
	(0.059)	(0.060)
Career FE		✓
Clustering	Career	Career

Notes: Two-stage least squares estimates in first differences at program-cohort level. Dependent variable in second stage is the change in average log wage. The endogenous variable is the change in log first-year enrollment, instrumented by the change in log program-level vacancies. Model 0 has no fixed effects; Model 1 includes career fixed effects. Standard errors are clustered at the career level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.4 Less Selective Universities (Privadas) Analysis

This section replicates the previous analysis by student and program-cohort. But restricts the sample to less selective private universities (clasificacion1 = "(b) Universidades Privadas").

4.5 OLS Regressions on Students - Private Universities

4.5.1 Levels Specification - Private Universities

This specification mirrors Set 4, using only students from less selective private universities. The regression equation is:

$$\ln(w_{ipt}) = \beta \ln(V_{pt-6}) + \alpha_{FE} + \varepsilon_{ipt} \tag{20}$$

where all variables are defined as before, but the sample is restricted to private university graduates.

Five specifications with progressively more saturated fixed effects are estimated:

- **Model 0 (No FE)**: No fixed effects, baseline correlation for private universities.

- **Model 1 (Cohort FE)**: Cohort fixed effects to control for aggregate time trends.
- **Model 2 (Career FE)**: Career fixed effects to absorb time-invariant career characteristics.
- **Model 3 (Cohort + Career FE)**: Both cohort and career fixed effects.
- **Model 4 (Cohort-Program FE)**: Most saturated specification with cohort-program fixed effects.

Table 9: Effect of Program-Level Vacancies on Entry Wages - Private Universities (Levels)

	(1)	(2)	(3)	(4)	(5)
ln(vacancies t-6)	0.180***	0.174***	0.120***	0.113***	0.105***
	(0.044)	(0.046)	(0.039)	(0.042)	(0.024)
Observations	87270	87270	87263	87263	86956
Cohort FE		✓		✓	
Career FE			✓	✓	
Cohort-Program FE					✓
Clustering	Career	Career	Career	Career	Career

Notes: Dependent variable is log wage in the first year after degree completion. The independent variable is the log of program-level vacancies lagged 6 years (measured when the cohort was entering the program). Model 0 has no fixed effects; Model 1 includes cohort fixed effects; Model 2 includes career fixed effects; Model 3 includes both cohort and career fixed effects; Model 4 includes cohort-program fixed effects (most saturated). Standard errors are clustered at the career level to allow for arbitrary correlation within career.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.5.2 First Differences Specification - Private Universities

The first differences specification for private universities is:

$$\Delta \ln(w_{ipt}) = \beta \Delta \ln(V_{pt-6}) + \alpha_{FE} + \varepsilon_{ipt} \tag{21}$$

Three specifications are estimated:

- **Model 0 (No FE)**: First differences with no fixed effects.
- **Model 1 (Cohort FE)**: First differences with cohort fixed effects.
- **Model 2 (Career FE)**: First differences with career fixed effects.

Table 10: Effect of Program-Level Vacancies on Entry Wages - Private Universities (First Differences)

	(1) No FE	(2) Cohort FE	(3) Career FE
$nDelta\ln(\text{vacancies})$	-0.022	-0.031	-0.003
	(0.030)	(0.027)	(0.017)
Observations	43359	43359	43350
Clustering	Career	Career	Career

Notes: Dependent variable is the change in log wage in the first year after degree completion. The independent variable is the change in log of program-level vacancies. Model 0 has no fixed effects; Model 1 includes cohort fixed effects; Model 2 includes career fixed effects. Standard errors are clustered at the career level to allow for arbitrary correlation within career.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.6 Two-Stage Least Squares (IV) Estimation - Private Universities

4.6.1 IV Specification in Levels - Private Universities

The two-stage system for private universities is:

First Stage:

$$\ln(M_{pt-6}) = \pi \ln(V_{pt-6}) + \alpha_{FE} + \nu_{ipt} \tag{22}$$

Second Stage:

$$\ln(w_{ipt}) = \beta \ln(\widehat{M_{pt-6}}) + \alpha_{FE} + \varepsilon_{ipt} \tag{23}$$

Three specifications are estimated:

- **Model 0 (No FE)**: IV with no fixed effects.
- **Model 1 (Cohort FE)**: IV with cohort fixed effects.
- **Model 2 (Career FE)**: IV with career fixed effects.

Table 11: IV Estimates: Effect of First-Year Enrollment on Entry Wages - Private Universities (Levels)

Panel A: Second Stage			
	(1)	(2)	(3)
	No FE	Cohort FE	Career FE
ln(enrollment t-6)	0.181***	0.175***	0.123***
	(0.042)	(0.044)	(0.039)
Constant	14.710***		
	(0.172)		
Observations	87056	87056	87050
Panel B: First Stage			
ln(vacancies t-6)	1.007***	1.009***	0.985***
	(0.020)	(0.020)	(0.022)
Cohort FE		✓	
Career FE			✓
Clustering	Career	Career	Career

Notes: Two-stage least squares estimates. Dependent variable in second stage is log wage in the first year after degree completion. The endogenous variable is log first-year enrollment (matricula primer ano), instrumented by log program-level vacancies lagged 6 years. Panel B reports the first-stage relationship between vacancies and enrollment. Model 0 has no fixed effects; Model 1 includes cohort fixed effects; Model 2 includes career fixed effects. Standard errors are clustered at the career level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.6.2 IV Specification in First Differences - Private Universities

The IV model in first differences for private universities:

First Stage:

$$\Delta \ln(M_{pt-6}) = \pi \Delta \ln(V_{pt-6}) + \alpha_{FE} + \nu_{ipt} \tag{24}$$

Second Stage:

$$\Delta \ln(w_{ipt}) = \beta \Delta \widehat{\ln(M_{pt-6})} + \alpha_{FE} + \varepsilon_{ipt} \tag{25}$$

Three specifications are estimated:

- **Model 0 (No FE)**: IV in first differences with no fixed effects.
- **Model 1 (Cohort FE)**: IV in first differences with cohort fixed effects.
- **Model 2 (Career FE)**: IV in first differences with career fixed effects.

Table 12: IV Estimates: Effect of First-Year Enrollment on Entry Wages - Private Universities (First Differences)

	(1)	(2)	(3)
Panel A: Second Stage			
$nDelta\ln$(enrollment)	-0.090	-0.121	-0.016
	(0.107)	(0.104)	(0.064)
Observations	43234	43234	43225
Panel B: First Stage			
$\Delta\ln$(vacancies)	0.304***	0.299***	0.293***
	(0.057)	(0.057)	(0.054)
Cohort FE		✓	
Career FE			✓
Clustering	Career	Career	Career

Notes: Two-stage least squares estimates in first differences. Dependent variable in second stage is the change in log wage. The endogenous variable is the change in log first-year enrollment, instrumented by the change in log program-level vacancies. Panel B reports the first-stage relationship. Model 0 has no fixed effects; Model 1 includes cohort fixed effects; Model 2 includes career fixed effects. Standard errors are clustered at the career level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.7 Program-Cohort Level Analysis - Private Universities

4.7.1 OLS Specification at Program-Cohort Level - Private Universities

The regression equation at the program-cohort level for private universities is:

$$\overline{\ln(w)}_{pc} = \beta \ln(V_{pc-6}) + \alpha_{FE} + \varepsilon_{pc} \tag{26}$$

Four specifications are estimated:

- **Model 0 (No FE)**: No fixed effects.
- **Model 1 (Cohort FE)**: Cohort fixed effects.
- **Model 2 (Career FE)**: Career fixed effects.
- **Model 3 (Cohort + Career FE)**: Both cohort and career fixed effects.

Table 13: Effect of Vacancies on Average Wages at Program-Cohort Level - Private Universities (Levels)

	(1)	(2)	(3)	(4)
ln(vacancies t-6)	0.054**	0.043*	0.049*	0.035
	(0.025)	(0.025)	(0.025)	(0.025)
Observations	4102	4102	4071	4071
Cohort FE		✓		✓
Career FE			✓	✓
Clustering	Career	Career	Career	Career

Notes: Dependent variable is average log wage by program-cohort in the first year after degree completion. The independent variable is the log of program-level vacancies lagged 6 years. Model 0 has no fixed effects; Model 1 includes cohort fixed effects; Model 2 includes career fixed effects; Model 3 includes both. Standard errors are clustered at the career level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.7.2 IV Specification at Program-Cohort Level - Private Universities

The IV system at the program-cohort level for private universities:

First Stage:

$$\ln(M_{pc-6}) = \pi \ln(V_{pc-6}) + \alpha_{FE} + \nu_{pc} \tag{27}$$

Second Stage:

$$\overline{\ln(w)}_{pc} = \beta \ln(\widehat{M_{pc-6}}) + \alpha_{FE} + \varepsilon_{pc} \tag{28}$$

Four specifications are estimated:

- **Model 0 (No FE)**: IV with no fixed effects.
- **Model 1 (Cohort FE)**: IV with cohort fixed effects.
- **Model 2 (Career FE)**: IV with career fixed effects.
- **Model 3 (Cohort + Career FE)**: IV with both fixed effects.

Table 14: IV Estimates: Effect of Enrollment on Average Wages at Program-Cohort Level - Private Universities (Levels)

	Panel A: Second Stage			
	(1)	(2)	(3)	(4)
ln(enrollment t-6)	0.072**	0.059**	0.072**	0.054*
	(0.029)	(0.029)	(0.030)	(0.030)
Constant	15.075***			
	(0.112)			
Observations	4009	4009	3977	3977
	Panel B: First Stage			
ln(vacancies t-6)	0.826***	0.832***	0.766***	0.777***
	(0.039)	(0.039)	(0.043)	(0.042)
Cohort FE		✓		✓
Career FE			✓	✓
Clustering	Career	Career	Career	Career

Notes: Two-stage least squares estimates at program-cohort level. Dependent variable in second stage is average log wage by program-cohort. The endogenous variable is log first-year enrollment, instrumented by log program-level vacancies lagged 6 years. Model 0 has no fixed effects; Model 1 includes cohort fixed effects; Model 2 includes career fixed effects; Model 3 includes both. Standard errors are clustered at the career level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.7.3 First Differences at Program-Cohort Level - Private Universities

The first differences specification at the program-cohort level for private universities:

$$\Delta\overline{\ln(w)}_{pc} = \beta\Delta\ln(V_{pc-6}) + \alpha_{FE} + \varepsilon_{pc} \tag{29}$$

Two specifications are estimated:

- **Model 0 (No FE)**: Pure first differences.
- **Model 1 (Career FE)**: First differences with career fixed effects.

Table 15: Effect of Vacancies on Average Wages at Program-Cohort Level - Private Universities (First Differences)

	(1)	(2)
Δln(vacancies)	-0.010	-0.019
	(0.029)	(0.029)
Observations	3196	3177
Career FE		✓
Clustering	Career	Career

Notes: Dependent variable is the change in average log wage by program-cohort. The independent variable is the change in log of program-level vacancies. Model 1 has no fixed effects (differencing removes time-invariant effects); Model 2 includes career fixed effects. Standard errors are clustered at the career level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.7.4 IV in First Differences at Program-Cohort Level - Private Universities

The IV specification in first differences at the program-cohort level for private universities:

First Stage:

$$\Delta \ln(M_{pc-6}) = \pi \Delta \ln(V_{pc-6}) + \alpha_{FE} + \nu_{pc} \tag{30}$$

Second Stage:

$$\Delta \overline{\ln(w)}_{pc} = \beta \Delta \widehat{\ln(M_{pc-6})} + \alpha_{FE} + \varepsilon_{pc} \tag{31}$$

Two specifications are estimated:

- **Model 0 (No FE)**: IV in pure first differences.
- **Model 1 (Career FE)**: IV in first differences with career fixed effects.

Table 16: IV Estimates: Effect of Enrollment on Average Wages at Program-Cohort Level - Private Universities (First Differences)

Panel A: Second Stage		
	(1)	(2)
Δln(enrollment)	-0.046	-0.069
	(0.092)	(0.087)
Constant	0.027***	
	(0.007)	
Observations	3116	3098
Panel B: First Stage		
Δln(vacancies)	0.313***	0.336***
	(0.038)	(0.039)
Career FE		✓
Clustering	Career	Career

Notes: Two-stage least squares estimates in first differences at program-cohort level. Dependent variable in second stage is the change in average log wage. The endogenous variable is the change in log first-year enrollment, instrumented by the change in log program-level vacancies. Model 0 has no fixed effects; Model 1 includes career fixed effects. Standard errors are clustered at the career level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5 Summary Statistics on Vacancies, Enrollment, and Graduation

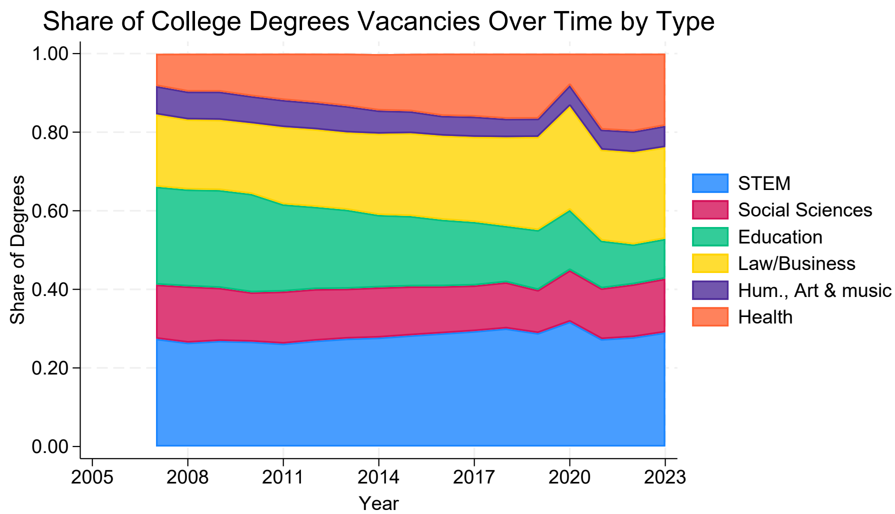


Figure 8: Share of College Degrees by Type Over Time

Table 17: Vacancies per program by education category

	Mean	Median	Std.Dev.	P10	P90	0bs.
Voc. STEM	39	30	30	19	70	25,717
Voc. Social Sciences	41	35	29	20	70	2,644
Voc. Education	44	35	35	20	80	6,456
Voc. Law/Business	42	30	44	15	80	17,992
Voc. Hum., Art and Music	34	30	22	15	60	2,796
Voc. Health	55	40	51	20	105	9,697
Coll. STEM	44	35	39	16	80	25,757
Coll. Social Sciences	47	40	34	20	80	7,939
Coll. Education	41	35	32	20	70	12,894
Coll. Law/Business	50	35	51	15	100	17,474
Coll. Hum., Art and Music	50	40	37	20	90	5,645
Coll. Health	63	54	37	30	105	8,797
Total	46	35	40	20	81	143,808

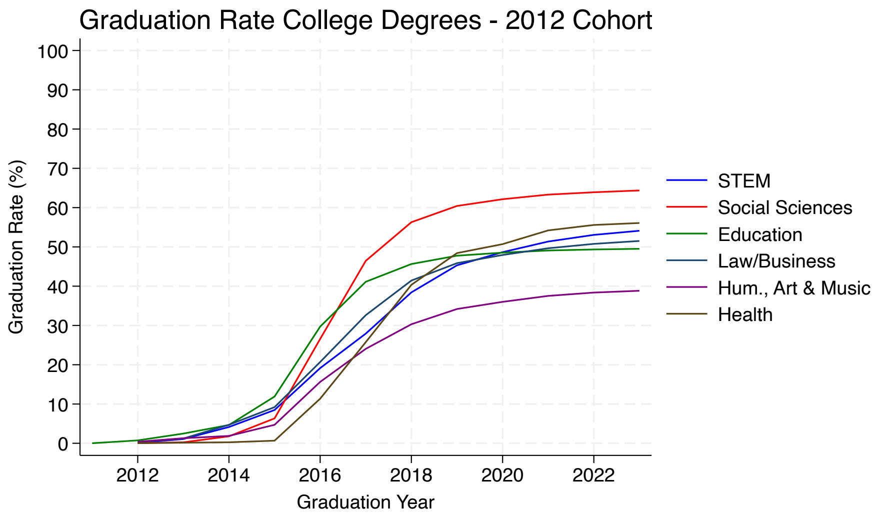


Figure 9: Graduation rates of college degrees for the 2012 cohort

Table 18: Year-on-year Change in Vacancies (%)

	Mean	Median	Std.Dev.	P10	P90
Porcentual Change in Vacancies	10.88	0.63	67.74	-34.69	56.38

5.1 College Vacancy Growth Rates by Degree and Regions

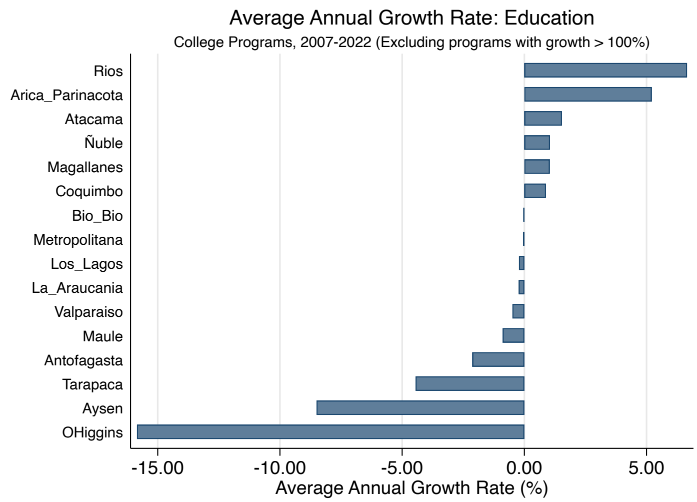


Figure 10

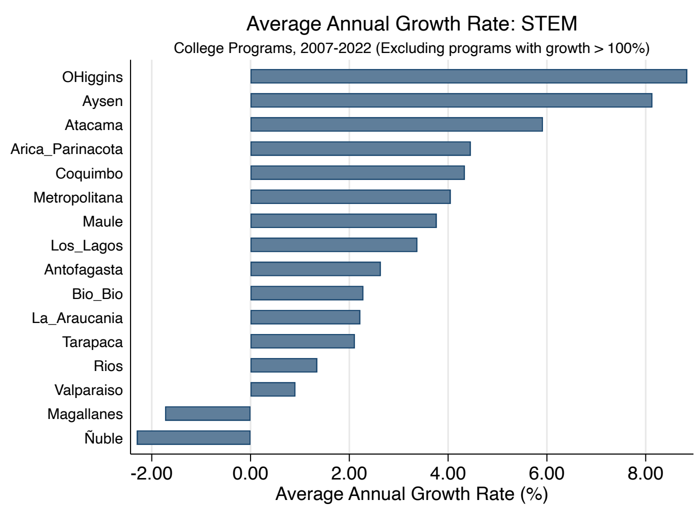


Figure 11

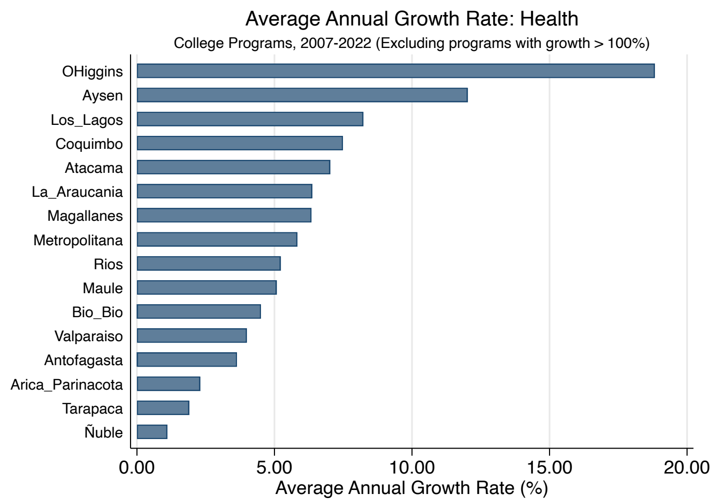


Figure 12

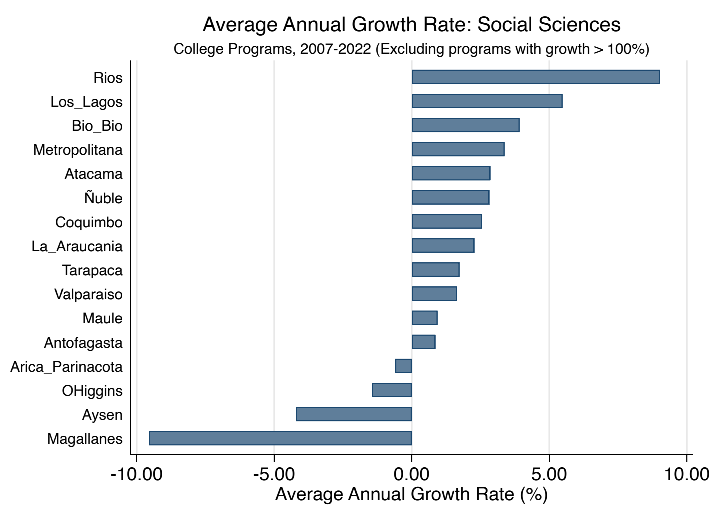


Figure 13

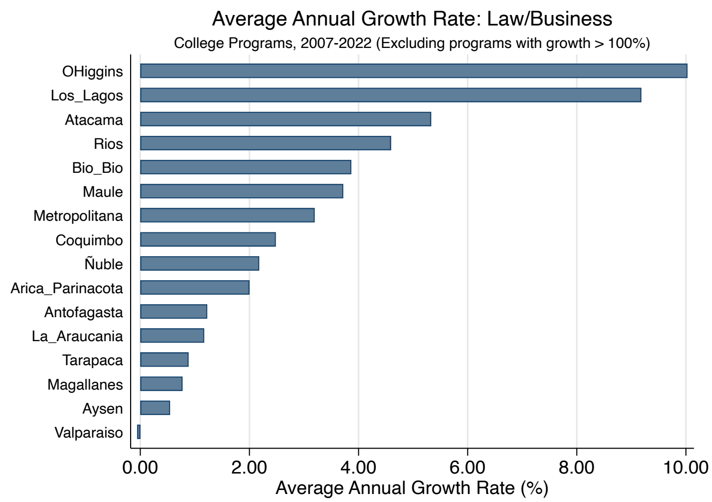


Figure 14

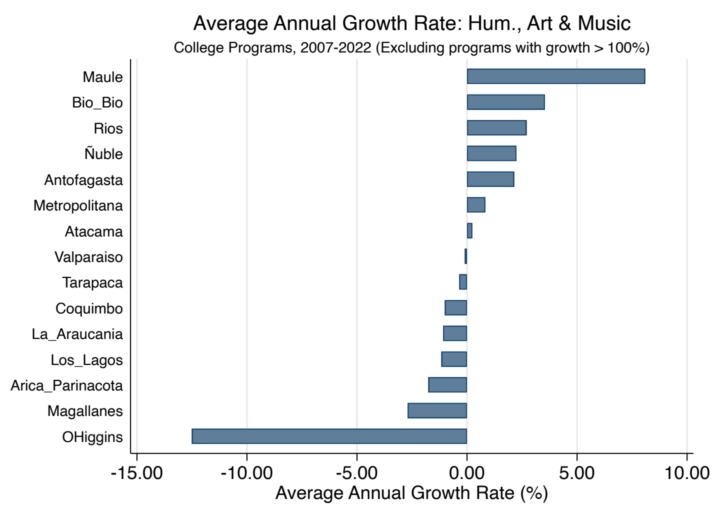


Figure 15

6 Average Growth Rate by Field and Quality

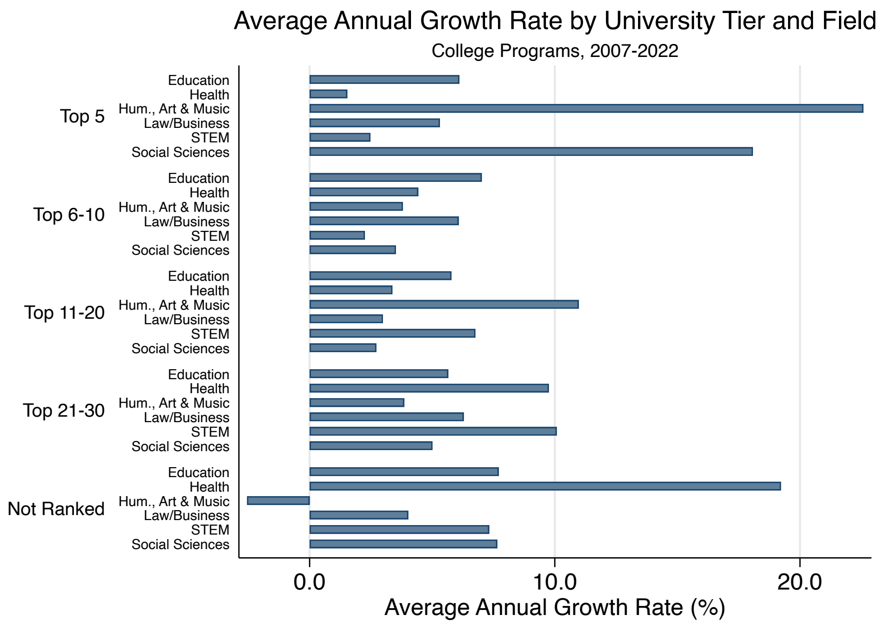


Figure 16

7 Motivating Regressions

The following regressions examine various partial correlations with vacancies. We look at vacancies for *both* bachelor and vocational programs.

7.1 Vacancies are associated with higher enrollment and graduation

All regressions are run at the province-major-year level, pooling bachelor's degree programs together. The first table shows that, after controlling for province and year fixed effects, a 100% increase in vacancies is associated with an 80% increase in enrollment. The second table shows that it is associated with a 14% increase in graduation six years later.

Table 19: Preliminary Regression: Vacancies and ln(enrollment) by Province

	No FE (1)	Year FE (2)	Prov FE (3)	Prov+Year FE (4)
Dependent variable: ln(enrollment)				
ln(vacancies)	1.136^{***}	1.133^{***}	0.797^{***}	0.767^{***}
	(0.028)	(0.028)	(0.054)	(0.047)
Constant	0.078	0.104	2.404^{***}	2.614^{***}
	(0.217)	(0.214)	(0.369)	(0.321)
Observations	682	682	681	681
R^2	0.950	0.952	0.976	0.979

Notes: Province-level OLS. "Year FE" adds year fixed effects; "Prov FE" adds province fixed effects; "Prov+Year FE" includes both. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 20: Preliminary Regression Vacancies on Graduates by Province

	(1) OLS	(2) Year FE	(3) Province FE	(4) Two-way FE
Dependent variable: ln(graduates) (province-year)				
ln(vacancies)	0.962***	0.971***	0.379***	0.140**
	(0.050)	(0.049)	(0.038)	(0.051)
Constant	-0.353	-0.419	3.906***	5.655***
	(0.416)	(0.408)	(0.279)	(0.373)
Observations	384	384	384	384
R^2	0.855	0.865	0.938	0.954

Notes: Vacancies are matched to graduates four years later (proxying average degree duration). Standard errors (in parentheses) clustered at the province level. Year FE = year fixed effects; Province FE = province fixed effects; Two-way FE includes both. Sample period: 2009–2023. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

7.2 Vacancies lower admissions cutoffs

As a program increases the number of vacancies, the admission score cutoff falls. The table below shows that a 100% increase in vacancies is associated with a 2.3% reduction in the admissions cutoff.

Table 21: Effect of Vacancies on ln(admission score cutoff)

	(1) OLS	(2) Year FE	(3) Prov FE	(4) Prog FE	(5) Year+Prog FE	(6) Full FE
Dependent variable: ln(admission score cutoff) (PSU/PDT)						
ln(vacancies)	0.048^{***}	0.048^{***}	0.037^{***}	-0.022^{***}	-0.023^{***}	-0.023^{***}
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Constant	6.065^{***}	6.065^{***}	6.106^{***}	6.334^{***}	6.336^{***}	6.336^{***}
	(0.006)	(0.005)	(0.006)	(0.008)	(0.008)	(0.008)
Year FE	✗	✓	✗	✗	✓	✓
Province FE	✗	✗	✓	✗	✗	✓
Program FE	✗	✗	✗	✓	✓	✓
Observations	29,071	29,071	29,071	28,665	28,665	28,665
R^2	0.039	0.044	0.078	0.815	0.820	0.820

Notes: Sample restricted to college/university programs. Years 2023–2024 excluded due to score scale change (PSU $\rightarrow$ PDT). Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

7.3 Vacancies are a supply constraint

The following three tables show that increases in vacancies for bachelor's programs in a given major-province-year are associated with increases in employment. The analysis pools bachelor's programs. For bachelor's programs, employment is measured 6 years later.

Table 22: Regression Log of Workers vs. Log of Degree Vacancies by Region

	Yr (1)	Yr+Prov (2)	Yr+Firm (3)	Yr+Prov+Edu (4)	Yr+Firm+Edu (5)	Yr×Edu (6)	Yr×Firm (7)
Dependent variable: ln(work)							
z_vacancies	0.106^{***}	0.228^{***}	0.318^{***}	0.108^{***}	0.135^{***}	0.039^{***}	0.359^{***}
	(0.008)	(0.072)	(0.083)	(0.025)	(0.029)	(0.011)	(0.093)
$L.$ z_vacancies	-0.026^{***}	0.120^{***}	0.171^{***}	0.038^{*}	0.033	0.002	0.180^{***}
	(0.008)	(0.034)	(0.026)	(0.021)	(0.025)	(0.011)	(0.035)
Observations	449,269	449,269	449,217	449,269	449,217	449,269	434,228

Notes: Each column reports a separate OLS specification. "Yr" = year FE; "Prov" = province FE; "Firm" = firm FE; "Edu" = education-group FE; interactions shown as indicated in column headers. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 23: Regression Log of Workers vs. Log of Degree Vacancies by Region, adding Log Wage

	Yr (1)	Yr+Prov (2)	Yr+Firm (3)	Yr+Prov+Edu (4)	Yr+Firm+Edu (5)	Yr×Edu (6)	Yr×Firm (7)
Dependent variable: ln(work)							
$\ln w$	0.379^{***}	0.349^{***}	0.479^{***}	0.341^{***}	0.424^{***}	0.326^{***}	0.581^{***}
	(0.003)	(0.045)	(0.030)	(0.055)	(0.035)	(0.047)	(0.027)
z_vacancies	0.081^{***}	0.193^{***}	0.284^{***}	0.103^{***}	0.133^{***}	0.028^{**}	0.309^{***}
	(0.008)	(0.060)	(0.065)	(0.024)	(0.026)	(0.011)	(0.065)
$L.$ z_vacancies	-0.026^{***}	0.105^{***}	0.162^{***}	0.044^{**}	0.042^{*}	-0.000	0.177^{***}
	(0.008)	(0.026)	(0.016)	(0.019)	(0.022)	(0.012)	(0.019)
Observations	449,269	449,269	449,217	449,269	449,217	449,269	434,228

Notes: Each column reports a separate OLS specification. "Yr" = year FE; "Prov" = province FE; "Firm" = firm FE; "Edu" = education-group FE; interaction structures as indicated by column headers. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 24: Dynamic Specification: Lagged ln(work) and Vacancies Across Fixed-Effect Sets

	Yr (1)	Yr+Prov (2)	Yr+Firm (3)	Yr+Prov+Edu (4)	Yr+Firm+Edu (5)	Yr×Edu (6)	Yr×Firm (7)
Dependent variable: ln(work)							
L. ln(work)	0.923^{***}	0.919^{***}	0.879^{***}	0.916^{***}	0.872^{***}	0.916^{***}	0.918^{***}
	(0.001)	(0.002)	(0.006)	(0.002)	(0.005)	(0.002)	(0.006)
z_vacancies	-0.002	0.014^{***}	0.036^{***}	-0.001	0.008	0.001	0.025^{***}
	(0.004)	(0.004)	(0.009)	(0.009)	(0.006)	(0.005)	(0.006)
L. z_vacancies	0.008^{*}	0.026^{***}	0.035^{***}	0.022^{**}	0.020^{***}	0.001	0.030^{***}
	(0.004)	(0.007)	(0.007)	(0.008)	(0.007)	(0.005)	(0.007)
Observations	362,148	362,148	362,037	362,148	362,037	362,148	346,166

Notes: Each column reports a separate OLS specification with a lagged dependent variable. "Yr" = year FE; "Prov" = province FE; "Firm" = firm FE; "Edu" = education-group FE; interactions as indicated. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Additional facts to document:

1. During the period under study, how does vacancy creation correlate with program quality? Do the programs with initially high quality expand vacancies by more or less than those with initially low quality?