ONLINE APPENDIX

for

CONNECTING STUDENT LOANS TO LABOR MARKET OUTCOMES: POLICY LESSONS FROM CHILE*

by

Harald Beyer
Centros de Estudios Publicos
haraldbeyerb@gmail.com

Justine Hastings
Brown University and NBER
justine_hastings@brown.edu

Christopher Neilson
New York University, Stern School of Business
cneilson@stern.nyu.edu

Seth Zimmerman
University of Chicago, Booth School of Business
seth.zimmerman@chicagobooth.edu

March 1, 2015

* Published as, Beyer, Harald, Justine Hastings, Christopher Neilson and Seth Zimmerman. "Connecting Student Loans to Labor Market Outcomes: Policy Lessons from Chile," *American Economic Review: Papers & Proceedings* 2015, 105(5). Noele Aabye, Phillip Ross, Unika Shrestha, Anthony Thomas and Lindsey Wilson provided outstanding research assistance. We thank the excellent leadership and staff at Mineduc and SII, including Fernando Rojas, Loreto Cox Alcaíno, Claudia Allende, Andrés Barrios, Fernando Claro, and Josefina Eluchans. Gene Amromin, Raj Chetty, Larry Katz, Brigitte Madrian and Jesse Shapiro provided helpful input and discussion. This project was supported by a grant from Brown University.

A. Additional Tables and Figures

Table A.1 shows summary statistics for students in the loan repayment sample versus the population of higher education enrollees in 2013.[1]

Table A.1. Descriptive Statistics

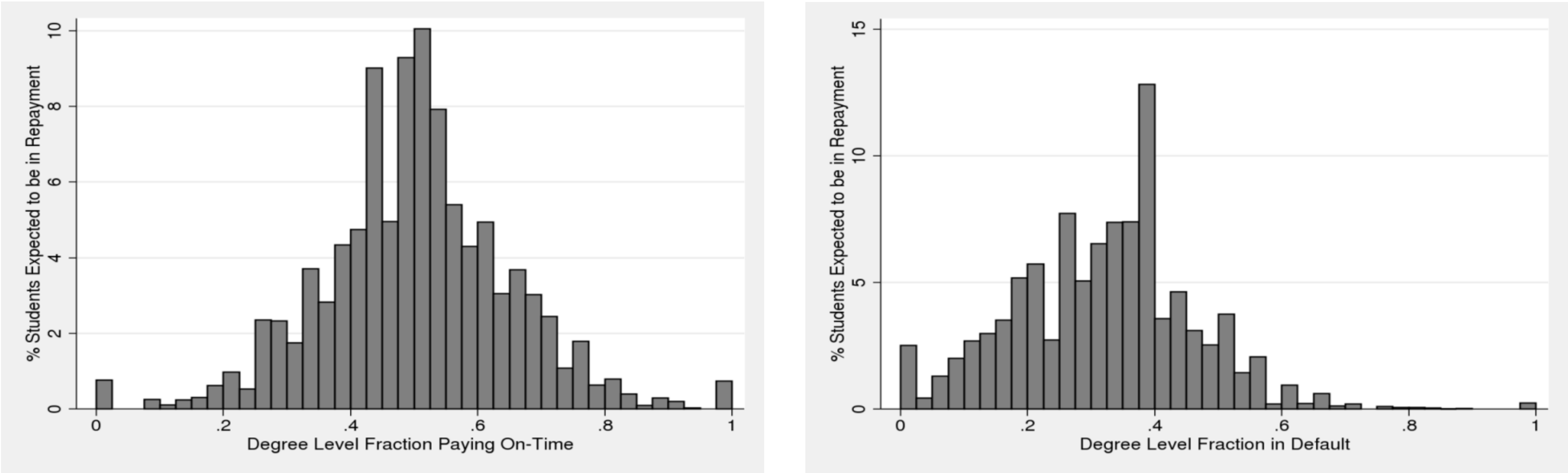
	2013 Enrollees	Repayment Sample
Individual Characteristics		
Percent low-SES	41.2%	48.4%
Ave. entrance exam score	527	500
Percent CFT/IPs	42.0%	52.9%
Percent in science/health/tech/business degrees	62.6%	53.2%
Percent in humanities/edu/arts/ag. degrees	20.3%	26.5%
Degree Characteristics		
Ave. degree graduation rate	53.6%	55.3%
Ave. past enrollee observed earnings exp. 2-7	5,943	4,910
(1,000's CLP)	(4,029)	(3,130)
Total expected tuition costs (1,000's CLP)	8,726	7,633
	(6,636)	(5,697)
Average Degree Value-Added by Selectivity Tier (1,000's CLP)		
Technical degrees	3,119	3,008
	(1,150)	(1,021)
Missing entrance exam or [0,450)	4,922	4,036
	(2,109)	(1,685)
Entrance exam [450,525)	5,213	4,883
	(2,366)	(1,920)
Entrance exam [525,600)	6,879	6,424
	(3,290)	(2,682)
Entrance exam [600,850]	9,226	8,682
	(4,474)	(4,113)
Observations	1,072,895	43,166

Notes: Standard deviations are in parentheses. Higher education enrollees in 2013 are defined as any student in the Mineduc database of higher education matriculation records in 2013. The loan repayment sample is defined as students with loans originating in 2006-2010 who should be in repayment as of April 2013. Students should be in repayment if they are not enrollees in 2011, 2012, 2013 and if their official degree length plus a 1.5 year repayment grace period should be completed by 2013. Low-SES is defined as students graduating from high schools in the lowest two (of five) poverty rating categories (RBD poverty ratings) according to Mineduc in 2010, 2008, 2006, 2003, 2001. Average entrance exam score is the average of the Math and Language components of the Prueba de Selecion Universitaria (PSU). Degree graduation percent is the percent of the 2000-2005 freshmen cohorts who graduate with a higher education degree. Earnings, tuition costs, and degree value-added are stated in thousands 2011 Chilean pesos ('000 2011 CLP). Average annual past enrollee earnings are the average earnings calculated using earnings data from the Chilean tax authority (SII) with 2-7 years of work experience. Expected tuition costs are calculated as the present value of the total tuition costs using 2013 tuition values for the length of degree (discounted at 2% APR). To calculate value-added, we estimate a flexible model of earnings in the tax authority (SII) as a function of field of study, selectivity tier of the degree, SES, PSU score, and gender controlling for tax years. We estimate fixed effects by degree. Using average characteristics by selectivity tier, we predict degree value-added.

[1] This disclosure is required by the Chilean government. SOURCE: Information contained herein comes from taxpayers' records obtained by the Chilean Internal Revenue Service (Servicio de Impuestos Internos), which was collected for tax purposes. Let the record state that the Internal Revenue Service assumes no responsibility or guarantee of any kind from the use or application made of the aforementioned information, especially in regard to the accuracy, validity or integrity.

Figure A.1. shows the fraction of students in the repayment sample as a function of these degree level on-time repayment rates and default rates. The majority of students in the loan repayment sample were previously enrolled in degrees with on-time repayment rates between 40-60 percent and default rates between 20-40 percent.

Figure A.1. Enrollment Distribution by Repayment Status



Notes: Figure presents the fraction of individuals in the loan repayment sample who were previously enrolled in degrees for each 0.025 bin of degree average repayment rate. Figure A.1 shows the distribution of loan repayment rates. The repayment sample is defined as those who are in the repayment data and expected to be in loan repayment as of April 2013. There are a total of 43,166 former enrollees in the repayment sample. These are students who were not enrolled in any HEI in 2011, 2012, 2013, who were expected to have completed the degree by 2013 based on published degree length plus a 1.5 year grace period, for all students who originated loans between 2006-2010. We calculate a degree level fraction paying their loans on time ("on-time repayment rate") as the fraction of students who were enrolled in a particular degree and could be in repayment who are current on their loan repayment as of April 2013. Likewise, the degree fraction in default ("default rate") is calculated as the fraction of a degree's past enrollees who are in repayment but are at least three loan payments behind as of April 2013.

Table A.2. Additional Degree Characteristics by 2013 Loan Repayment Status, Weighted by 2013 Enrollment

	On-Time Repayment Rate		Default Rate
	Above Median	Below Median	Above Median
Average Value Added by Selectivity Tier			
Technical degrees	3,161	3,057	2,941
	(1,051)	(998)	(863)
Missing entrance exam or [0,450)	4,770	4,696	4,623
	(1,554)	(2,003)	(2,174)
Entrance exam [450,525)	5,349	4,863	4,638
	(2,111)	(2,320)	(2,445)
Entrance exam [525,600)	7,697	5,916	6,106
	(2,557)	(3,234)	(4,284)
Entrance exam [600,850]	10,152	7,505	7,045
	(4,392)	(4,110)	(3,009)

Notes: All values are stated in thousands 2011 Chilean pesos ('000 2011 CLP). All degree-level characteristics are weighted by 2013 enrollees. Standard deviations are in parentheses. The loan repayment sample is defined as students with loans originating in 2006-2010 who should be in repayment as of April 2013. Students should be in repayment if they are not enrollees in 2011, 2012, 2013 and if their official degree length plus a 1.5 year repayment grace period should be completed by 2013. Low-SES is defined as students graduating from high schools in the lowest two (of five) poverty rating categories (RBD poverty ratings) according to Mineduc in 2010, 2008, 2006, 2003, 2001. To calculate value-added, we estimate a flexible model of earnings in the tax authority (SII) for the first seven years of labor force experience (excluding the first year) as a function of field of study, selectivity tier of the degree, SES, PSU score, and gender controlling for tax years. We estimate fixed effects by degree. Using average characteristics by selectivity tier, we predict degree value-added.

Table A.3. Additional Earnings-Based Loan Caps as a percentage of Baseline Loan Caps

	P25	Mean	P75
Enrolling Students with Entrance Exam Scores in Range			
[0,475)	42.6	94.9	122.6
[475,600)	48.4	87.2	115.0
[600,800]	72.4	117.0	150.0
Average Past Enrollee Observed Earnings			
Above median	82.3	129.5	150.0
Below median	31.2	58.3	79.3
Tuition Costs			
Above median	57.7	94.6	124.5
Below median	44.7	97.9	129.9
Average Predicted Earnings by Selectivity Tier			
Technical Degrees			
Above median	57.7	101.5	137.5
Below median	22.2	64.1	91.5
Missing Entrance Exam or [0,450)			
Above median	75.7	123.0	154.2
Below median	27.7	50.8	57.1
Entrance Exam Score [450,525)			
Above median	70.5	119.2	136.6
Below median	31.5	47.1	54.2
Entrance Exam Score [525,600)			
Above median	77.5	120.1	136.4
Below median	42.8	64.7	74.4
Entrance Exam Score [600,850)			
Above median	106.2	136.6	158.5
Below median	56.9	85.4	103.1

Notes: All values stated as percentages. P25 and P75 are the 25th and 75th percentiles of percent changes at the degree level for the type listed in the row label, weighted by 2013 enrollment. The loan repayment sample is defined as students with loans originating in 2006-2010 who should be in repayment as of April 2013. Students should be in repayment if they are not enrollees in 2011, 2012, 2013 and if their official degree length plus a 1.5 year repayment grace period should be completed by 2013. Low-SES is defined as students graduating from high schools in the lowest two (of five) poverty rating categories (RBD poverty ratings) according to Mineduc in 2010, 2008, 2006, 2003, 2001. Average entrance exam score is the average of the Math and Language components of the Prueba de Selecion Universitaria (PSU). Degree graduation percent is the percent of the 2000-2005 freshmen cohorts who graduate with a higher education degree. Earnings, tuition costs, and degree value-added are stated in thousands 2011 Chilean pesos ('000 2011 CLP). Average annual past enrollee earnings are the average earnings are calculated using earnings data from the Chilean tax authority (SII) with 2-7 years of work experience. Expected tuition costs are calculated as the present value of the total tuition costs using 2013 tuition values for the length of degree (discounted at 2% APR). To calculate value-added, we estimate a flexible model of earnings in the tax authority (SII) as a function of field of study, selectivity tier of the degree, SES, PSU score, and gender controlling for tax years. We estimate fixed effects by degree. We use the coefficients from these regressions to predict earnings as a function of personal and degree characteristics of enrollees. See Appendix B for a detailed description of variable construction.

B. Data Construction

B.1. Data Sources

High school graduation data

High school graduation data are files from the Ministry of Education (Mineduc) for the years 2002-2012. They contain year of high school graduation and high school identification numbers (RBD) for each student.

High school standardized test scores

All students take standardized tests, known as Sistema Nacional de Medición de la Calidad de la Educación (SIMCE).These files contain standardized math and language scores for students in their second year of high school. These files also contain high school identification number (RBD), gender, and student reported parental education levels. The standardized test scores files span 2001-2011.

Enrollment & graduation data

Enrollment and graduation data are official files from the Ministry of Education (Mineduc). These files contain major, major type (professional or technical), institution, institution type, freshman year, tuition costs, and other individual variables for all enrollees and graduates. Enrollment files are for the years 2000-2013. Graduation files are 2000-2011.

Entrance exam data

The college entrance exam in Chile is called the Prueba de Selecion Universitaria (PSU). The exam consists of math, language, history, and social science components. Math and language are the two mandatory components. Scores are normalized to have a mean of 500 and standard deviation of 110. The scores referred to in this analysis are defined as the individual's average of the math and language components. The PSU files also contain the student's reported high school graduation year. The PSU has been the entrance exam administered since 2004; we have scores for 2004-2013.

Loan repayment data

Loan origination data is the official data for all loan originations in years 2006-2010. The loan repayment status data is the official loan repayment data for all those that are in loan repayment as of April 2013. There are three definitions of loan repayment status from the official repayment data, as follows: on-time, one or two payments behind, and at least three payments behind. All loan origination and repayment data are obtained from the student loan authority, Ingresa - Sistema de Crédito Estudios Superiores.

B.2. Key Variable Construction

Socioeconomic status

Socioeconomic quintiles are calculated based on family average monthly per capita income. Each high school is assigned to one of five poverty rating categories defined by the Ministry of Education (Mineduc) based on the socioeconomic composition of its students. High school poverty ratings are used to construct an indicator for low-SES. A student is defined to be low-SES if the student's high school poverty rating is in the lowest two income quintiles. High school poverty rating is obtained from official high school graduation files and standardized test score files.

Earnings experience years

Experience is counted as any year that the individual is not fully enrolled in an HEI after having graduated high school. Earnings experience is set to zero for all years that a student is enrolled in an HEI.

Degree graduation percent

Degree graduation rates are constructed as the percent of enrollees who were freshmen in the years 2000-2005 who were recorded by the HEI as having graduated their degree.

Total expected tuition costs

Official reported tuition costs from Mineduc in the enrollment files for the 2013 academic year (in 2011 Chilean pesos) are used to construct expected tuition costs. The tuition is assumed constant for the expected degree length at current (2013) tuition levels. Total tuition costs are the present-value over the expected length of degree enrollment, discounted at 2% APR. The expected length of degree enrollment is calculated as the average enrollment length of time (in semesters) for past enrollees (from 2000-2005 freshman cohorts).[2]

Institution type

There are four types of HEIs in Chile. They are: Centros de Formación Técnica (technical institutions), Institutos Profesionales (professional institutions), Universidades Privadas (private universities), and

[2] We assume time of enrollment is 2013. Expected tuition costs are reported as the expected tuition costs for 2013 enrollees.

Consejo de Rectores de Las Universidades Chilenas (CRUCH universities). This variable is obtained from Mineduc enrollment data.

Average annual past enrollee earnings

Average annual past enrollee earnings are the average earnings by degree across earnings experience years 2-7, obtained from the tax authority (SII). The sample is all individuals in a base file that contains all high school graduates from 2002-2011, HEI enrollees from 2000-2013 and HEI graduates from 2000-2011. Construction of this base file is discussed below in section B.3.

Loan on-time repayment & default rates

Using loan origination information for 2006-2010, a determination is made whether or not a student should be in loan repayment. We start with the student's freshman year in the degree, add the official duration of the degree and a grace period of 1.5 years.[3] If this resulting year is any year prior to 2013 and the student has not re-enrolled in any HEI degree in the years 2011-2013, the student is included in our sample of those that should be in repayment as of 2013. Using the official sample of those that are actually in repayment, ~90% of those defined as "should be in repayment" are in repayment as of April 2013. On-time repayment rate by degree is defined as the percent of those that should be in repayment who are classified as being on-time in their loan payments. Default rate by degree is defined as the percent of those that should be in repayment who are classified as being at least three payments behind.

Selectivity tier

Each degree is assigned to one of five selectivity tiers. All technical degrees are defined in a single selectivity tier. All other degrees are classified based on median historical entrance exam scores for enrollees 2000-2013. The remaining four selectivity tiers are as follows: missing a median historical score or [0,450); [450,525); [525, 600); and [600,850].

Average value-added

Average value-added uses data from freshman enrolling cohorts from 2000 through 2005 and links it to tax return data for the tax years of 2005 through 2012.

Let earnings in year t for individual i who enrolled in degree j be a function of individual characteristics, the degree field of study and higher education institution they enrolled in and an idiosyncratic error term.

[3] This is the official grace period for all CAE loans. Law 20,027, as revised by Law 20,634 on 4 October 2012. Article 12.

$$(1) \qquad y_{it} = X_{it}\beta + Z_j\gamma + W_{itj}\delta + \mu_{jc} + \varepsilon_{it}$$

Where X_{it} includes baseline characteristics (socio-economic status, entrance exam scores, gender), years of labor market experience measured as number of years working but not enrolled in any higher education degree (which we measure using our comprehensive enrollment data), interactions between experience and baseline characteristics, and tax year dummies; Z_j are dummy variables for majors, and W_{ijt} are interactions between majors and baseline characteristics and majors and experience years, μ_{jc} is component of residual earnings that varies by degree (institution and major combination) and freshman cohort, and ε_{it} is a mean zero idiosyncratic error term.

The straight mean fixed effect (residual) is calculated from (1) for degree *j* over all cohorts 2000 to 2005. We estimate (1) by degree and selectivity tier securely within the tax authority in Chile, taking out parameter estimates and freshman-cohort-level mean residuals needed to estimate $\hat{\mu}_{jc}$. We then use the selectivity tier average X's for the freshman cohorts from 2000-2005 and the major-selectivity and field-selectivity specific regression estimates from (1) above, to predict degree value-added in experience years 2-7. We use the average degree-specific fixed effect estimated from 2000-2005 cohorts. We obtain the value-added estimate by taking the average predicted value-added across experience years 2-7.[4]

B.3. Student Panel Construction

We create a panel of data from the datasets and variables discussed in sections B.1. and B.2. We use student level high school graduation records for the years 2002-2012. We match this by student to enrollment data for the years 2000-2013, graduation data for the years 2000-2011. We define no-college individuals as those who graduated from high school in our records in years 2002-2012, but who never enroll or graduate from any higher education institution through 2013. We then match this data by individual with standardized entrance exam scores for the years 2000-2013 (entrance exam is PAA prior to 2004, and PSU thereafter).

At the degree level, we created a corrected and unified version of Mineduc's categorization of majors to categorize each degree into a major (e.g. the degree Social Work and Service at University A and Social Service at University B are both majors in Social Service) consistently across all years in the

[4] For specification checks we also calculated degree-specific mean effects using the methodology in Chetty et al. 2014. We did not find economically meaningful differences in results versus degree fixed effects.

data. We use data from our enrollment files on the tuition charged for each degree, the type of degree awarded (professional or technical), and the type of institution to categorize degrees and higher education institutions by major.

Most students enroll in one degree. Some students enroll in multiple distinct degrees successively (i.e. they switch majors or institutions through degree progression). For these students, we only attribute earnings to their most recently matriculated degree. If the student has more than one degree in the most recent year, we do not count their earnings towards either degree unless they are in the graduation data for one of the degrees (this affects less than 0.5 percent of our sample). If a student enrolls in multiple degrees, but graduates from only one, we only attribute earnings from the degree they graduated from.

We merge all of the above data at the individual level on to tax data within the tax authority (SII) for employment years 2005-2012 (tax filing years 2006-2013). From SII, we are able to extract degree average earnings by gender, socioeconomic status, and earnings experience year.[5] We also extract coefficients from regressions discussed above (in construction of "average value-added") and regressions discussed in detail in Appendix Section C.1.

[5] Per the data agreement with SII, we are only allowed to extract average earnings for cells of data in which there are at least 11 individual tax records.

C. Computing loan caps

C.1. Construct regression-adjusted average earnings

The base model used to calculate our version of earnings-based loan caps uses individual tax data from SII for all individuals in our base file of all high school graduates from 2002-2011, HEI enrollees from 2000-2013 and HEI graduates from 2000-2011. Per requirements from the Ministry of Education, if the individual has an income less than the full-time minimum wage in the tax data, the individual's observed wage is replaced with the full-time minimum wage before running programs to compute loan caps.

A "short-run" regression is used to estimate earnings using gender, socioeconomic status, earnings experience years 2-4 in the most recent four tax years (2009-2012), and degree fixed effects. A "long-run" regression is also estimated that includes 2-7 years of earnings experience and tax years 2005-2012, while still including gender and socioeconomic status controls. We ran these regressions at several concentric levels of degree-level comparison groups as follows: major-selectivity, major, narrow field-selectivity, and broad field-selectivity.

There are four components needed to calculate regression-adjusted average earnings: (1) observed average earnings by degree (2) degree composition of: experience year, tax year, gender, and socioeconomic status (3) regression coefficients on demographic and control variables and (4) a benchmark mean value of degree demographic composition (experience year, tax year, gender, and socioeconomic status).

We regression-adjust mean earnings for enrollees in degree *j* with 2-4 years of experience according to the following formula:

$$\hat{y}_j = \bar{y}_j - \hat{\beta}_{fs}\left(\bar{X}_j - \bar{X}_{fs}\right)$$

where $\bar{y}_j$ is the observed mean earnings for past enrollees of *j* with 2-4 years of experience within the past four tax years, $\bar{X}_j$ are the average characteristics those students. For the benchmark mean demographic composition, $\bar{X}_{fs}$, we use average characteristics of enrolling students in the same field of study and selectivity group as j. X_j include gender, SES (measured at the high school level), dummies for tax years, and dummies for years of labor market experience to control for cohort size fluctuations. $\hat{\beta}_{fs}$ are estimated effects of student characteristics on earnings using the long-run estimation panel of individuals with two to seven years of labor force experience for all tax years. This maximizes the sample used to identify the impacts of gender and SES on earnings by field and selectivity tier and controlling for tax year mean effects.

Degree average earnings

If a degree has at least 30 individual-year observations in our tax data, the average earnings from the "short-run" major-selectivity sample are reliable and are used as the base earnings that will be adjusted as discussed above. If not, the average earnings are obtained from the corresponding "long-run" model.

Degree composition

We use the same sample to calculate degree demographic composition as the sample used to calculate average earnings.

C.2. Calculating total earnings-based loan caps

The earnings-based loan cap for degree *j* is calculated as follows:

$$l_j = \rho \hat{g}_j = \rho \left(\sum_{t=1}^{15} \delta^t \hat{y}_{jt} - OC_j \right)$$

where ρ is the fraction of earnings gains reasonably dedicated to loan repayment and $\hat{g}_j$ is the estimated earnings gain from enrolling in degree *j* over not enrolling in college. Initially, ρ was set so that the overall amount of loans under the new loan cap system would equal the amount in the existing system given current enrollment (currently ρ is set to 11%). $\hat{g}_j$ is calculated as the present discounted value of earnings conditional on enrolling at degree *j*, $\hat{y}_{jt}$, over fifteen years of labor market participation, discounted at rate δ, less OC_j, a measure of forgone earnings for students enrolling in degree j. In practice, OC_j was set equal to the present discounted value of fifteen years of mean earnings for students who complete high school but do not enroll in college for every degree. Thus comparing across degrees, differences in regression-adjusted earnings determine differences in loan caps, ρ and OC_j simply determine the overall level of loans being given. For the small fraction of degrees where OC_j exceeded the present discounted value of earnings, the l_j were set to zero.

Earnings growth rate to project earnings

To scale earnings predictions to 15 years of earnings for the earnings-based loan cap calculation, we add to each short-run earnings prediction an expected earnings linear growth rate which we estimate from the tax data within the tax authority. Specifically, we estimate the following equation separately by major.

$$y_{it} = \alpha_j + X_{it}\beta + E_{it}\gamma + \varepsilon_{it} \tag{2}$$

Where α_j are degree-specific fixed effects, X_{it} includes tax year dummies and individual, time-invariant baseline characteristics (entrance exam score, gender and socio-economic status), and E_{it} is a linear term for years of experience. γ is the major-specific linear earnings growth rate used to construct earnings through 15 years of earnings experience. We chose a linear term after comparing models where we used dummy variables for experience years and determining that a linear growth term appeared reasonable for most degrees. Per Mineduc request, we constrained γ to be the same within a major and a degree-level – it did not vary with selectivity tier. More selective university degrees in civil engineering, for example, would have the same growth rate as less selective ones.

Calculate total earnings-based loan caps by degree

The regression-adjusted average earnings discussed in C.1. is used as the earnings for each labor market experience year 1-4. After experience year 4, earnings are grown linearly through experience year 15, using the major-specific linear growth rate (γ). After projecting earnings through experience year 15, the present-value is calculated by discounting annual earnings using a 2% annual rate (δ).[6]

We calculate the opportunity cost (OC_j) as 15 years of earnings for those in our base dataset who never attended an HEI. This is simply the average earnings from the "short-run" sample set to earnings for experience years 1-4, then projected using a linear growth rate (as discussed above) through experience year 15. The present-value is calculated by discounting annual earnings using δ. The final earnings-based loan cap is ρ times the value of the sum of the present-value for degree earnings less the sum of the present-value for no-college earnings.

[6] We assume a 2% APR because this is the interest rate charged on student loans in Chile.