Dynamic Earnings Simulation for Income-Contingent Student Loan Programs

Nora Delaney

Master's Thesis
Quantitative Methods in Social Sciences
Columbia University Graduate School of Arts and Sciences
December 27, 2019

https://github.com/n-delaney/QMSS-Thesis

Contents

1 Introduction

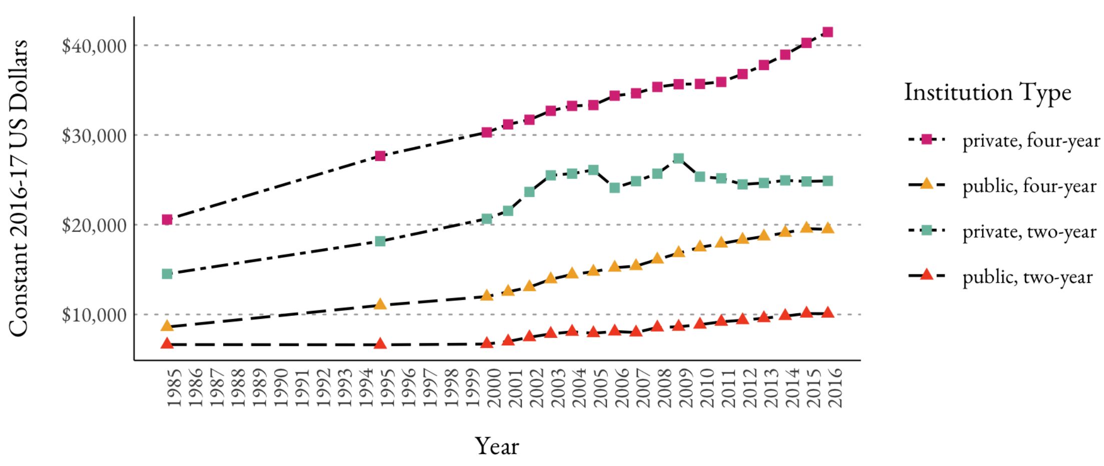
The goal of this thesis is to evaluate a hypothetical national-scale income-contingent student loan (ICL) program as an alternative to conventional mortgage-style federal student loans in the United States. First, I test whether revenues from the existing REPAYE income-based repayment plan for federal student loans would cover disbursed loan amounts if expanded as a mandatory policy for all borrowers seeking federal student loans. Subsequently, I propose alternative ICL policies to the REPAYE program that both satisfy or exceed revenue neutrality and carry more equitable outcomes for both high and low-earning borrowers. Income-contingent loans offer income-smoothing benefits to borrowers, who are exempted from repayment if their annual income falls below a certain threshold. While income-contingent loans may be favorable to borrowers experiencing periods of low income, the public cost of a national-scale income-contingent student lending program is a key consideration. Fiscal projections of an income-contingent student loan program rely on accurate, longitudinal measures of lifetime income. While revenues for conventional loans represent fixed payments with some variations in cases of deferral or default, repayment revenues drawn from income-contingent loans vary with borrowers' income. As a result, evaluating the viability of a federal income-contingent student loan program requires credible, nationally-representative, and longitudinal lifetime earnings data across a range of borrower profiles. In this thesis, I focus on both dynamic earnings simulations and their applications in designing a viable federal ICL program.

In section 2, I provide context on higher education financing in the United States and introduce existing empirical and theoretical work on student loan policy. In section 3, I discuss the two sources of income data I use for empirical modeling of income-contingent student loan programs, the National Longitudinal Survey of Youth and the Current Population Survey. I conduct the empirical modeling portion of this thesis in two parts. In section 4, I create a statistical model of dynamic earnings by fitting vine copulas to observed earnings rank trajectories in National Longitudinal Survey of Youth panel data. I use copula parameters to simulate unique earnings trajectories for 10,000 female and 10,000 male pseudo-individuals. In section 5, I use earnings simulations to numerically model three hypothetical ICL programs. Each program assumes mandatory enrollment in income-based repayment for students taking out federal student loans. I evaluate the distributional characteristics and public cost of each hypothetical ICL program as a replacement for current federal student loan programs. In in section 6, I present policy recommendations and opportunities for future research.

2 Student Loan Policy: Theory and Empirical Evidence

In the United States, student lending programs play a crucial role in financial decision-making surrounding college enrollment. Data from the United States Federal Reserve places public and privately sourced student debt at $1.6 trillion in 2019.[1] While federal grants are available to a small fraction of students, loans provided through William D. Ford Federal Direct Loan Program comprise the vast majority of federal student financing, including $91.4 million dollars in loans in 2018.[2] Eligible undergraduate students can access up to $12,500 in federal loans annually, which in some cases may be insufficient to cover the full cost of attendance. Tuition, fees, and room and board expenses have increased steadily at U.S. colleges over the past two decades (figure 1), accompanied by increases in student debt levels. Data provided by the Institute for College Access & Success indicates that individuals graduating with a 4-year degree in 2018 held an average of $29,650 in student debt relative to an average of $12,750 in 1996 after inflation adjustments (Gonzalez et al., 2019). Lochner and Monge-Naranjo (2016) find that increased labor market uncertainty and funding constraints are powerful barriers to enrollment as the cost of attendance rises, even as earnings returns to higher education have increased. Crucially, Bailey and Dynarski (2011) find that, between 1979 and 2000, increases in the rate of college enrollment among 19-year-olds cohorts in the United States have outpaced the rate at which the same students cohorts complete college by age 25, suggesting that a growing portion of US students may be carry student loans without the benefit of the higher earnings observed among college graduates. The authors find the discrepancy to be particularly pronounced among low-income and minority students. The complex student borrowing market and increasing levels of student debt held by US graduates calls for an examination of student lending programs internationally, in search of efficiency gains that may alleviate barriers to entry and stymied human capital development.

Figure 1: Average Cost of Full-time Attendance at U.S. Colleges
(Tuition, fees, room and board)



Source: U.S. Department of Education, National Center for Education Statistics

[1]Source: https://www.federalreserve.gov/releases/g19/current/
[2]Source: Department of Education Fiscal Year 2018 Annual Report on Federal Student Aid

2.1 Income-Contingent Student Loan Policies

In evaluating student finance, income-contingent loans (ICL) offer notable advantages over the conventional mortgage-style loans that currently dominate student lending in the United States. While conventional student loans require fixed monthly or quarterly payments regardless of the borrower's ability to pay, income-contingent loans collect a portion of the borrowers income if it falls above a certain threshold. This builds in income-smoothing and insurance components to the loan, reducing repayment burden as a proportion of income many cases. Low-earning borrowers are additionally protected from default during periods of financial stress, since loan repayment is not required when annual income falls below the repayment threshold.

In designing an ICL program, policymakers may tune a selection of loan characteristics, each impacting the degree of income smoothing and overall repayment amount for borrowers of various financial capacities:

- **Earnings threshold -** the level of income above which some portion of income is claimed towards loan repayment.
- **Repayment rate -** the percentage of an individual's income claimed towards loan repayment, which may be static or graduated based on the individual's income.
- **Interest rate -** the interest rate applied to the outstanding loan balance.
- **Repayment period -** the maximum period of time an individual is beholden to the loan scheme, after which the loan is typically forgiven.
- **Repayment cap -** the maximum dollar amount which, if met before the end of the repayment period, exempts the individual borrower from additional payments.

ICL programs that pool debt among borrowing cohorts additionally pool risk among individual borrowers, with high-earners taking on risk and expense on behalf of low-earners. This creates potential for both adverse selection, in terms of which individuals choose to enroll in an ICL, and moral hazard, as some borrowers may "game" ICL policy characteristics. These issues have been explored through theoretical work and have also become evident in early experiments with ICL policies, such as the Yale University lending program initiated in 1971. The short-lived program, designed by Milton Friedman in keeping with his theoretical work on the subject (Friedman, 1955), failed largely due to the school's inability to enforce repayment among participating graduates[3]. Borrowers who continued making payments to the program ended up paying down the majority of the debt, highlighting the potential for moral hazard and payment defection in poorly administered ICL programs. More recently, ICL programs in Australia and the United Kingdom both rely on federal tax authorities to enforce and administer payment collection, reducing opportunities for moral hazard among borrowers (Murphy et al., 2017; Chapman, 2005). Since the ICL policies proposed in this paper are assumed to be similarly administered by a central governmental authority, issues of moral hazard and payment enforcement will not be an active consideration in the empirical portion of the paper.

However, issues of adverse selection remain a key consideration in efficient ICL design. As Nerlove (1975)

[3]https://www.chronicle.com/article/The-Yale-Experiment/143185

points out, programs such as the Yale experiment rely on enrollment of both high-capacity and low-capacity earners to achieve revenue neutrality. Students who expect to receive large salaries after graduation may choose to attend a school that does not require ICL enrollment, leaving a school like Yale with primarily low-capacity borrowing students. Similarly, Hanushek et al. (2004) find that non-mandatory ICL programs will select for high-capacity students from poor backgrounds and low-capacity students from more privileged backgrounds, since high-capacity wealthy students will have the option to pay for higher education out of pocket or through private financing. As a result, poorer high-capacity students will subsidize their lower-capacity peers. For these reasons, the ICL programs I propose address adverse selection by requiring mandatory enrollment in income-based repayment for students taking out federal student loans.

2.2 Distributional Consequences of ICL Policies

Designing a student loan policy requires sensitivity to both public and private returns to higher education. Collectively, the public benefits from positive externalities associated larger numbers of college-educated healthcare practitioners, educators, technologists, and others. From the private perspective, higher education is linked to higher salaries among college-educated individuals. However, private returns are tempered by risks, such as labor market uncertainty, regional factors, and an evolving demand for certain types of skilled labor. Additionally, some careers requiring a college degree, such as teaching and other forms of public service, yield consistently low private returns. As a result, a successful public student loan policy should link the cost of higher education to associated private returns, while also tempering individual responsibility for repayment against labor market risk and low-paying public service careers.

From the perspective of public cost, public responsibility to pay for higher education should follow from public benefit associated with an educated and productive workforce. The identification and valuation of public benefits to education raise questions over whether tuition-free public higher education is favorable to student-financed alternatives. However, income-contingent lending policies compare favorably to ”free” public higher education, both in terms of overall access and progressivity. Resource constraints on free public universities have resulted in enrollment caps, admitting only the best-prepared and most qualified students (Chapman, 2005). Well-resourced students are more likely to attend college than less privileged peers and, in a public university system, a higher proportion of enrollment among middle- and high-income students has the potential to create a regressive distribution of tax-financed educational resources (Diris and Ooghe, 2018), especially if enrollment caps limit access to only the most elite students. Additionally, research on ICL programs in Australia and England indicates that, while students accumulate larger amounts of debt under ICL programs, enrollment rates among low-income students have remained stable relative to previous tuition-free higher education programs. This can be attributed to the reduction or elimination of liquidity constraints covering costs associated with college attendance (Chapman, 2005; Murphy et al., 2017).

Funding constraints play a crucial role in both college selection, enrollment, and graduation rates. Risk-averse students may be wary to take on excess financing when weighing enrollment or choosing between colleges, potentially with profound effects on future earnings. Chetty et al., (2017) demonstrate that not only a college degree but also institution of enrollment play an outsized role in predicting graduate earnings relative to parental earnings; graduates hailing from the bottom quintile of parental income and graduating from top-tier schools earn comparably to wealthier peers (Chetty et al., 2017). This suggests that enrolling students' decisions between colleges carries profound implications if systematically influenced by funding constraints. As a result, income-contingent student loans have the potential to facilitate enrollment by expanding access to low-risk funding.

2.3 Student Loan Policies in the United States

In the United States, an income-contingent loan policy may serve to address issues of risk aversion for individuals weighing the cost of college attendance against adverse outcomes associated with student loan repayment, such as default and excess repayment burden during periods of financial hardship. Undergraduate students taking out federal student loans under any repayment plan may borrow up to \$9,500 in their first year of enrollment, \$10,500 for the second year, and \$12,500 for subsequent years. An individual student can borrow up to a total amount \$57,500 over their entire period of study. Students who demonstrate sufficient financial need may borrow limited quantities under subsidized loans that do not accrue interest during the period of study or in the six months following graduation. The loan balances of students who do not qualify for subsidized student loans accumulate interest throughout the enrollment period. The majority of student loans currently available in the United States are conventional, time-based repayment loans. Students borrowing through the Stafford Loan Program may select from several repayment options. Loans repaid under the standard repayment plan require fixed monthly payments, with minimum monthly payment of \$50 or accrued interest. Borrowers may choose to enroll in several alternatives to the standard repayment plan, including an income-based repayment plan.

Income-based repayment programs for federal student loan debt were initially introduced as part of the College Cost Reduction and Access Act passed by congress in 2007 as a first step towards reducing the default risk and repayment burden associated with fixed repayment schemes. The Obama administration updated the terms of the income-based repayment plan, creating a repayment option known as the Revised Pay as You Earn plan (REPAYE), which remains available today. The existing REPAYE program is popular, with more than a quarter of U.S. student borrowers currently enrolled in an income-contingent repayment plan (Britton et al., 2019). Individuals enrolling in the income-based repayment plan after 2014 pay 10% of their income above a repayment threshold. During the repayment period, borrowers earning less than 150% of the federal poverty line, \$18,210 for a single-person household in 2018, are exempt from making loan repayments. Borrowers earning above that threshold pay 10% of their discretionary earnings, defined as the difference between after-tax income and the repayment threshold. Borrowers continue to make monthly payments relative to their income throughout the REPAYE repayment window, regardless of their outstanding loan balance. The repayment period is 20 years for borrowers financing undergraduate degrees and 25 years for those enrolling in graduate studies. After the repayment period ends, borrowers who have not paid down their full loan balance (and accruing interest) are eligible for loan forgiveness, although they must pay income tax on the loan balance forgiven.

The empirical portion of this thesis uses the REPAYE plan as a basis for a national-scale policy simulation. In this policy simulation, income-based repayment would become a mandatory feature of all federal student loan programs. Currently, enrollment in the REPAYE plan is optional and likely appeals primarily to borrowers in low-wage jobs, such as teachers, or who are otherwise uncertain of their earnings prospects. Additionally, borrowers may shift between repayment options over the course of repaying a single loan and must re-certify enrollment in income-based repayment each year. Under the current REPAYE policy, individuals who observe that their income-based payments will exceed the fixed amount associated with conventional repayment can be predicted to exit the income-based repayment option. This raises concern over adverse selection, with predominantly low-earning student-borrowers subscribing to the income-based repayment plan. Conversely, an ICL program in England integrates income-based repayment as a universal feature of federal student loans, ensuring that borrowing students with a range of professional interests and future incomes participate in the program. Mandatory enrollment in income-based repayment for all borrowing students reduces concerns over adverse selection among borrowing students. Additionally, pooling high and low-earnings borrowers creates the potential for a revenue neutral federal ICL program that offers income-smoothing benefits to borrowers experiencing financial hardship. Crucially, a revenue neutral ICL program ensures a progressive distribution of the cost of higher education among the individuals (college graduates) who benefit from it, without requiring tax revenues drawn from the larger population to subsidize the costs of federal student loan financing.

2.4 Dynamic Income Modeling

Existing work to simulate dynamic income for use in student loan policy modeling has taken several approaches. Higgins and Sinning (2013) have demonstrated the success of multivariate regression analysis using a range of demographic indicators and individual characteristics in predicting observed income dynamics in Australia. Unconditional quantile regression has been used to predict earnings as a function of age among college graduates in China (Cai et al., 2019). However, while multivariate regression analysis has the potential to capture the complexities relationships of individual characteristics and predicted earnings, regression coefficients (which remain static) may not capture evolving relationships between covariates as labor markets and other factors change over time throughout the period of income projections. As a result, the appropriateness of predicted future income dynamics over the span of a proposed student loan policy break down as observed earnings dynamics may diverge from static regression coefficients. Regression-based approaches also have severe limitations in simulating income data, since income simulations using this approach will also require simulated covariate values to plug into regression coefficients. Additionally, regression-based approaches capture the central tendencies of earnings or income, whether at the mean, median, or quantile. While such measures are suitable for predicting earnings on average, they all not well-suited to measuring dynamic transitions in earnings across individual earnings trajectories (which rarely intersect with the central tendency) over time.

Some approaches to dynamic income modeling retain regression based approaches but add a stochastic component to capture movement in individual earnings from one period to the next. For example, Bonhomme and Robin (2009) combine a fixed-effects regression with transitory component drawn from copulas to fit panel earnings data observed in the French Labor Force Survey. In doing so, they introduce heterogeneity into earnings modeling, allowing for greater insight into labor force mobility at the individual level rather than at the central tendency. Similarly, Geweke and Keane (2000) model dynamic earnings among male participants in the U.S. Panel Survey of Income Dynamics as a function of demographic factors, unobserved characteristics, and a randomly distributed shock. In this case, by decomposing variations in individual earnings dynamics over time, the authors demonstrate that transitory shocks account for most variation in earnings in a given year, after controlling for race, education, and age. However, the authors find that unobserved individual heterogeneity plays the largest role in determining lifetime earnings. It follows that, for predictive purposes, it would be appropriate to model dynamic earnings as a series of transitory shocks fitted to individual-specific earnings trajectories.

In this vein, the use of copulas to capture income volatility as a function of age and sex has also been successful in simulating observed earnings, carrying some advantages over regression and mixed-regression modeling approaches. While regression-based models may be combined with dynamic components, these models are ill-suited to simulation, since they require credible simulated coefficient inputs of demographic and individual characteristics. By contrast, copulas may be fit to samples of panel earnings disaggregated by sex, race, education, or other demographic characteristics as a means of controlling for major drivers of earnings without creating the need for simulated input data. For purposes of modeling income-contingent student loan programs, several papers have demonstrated that dynamic income simulations generated by copula models can effectively model the number of borrowers experiencing excess repayment burden in a given period of loan repayment (Dearden, 2019; Barr et al., 2019; Armstrong et al., 2019). Since copula modeling of dynamic income introduces a stochastic element to income predicted on the basis of sex and age, dynamic income simulations offer a high granularity of insight into the proportion of individuals falling below an income threshold in an income-contingent student loan policy. This approach allows for credible estimates of low-earnings periods, even for individuals who may be at or above average in terms of total lifetime earnings. Since this better accounts for individual movement in income or earnings aroung an ICL repayment threshold, Dearden (2019) shows that this allows for modeling of the insurance and consumption-smoothing characteristics of an income-contingent loan policy at the level of individual observations, as well as a more accurate prediction of the degree public subsidy at the aggregate level.

Since the purpose of my research is to simulate an ICL program, I select the a pure copula model as the best approach for generating credible earnings simulations. By using copula modeling, I measure observed transitions in earnings ranks from ages 22 to 53 in National Longitudinal Survey of Youth panel data and then use the copulas to simulate artificial earnings trajectories. By dis-aggregating the panel data by age, sex, and college degree, I control for these demographic factors in fitting the copula models, yielding Student-t distributions fitted to transitory movements in earnings between adjacent ages. Resulting simulations yield unique trajectories for each pseudo-individual, with movement at each age transition clustered around (but not constrained to) a central tendency measure.

3 Data Sources

The availability and characteristics of earnings data play a crucial role in the estimation of a dynamic earnings model, since longitudinal earnings trajectories at the individual level are of primary interest. In particular, panel data must contain a minimum of two observations of earnings per individual to begin modeling an individual's earnings trajectory over time. Sample sizes at each age must also be sufficiently large to ensure the robustness of model fit. For this reason, I use earnings data from two publicly-available data sets to fit and scale dynamic earnings simulations. First, I fit dynamic earnings model to data observed in the National Longitudinal Survey of Youth (NLSY), which contains up to 21 observations of earnings per individual. Then, I control for cohort effects by scaling simulated dynamic earnings ranks to cross-sectional earnings in the Current Population Survey (CPS).

3.1 National Longitudinal Survey of Youth

The National Longitudinal Survey of Youth (NLSY) is conducted by the U.S. Bureau of Labor Statistics and contains information on a broad range of sociological and economic variables. I concatenate data on college graduates from the 1979 and 1997 NLSY cohorts to measure earnings trajectories over ages 22 to 53. Table 1 shows sample size and characteristics of each cohort. Due to limitations in the construction of a comprehensive measure of total income using NLSY data, I focus on the study's records of wage and salary earnings. While an income-contingent repayment program will be best modeled on a measure of total income, earnings represent a conservative substitute for total income. Primarily, using earnings in place of total income will necessarily understate income-based repayments under an ICL program, leading to a conservative estimate of an ICL policy's public cost. Additionally, since a relatively small subset of individuals enjoy substantial income outside of wage and salary earnings, tailoring an ICL model to income volatility among individuals vulnerable periods of low or nonexistent earnings is of principal importance. For these reasons, I choose NLSY data for its longitudinal insights into earnings patterns over time and use the earnings variable the study provides.

Because NLSY follows individuals longitudinally, it offers relatively long panels of individual-specific earnings trajectories. However, cohort effects within the NLSY sample constrain the generalizability of NLSY earnings data. Figure 2 shows cohort ages where at least some individuals experience economic recessions. Male earnings in the 1979 NLSY cohort show sensitivity to the Great Recession from 2008 to 2009, when cohort members ranged between 44 and 50 years of age. Male earnings at ages 44 through 50 dip and then continue increasing in each quintile, in contrast to smooth parabolic earnings trajectories observed in cross-sectional data shown in figure 3. Similarly, is possible that a recessionary cycles in 1980-1982 and 2008-2009 impact early career earnings for the 1979 and 1997 NLSY cohorts respectively. To address cohort effects in the NLSY data, I model earnings volatility at a given age as a function of earnings rank. Optimistically, measuring earnings as ranks instead of dollar amounts may remove some bias attributable to recession. If recessions are assumed to impact all cohort members similarly, then volatility in dollar earnings should not appear in earnings ranks at a given age. However, if recessions impact individuals in various earnings rank segments differently, it is possible that bias remains in the NLSY earnings rank data. Since I aim to create a conservative models of ICL policies, overestimates of early career earnings volatility in the NLSY data may still be appropriate for use in modeling ICL policies that are robust to recessionary or otherwise highly volatile earnings in early career repayment. Finally, while a statistical model of earnings dynamics is fit using NLSY data, simulated earnings ranks are scaled using contemporary cross-sectional data from the Current Population Survey, ensuring the highest degree of representativeness at the population level.

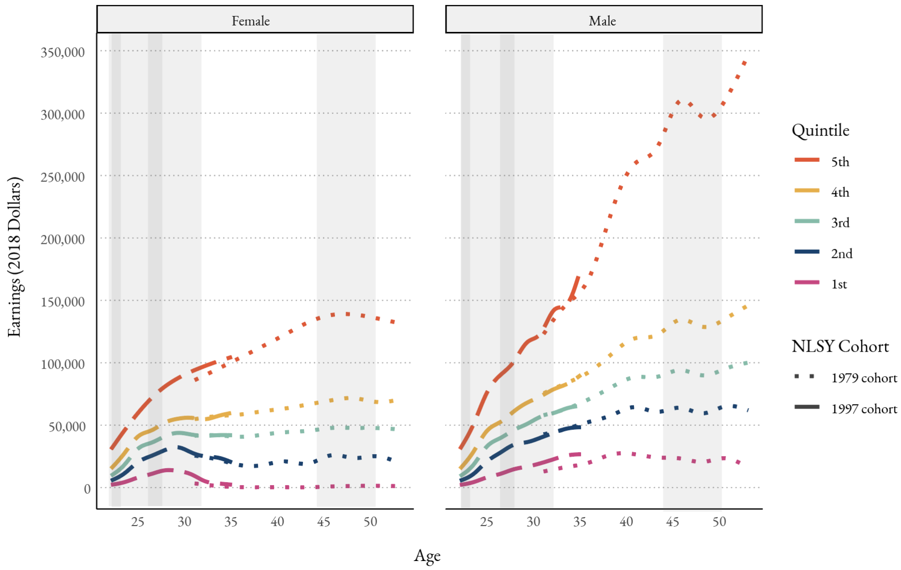
Table 1: NLSY Cohort Characteristics

	NLSY 1979	NLSY 1997
Study participants	12,686	9,000
Birth years	1958 - 1964	1981 - 1984
Age at last observation	51 - 57	31 - 34
Max. earnings observations	21	12

3.2 Current Population Survey

The Current Population Survey (CPS) is jointly sponsored by the US Census Bureau and the US Bureau of Labor Statistics. Roughly 60,000 households throughout the United States participate in the CPS and are tracked over a 16-month interval. Respondents surveyed in March participate in the the Annual Social and Economic Supplement (ASEC), a larger and overlapping study investigating income, educational attainment,

Figure 2: NLSY Cohorts: Earnings and Recessions



Shaded areas represent ages where some cohort members (in staggered birth years) experience recessions.

workforce specialization, and related fields (Flood and Pacas, 2016). I make use of publicly available ASEC samples accessed through the Integrated Public Use Microdata Series (IPUMS) database. While CPS data offers up to two panel observations per individual, earnings panels constructed from CPS data yield lower sample sizes by age and only two earnings observations, relative to panels of up to 21 observations for NLSY data.

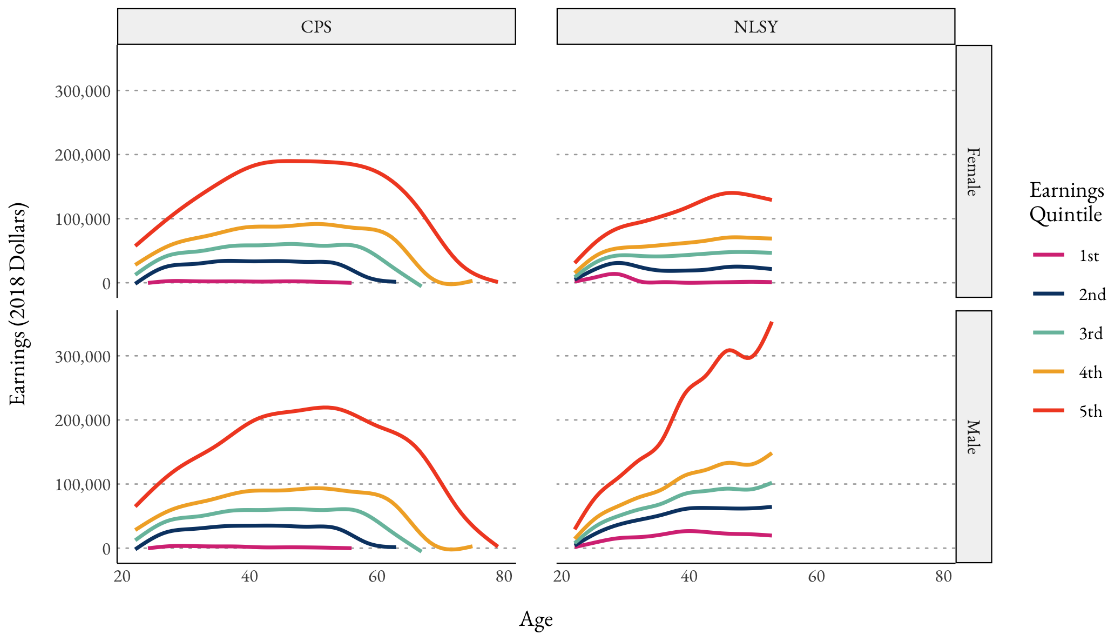
Because CPS data is nationally representative and cross-sectional, it is appropriate for use in scaling earnings rank simulations discussed in Section 4.4. A comparison of CPS and NLSY earnings data is shown in figure 3, indicating that CPS does not suffer from the same cohort-effects as NLSY data. I select a sample of 160,574 college graduates aged 22 to 79 reporting earnings in 2015, 2016, 2017, and 2018. All earnings data is scaled to 2018 dollars.

4 Copula Modeling

Copulas introduce flexibility into multivariate dependency modeling, allowing for a multivariate joint distribution to be described by uniform marginal distributions and a copula function (Demarta and McNeil, 2005). As a result, I use copulas to model dependencies between income observed over adjacent periods. For continuous random variables X_1, X_2 with joint and marginal distributions,

$$F(X_1,\ X_2) = P(X_1 \leq x_1,\ X_2 \leq x_2)$$

Figure 3: Comparing Cross-Sectional and Panel Earnings Data



$$F(X_j) = P(X_j \leq x_j), \quad j = 1,\ 2$$

Sklar's Theorem (Sklar, 1959) states that there exists a unique copula, $C : [0,1]^2$, such that

$$F(X_1,\ X_2) = C(F(X_1),\ F(X_2))$$

In order to measure patterns of earnings volatility as individuals age, I fit Student-t vine copulas to short panels of observed earnings data at three adjacent ages, measuring earnings rank dependence across ages 22 to 53. I then use fitted copulas parameters to generate earnings rank trajectories for pseudo-samples of 10,000 men and women. Simulated earnings ranks are then scaled to observed earnings data, producing simulated earnings in dollar terms. The resulting sample of 20,000 earnings trajectories is sufficiently large for use in numerical models of ICL policy scenarios discussed in section 5. My workflow follows the steps described in sections 4.1, 4.2, 4.3, and 4.4.

4.1 Transforming Earnings Panels

Because copulas are modeled as a function of uniform marginals, copula models estimating earnings volatility rely on measures of earnings distributed $U[0,1]$ rather than dollar values. Additionally, copulas cannot be fit to missing data where an individual may be ranked in one period but not the next. For example, an individual with income recorded at age 25 but not at age 26 may not be included in the copula model. I use two strategies to address missing data and ensure that copula estimates satisfy both internal validity relative to the NLSY cohorts and external validity at the population level.

First, I rank individuals' earnings based on the complete sample of individuals at a given age, including those who will not ultimately be included in the data inputted to the copula model. Earnings ranks are assigned using a weighted empirical cumulative distribution function, which assigns earnings ranks weighted by a measure of how representative each individual is at the population level. As a result, individuals whose earnings ranks are included in the copula model are ranked relative to the full sample of their peers at a given age, both within their NLSY cohort and relative to the population. Second, I choose a three-observation earnings panel as the basis for the vine copula model. In theory, a longer panel should yield greater certainty over measures of earnings volatility at a given age within the panel, and the NLSY 1979 and NLSY 1997 cohorts contain up to 21 and 12 earnings observations per individual respectively. However, few individuals report earnings over the entire timeline of the NLSY study due to attrition. As a result, a trade-off emerges between sample size of individuals included in a given panel and the length of that same panel. The three-observation earnings rank panels I use extend beyond the two-observation panel used in previous dynamic earnings modeling by Dearden (2019) and Barr et al. (2019), allowing me to reduce bias due to unobserved individual-level latent variables that may also influence earnings volatility across the earnings panel.

To create earnings panels, I dis-aggregate NLSY data by sex and arrange male and female subsets into 32 three-observation panels, each containing earnings rank data for three adjacent ages. This models two "age transitions" for each individual in the panel, as individuals' earnings ranks shift from one age to the next. Individuals with one or more instances of missing earnings data are omitted from the panel. Each earnings panel is centered at an age between 23 and 52. For example, the panel centered at age 27 contains earnings data for individuals aged 26, 27, and 28. However, panels centered at ages 32 to 53 differ slightly. NLSY records earnings every year for individuals aged 22 to 31 in the 1997 cohort, and every other year for individuals aged 32 and up in both the 1979 and 1997 cohorts. As a result, adjacent ages for panels centered on ages 32 and above span a 5-year period. For example, the panel centered at age 44 contains earnings ranks at age 42, 44, and 46. Despite the staggering of earnings every other year, panel samples are generated for each age between 23 and 53 due to varying birth years within each cohort. Earnings panels centered at ages 53 and above cannot be constructed due to the maximum age of individuals in the NLSY 1979 cohort, who were born in the years 1958 to 1964. For computational reasons, assigned earnings ranks are further processed using the *pobs()* function in the VineCopula package to ensure earnings ranks are continuous and unique on the interval $[0, 1]$. Sample sizes for each age transition are listed in table 2.

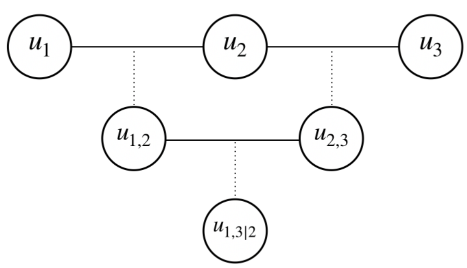
4.2 Vine Copula Structure and Family

For each three-observation panel, I fit a vine copula to model the multivariate distributions as a function of three bi-variate copulas. As a result, I create 64 vine copula models (32 male and 32 female) corresponding to age transitions from age 22 to 53, each composed of three bi-variate copulas. For each vine copula, the uniformly distributed set of earnings ranks, u_1, u_2, and u_3 correspond to age_1, age_2, and age_3. Bi-variate copulas are fitted to the joint distributions of u_1, u_2 and u_2, u_3, denoted $C(u_2,\ u_1)$ and $C(u_3,\ u_2)$ respectively. To complete the multivariate dependency structure, $C(u_1,\ u_3 \mid u_2)$ is fitted to the joint distribution of $C(u_2,\ u_1)$ and $C(u_3,\ u_2)$. The resulting dependency structure is translated to matrix form below and visualized in figure 4.

Next, I determine the copula family that best fits the dependency structure of earnings rank transitions by age. Using the vine copula matrix shown in figure 4, I fit range of candidate bi-variate copula structures to pairs of uniform marginals within each age transition panel. The best fit is returned based on values of Akaike Information Criterion (AIC) for each copula. Table 3 shows tallies of the most commonly assigned copula families. Since Student-t copulas are the most commonly fit copula type, I use this copula family throughout all subsequent copula modeling in this paper. Choosing a single copula family for use across all copula fits allows for straightforward interpretation of and simulation from fitted parameter estimates.

Student-t copulas take parameters rho, ρ, and degrees of freedom, ν. Student-t copulas are appropriate for modeling rank dependence in earnings, since ρ offers a measure of correlation between earnings rank at adjacent ages while ν allows for flexible modeling of tail dependence. Allowing ν to vary enables the model to capture the volume of individuals experiencing high earnings rank volatility at younger ages and at other times when earnings and career prospects may be uncertain.

Figure 4: Vine Copula Structure



$$= \begin{bmatrix} 1 & 0 & 0 \\ 3 & 2 & 0 \\ 2 & 3 & 3 \end{bmatrix}$$

u_1, u_2, and u_3 represent earnings rank at adjacent ages 1, 2, and 3

Table 2: NLSY Sample Sizes by Age Transition

Age	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Female	885	840	931	986	985	1022	818	829	577	1317	1119	989	788	652	476	412
Male	643	655	711	763	761	763	626	617	488	1055	910	810	635	560	418	341
Age	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53
Female	374	406	380	412	371	410	377	414	379	409	294	309	190	184	82	93
Male	333	330	329	331	331	332	335	339	338	339	260	233	166	139	78	73

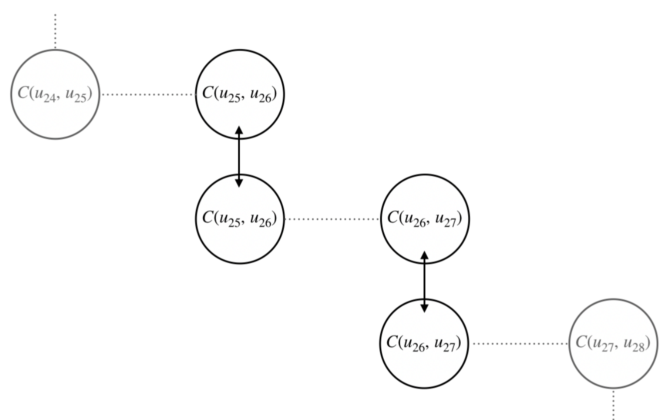
Table 3: Counts of Most Common Copula Families for NLSY Data

Copula Family	Female	Male
Student-t copula	34	28
Rotated Tawn type 1 copula (180 degrees)	19	2
BB8 copula	8	11
Frank copula	2	11
Tawn type 1 copula	8	4
Rotated Tawn type 2 copula (180 degrees)	6	5

4.3 Student-t vine Copula Estimation

For each of the 32 earnings panels for men and women respectively, I fit a vine copula model consisting of three bi-variate Student-t copulas. Copula parameters are estimated using maximum likelihood estimation.[4] Fitted values of bi-variate Student-t parameters are shown in figure 6. Due to the three-observation panel structure, vine copula models for ages 24 through 52 contain bi-variate copulas corresponding to overlapping age transitions. For example, the vine copula fitted to the age transition centered at age 25 contains $C(u_{26},\ u_{27})_1$, while the following age transition centered at age 26 also contains a copula modeling the same ages, $C(u_{26},\ u_{27})_2$. To control for panel-specific effects, I include parameter estimates for both overlapping copulas $C(u_1,\ u_2)$ and $C(u_2,\ u_3)$ when possible. Raw parameter estimates are smoothed using loess regression.

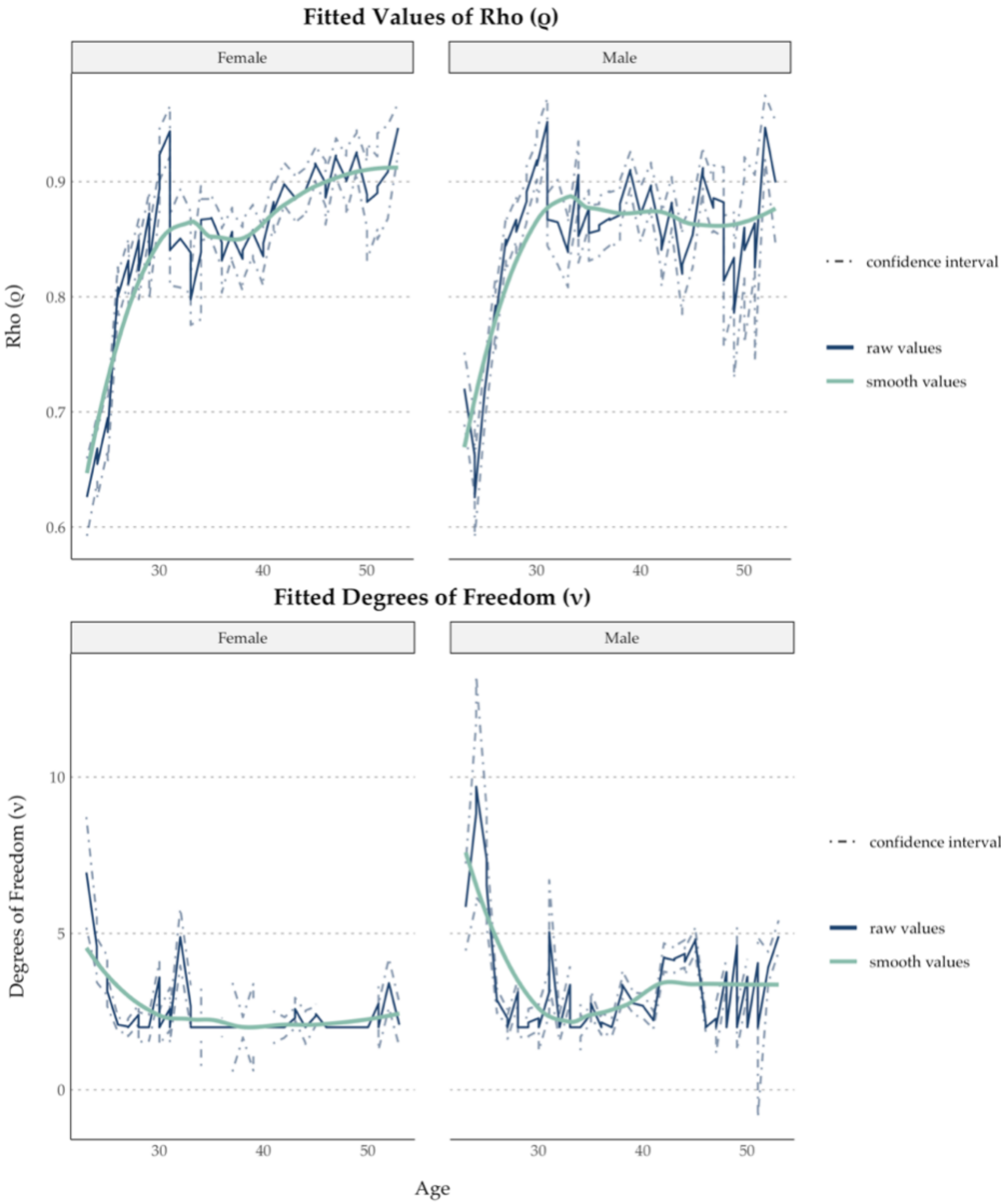
Figure 5: Overlapping Copulas



For both male and female observations, ρ increases substantially in the years following graduation. This is attributable to increasing earnings rank correlation as college graduates establish career trajectories. Increasing rank correlation in female earnings stalls around age 30, potentially due to child-rearing responsibilities, but continues to increase after age 40. A similar trend is not apparent in measures of ρ for male observations

[4] I use the RVineCopSelect() function in VineCopula to create a Student-t RVineMatrix object containing earnings rank data and model hyperparameters. I then use the RVineMatrix to seed the final copula fits using the RVineSeqEst() function.

Figure 6: Student-t Copula Parameters



at the same ages, since male estimates of ρ stabilize after age 30. Degrees of freedom, ν, are highest at ages directly following graduation for both male and female observations, suggesting a large volume of movement between upper and lower earnings ranks early in college graduates' careers. Degrees of freedom remain low throughout female earnings trajectories with a notable spike around age 30, likely corresponding to increased child-bearing frequency. Smoothed estimates of male degrees of freedom increase after age 40, suggesting greater tail movement later in male earnings trajectories.

4.4 Lifetime Earnings Simulation

The goal of lifetime earnings simulation is to generate an earnings sample large enough to evaluate individual repayment burden and public cost of income-contingent loan scenarios. Copula parameters determine patterns of earnings rank volatility at each age, allowing me to scale earnings dynamics observed among a limited number of participants in the NLSY study to much larger pseudo-sample of 20,000 individuals at ages 22 to 79. I simulate earnings samples following the workflow outlined in Dearden (2019). All simulations are conducted with smoothed parameter values, and earnings ranks are generated recursively with age-specific parameters according to the steps outlined below. Because the earnings simulation process contains a stochastic element, each simulated earnings trajectory is unique and follows a jagged path mimicking observed individual earnings trajectories. A sample simulated earnings trajectory is shown in figure 7.

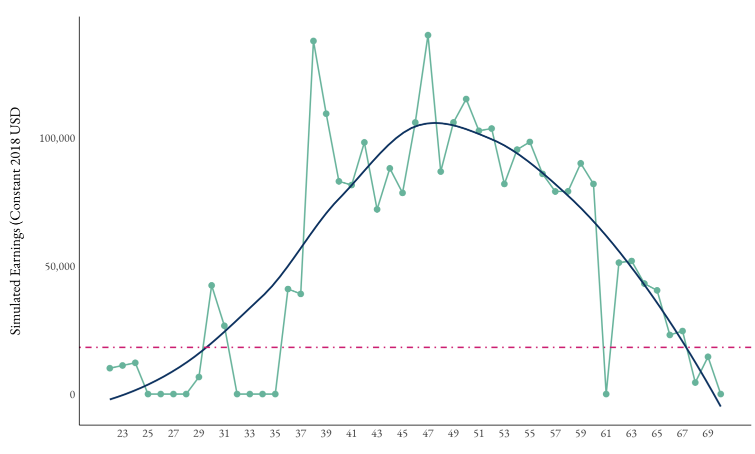
1. Create samples of 10,000 female and 10,000 male 22-year-old pseudo-individuals with uniformly assigned earnings ranks on the interval $(0, 1]$. Each pseudo-individual is assigned a unique identifier.

2. Generate an earnings rank for each pseudo individual at age 23 using a conditional draw from $C(u_{23}|u_{22})$. This is done using the *BiCopCondSim()* function.

3. Repeat step 2 recursively for ages 24 and above, generating an earnings rank at age i based on a conditional draw from $C(u_i|u_{i-1})$. Because of the limited age of individuals participating in the 1979 NLSY cohort, I simulate earnings ranks for individuals aged 54 and above using copula parameters from $C(u_{53}|u_{51})$.

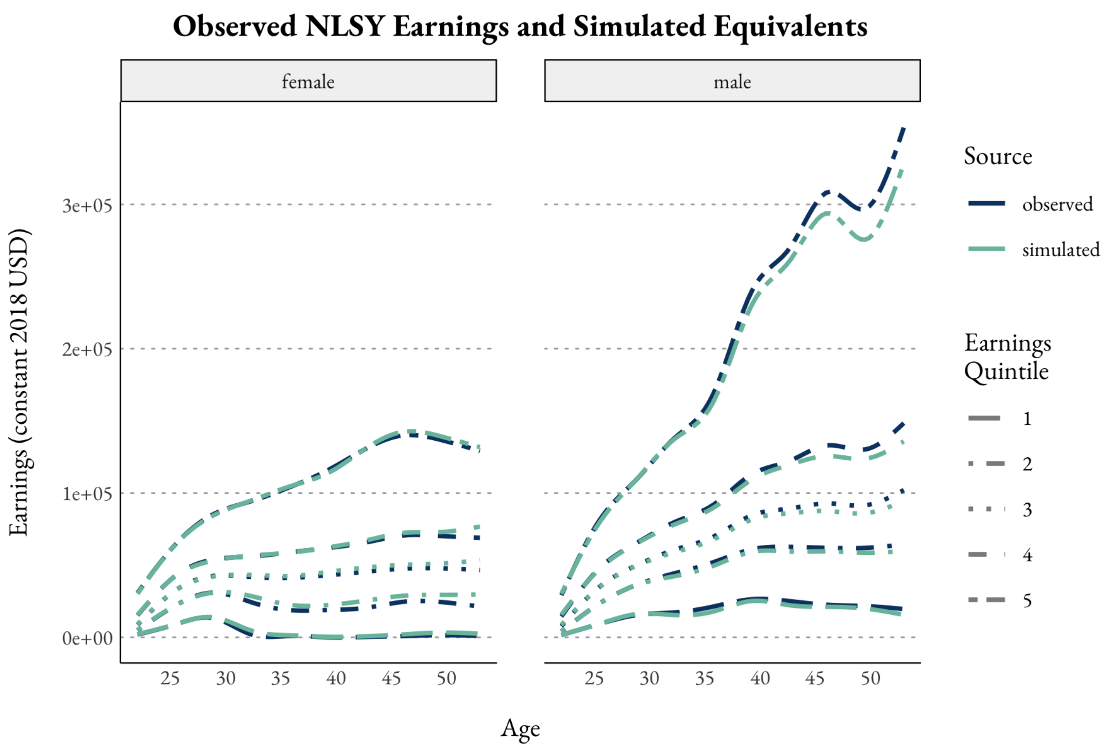
4. Scale earnings ranks to dollar values by applying the quantile() function to CPS earnings data at each age between 22 and 79.

Figure 7: Sample Simulated Female Earnings Path for Pseudo-individual f_4624



To evaluate the fit of simulated earnings, I scale simulated earnings ranks to NLSY data and over-plot simulated and observed earnings quintiles. As figure 8 shows, simulated and observed earnings are very similar. Some slight deviation of simulated earnings from observed earnings at later ages of males in the fourth and fifth earnings quintiles is likely due to relatively small sample sizes at these ages. However, since simulations underestimate earnings relative to observed values, simulations at these ages and quintiles will result in underestimates (rather than overestimates) of public subsidy for ICL programs. Note that figure 8 is used only for model evaluation purposes and that earnings simulations used in the Policy Modeling section use CPS-scaled values of dollar earnings.

Figure 8: Simulated and Observed NLSY Earnings



u_1, u_2, and u_3 represent earnings rank at adjacent ages 1, 2, and 3

5 Income-Contingent Policy Modeling

I use the generated sample of 20,000 simulated pseudo-individuals to evaluate a national-scale income-contingent student loan policy for the United States, as an alternative to existing federal student loan programs. In doing so, I aim to create a policy that is both self-financing and forgiving to student borrowers experiencing financial hardship. In this section, I first examine the characteristics of an ICL program similar to the existing REPAYE income-based repayment plan for US federal student loans, discussed in section 2.3. Finding that revenues collected in the modified REPAYE policy scenario far exceed the value of original loan disbursements, I proceed to propose two addition self-financing ICL programs that address undesirable distributional characteristics of the modified REPAYE program.

ICL policy design should be sensitive to both public cost and distributional characteristics, which tune the relative size of individuals' repayment contributions among a pool of borrowers. I use a self-financing constraint as a measure of the public cost of ICL programs. For my purposes, I define a self-financing program as one that satisfies revenue neutrality, with the present value of revenues precisely equal to loans disbursed, or otherwise exceeds revenue neutrality, with the present value of revenues exceeding the amount of loans disbursed. In addition, I calculate the internal rate of return (IRR) of a given loan program as a method of evaluating the relative costs of proposed ICL programs. To assess the distributional characteristics of ICL programs, I tally the frequency of individuals in non-payment across the repayment period and additionally evaluate total repayment by decile of lifetime earnings.

5.1 Modified REPAYE Program

The first hypothetical ICL policy I evaluate is an expansion of the existing REPAYE repayment plan for federal student loans. Evaluating an extension of an existing income-based repayment program provides a baseline for the public cost and distributional characteristics of a national-scale ICL program. I adapt the REPAYE program as a mandatory national-scale student loan program and simulate the policy using the earnings of the 20,000 pseudo-individuals generated in section 4. I make several conservative assumptions to simplify the policy model and ensure that it is generalizable. First, I assume that all students financing higher education with federal student loans must enroll in the modified REPAYE program. This addresses issues of adverse selection, since individuals are enrolled in the program regardless of earnings potential or planned career paths. None of the borrowers receive interest subsidies under this model, and all borrowers are treated as single-person households for tax purposes. Additionally, I assume that all students graduate after four years and take out the maximum loan amount for each year of enrollment, amounting to $45,000 in financing. Under the modified REPAYE program, borrowing students graduate with loans totalling $50,024 due to accruing interest over a four-year period of study.

Additionally, I remove the requirement that borrowers pay income tax on forgiven loan balances at the end of the repayment period; instead all revenues associated with the model are drawn from income-contingent repayments. This locks in the insurance component of the income-based repayment scheme, since borrowers need not worry that they will be taxed on loan balances accruing interest over the repayment period. To highlight an extreme case of taxed loan forgiveness under the current REPAYE program, an individual finishing college with a loan amount of $50,024 but always falling below the repayment threshold would have a loan balance of $120,644 forgiven after the 20-year REPAYE repayment period. According to the 2018 federal income tax schedule, this individual would be responsible for paying $28,954, amounting to 24% of the loan balance and more than half of the original loan amount, in income taxes. Eliminating the tax on loan forgiveness ensures that low-income borrowers are not disproportionately penalized at the conclusion of the repayment period.

The characteristics of the modified REPAYE program are as follows:

- Cohort size: 20,000 borrowers
- Loan balance at graduation: $50,024
- Repayment threshold: $18,210 (150% of the federal poverty line for an individual)
- Interest rate: 4.5%
- Repayment rate: 10%
- Repayment period: 20 years
- Repayment cap: none

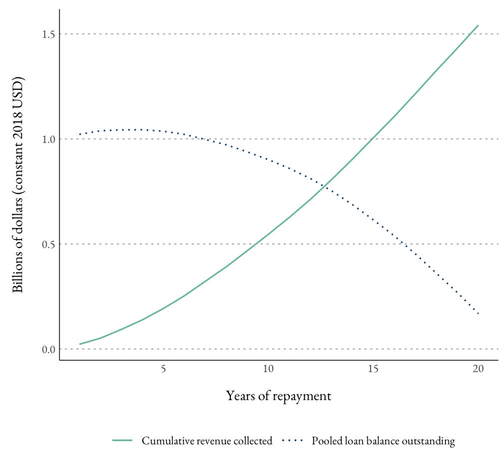
5.1.1 Simulating the Modified REPAYE Program

To simulate the modified REPAYE policy, I model the loan repayment process using the simulated cohort of 10,000 female and 10,000 male college graduates. The total sample size, 20,000, is sufficiently large to mimic the financial performance of a national ICL program, since estimates of revenues and measures are moderated by a large number of individual earnings trajectories. To run the policy simulation, I assume that all borrowers take out the maximum federal loan amounts, ranging from $9,500 to $12,500 annually, over a four-year period of study culminating in graduation. For the cohort of 20,000 borrowers, this results in a total loan disbursement of 900 million USD. Borrowers begin repayment at age 23. I first calculate federal income taxes for each yearly earnings observation. Then I multiply the difference between post-tax earnings and the repayment threshold by 10% to find the annual repayment amount for each borrower at each age throughout the repayment period. Repayment dues for individuals with earnings below $18,210 are set to zero. To ensure accurate accrual of interest, I divide each annual repayment amount into equal monthly payments before calculating interest and outstanding loan balances. Since the REPAYE plan does not include a repayment cap, all individuals continue payments throughout the 20-year repayment period, subject to the income threshold. All calculations use constant 2018 US Dollars.

5.1.2 Evaluating the Modified REPAYE Program

The modified REPAYE program sheds light on opportunities for innovation in crafting a national-scale ICL program. From the perspective of public cost, the program is self-financing, carrying an internal rate of return (IRR) of 4.76%. Total repayment revenue of 1,542 million USD exceeds 900 million USD in disbursed loans. However, the pooled fund of student loans under the modified REPAYE program carries a positive outstanding balance of 168.2 million USD at the end of the 20-year repayment period, due to accrual of interest over the repayment period. Figure 9 shows the discrepancy between cumulative revenue received and the collective outstanding loan balance of the borrowing pool. Because outstanding loan balance (and the interest rate it carries) is distinct from revenue in an ICL program with a positive interest rate, revenue should be used to evaluate public cost instead of outstanding loan balance. For the same reason, the IRR implied by revenues and loan disbursement provides a superior measure of a program's financial robustness relative to the interest rate calculated on loan balances.

Figure 9: Modified REPAYE
Cumulative Repayment vs. Outstanding Loan Balance

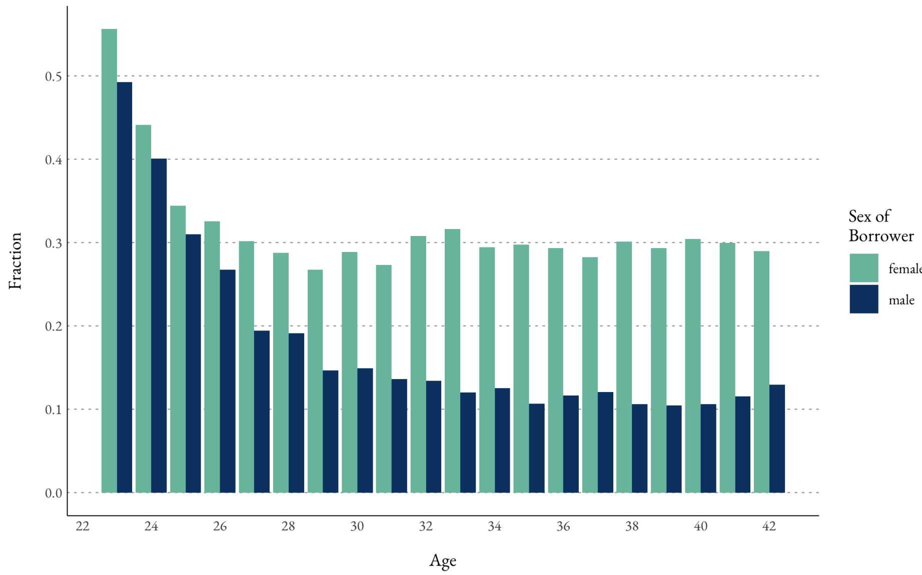


From a distributional perspective, the REPAYE program relies heavily on high earners for repayment revenue, while extending only minimal repayment insurance to borrowers with low earnings. As figure 10 shows, non-repayment rates due to earnings below the threshold of $18,210 are highest among individuals aged less than 25 years of age, reflecting greater employment uncertainty or volatility at these ages. The fraction of female borrowers in non-repayment at each age exceeds that of male borrowers, reflecting overall lower earnings as well as the higher volatility in female earnings observed in section 4. Importantly, while male non-repayment rates hover just above 10% throughout most of the repayment period, female repayment rates never drop below 25%. As a result, male borrowers pay down more than half the loan balance, despite representing only half of the borrowing cohort. Attempts to innovate on the modified REPAYE ICL scheme should balance equity of repayment burden between male and female borrowers against an insurance component for low earners, who are disproportionately female.

5.2 Proposed ICL Policies

The results above show that the REPAYE program has the potential to satisfy, and in fact exceed, revenue neutrality if implemented as an inclusive and national-scale replacement for existing federal student loan programs. The modified REPAYE program's IRR of 4.76% suggests that there is opportunity to further reduce repayment burden on low-income borrowers without requiring a subsidy from the general tax base

Figure 10: Modified REPAYE, Fraction of Borrowers in Non-payment
(Post-tax earnings less than $18,210)



to cover the costs of federal student loan financing. To evaluate ICL policy alternatives to the modified REPAYE program, I simulated eight policy scenarios. From these ICL program scenarios, I select two policies to evaluate as alternatives to the modified REPAYE program. First, a graduate tax scenario extends the repayment period until retirement, while maintaining a relatively low repayment rate of 3% above a threshold of $25,000. Second, a capped repayment plan carries a repayment rate of 10% above an earnings threshold of $30,000 and limits repayments to two times the amount borrowed. Parameters of each model are outlined in table 4 and discussed in detail in sections 5.2.1 and 5.2.2.

Table 4: Simulated Policy Scenarios

	Threshold	Interest Rate	Repayment Period	Repayment Rate	Payment Cap
Modified REPAYE	$18,210	4.45%	20 years	10%	none
Graduate Tax	$25,000	0%	43 years	3%	none
Capped Repayment	$30,000	0%	25 years	10%	200% of loan amount

In creating the graduate tax and capped repayment scenarios, I manipulate a number of policy parameters to tune relief from high repayment burden at low incomes, as well as the distribution of total loan cost among repaying individuals. Each scenario is calculated in the same manner as described in section 5.1, subject to the following policy parameters:

- Repayment threshold
- Interest rate
- Repayment rate

- Repayment period
- Repayment cap (as a multiple of the original loan amount)

Moving forward, all proposed ICL programs carry zero interest rate. As the modified REPAYE program simulation shows, interest rates are relevant to outstanding loan balances but not overall revenue in an ICL program. Since borrowers will not be taxed on loan forgiveness, there is no reason to assign penalties to forgiveness on relatively large sums that have been compounded by interest. Additionally, since revenue far exceeds loan disbursements in the modified REPAYE program, I raise the earnings threshold for repayment on both the graduate tax and capped repayment scenarios. The modified REPAYE plan's repayment threshold of $18,210 is low compared to thresholds of GBP 21,000 and AUD 57,000 in the United Kingdom and Australia respectively. While raising the earnings threshold may drive down overall revenue, it also expands the ICL insurance component for low-earners. Similarly, a policy's repayment rate interacts with the repayment threshold to moderate the repayment amounts. Lowering repayment rates as a means of reducing repayment burden for low-income borrowers must be considered relative to the repayment threshold. Because payments are calculated as the difference between after-tax earnings and the repayment threshold, loan payments are initially quite small as individuals cross the earnings threshold but balloon quickly as higher levels of earnings diverge from the repayment threshold.

5.2.1 Graduate Tax Scenario

In the graduate tax scenario, I seek to understand how an extended repayment period and low repayment rate interact with distributional concerns and income smoothing in ICL design. By extending the repayment period until retirement at age 65, earners have a larger number of periods during which they can pay down the pooled loan balance. From the perspective of gender equity, this allows for women to pay larger quantities into the pooled loan balance after they move through earnings volatility associated with child-bearing years. Because of the 43-year duration of the graduate tax scenario, more than twice that of the REPAYE program, I set the repayment rate to a relatively low level of 3% to moderate overall repayment contributions. Because the repayment rate is sufficiently low, and with the goal of including more women in repayment, the threshold for this scheme is set to $25,000, which is higher than the REPAYE threshold but lower than the capped repayment scenario.

5.2.2 Capped Repayment Scenario

Like the graduate tax scenario, the capped repayment scenario attempts to address the only modest income-smoothing characteristics of the modified REPAYE scenario. Additionally, the capped repayment plan seeks to limit overall repayment burden among high earners to a predetermined level. Under this policy, individuals who have payed two times the loan amount borrowed ($90,000 in this case) are exempted from additional payments. Additionally, the capped repayment plan's high repayment threshold of USD $30,000 offers the greatest degree of protection for low earners among the proposed ICL policies. The plan requires a repayment rate of 10% above the earnings threshold. Importantly, unlike the REPAYE and graduate tax scenarios, the capped repayment scenario includes a mechanism linking the value of the loan to the total repayment amount. High earners who disproportionately pay down their cohort's pooled loan balance thus have assurance that, while they benefit from same insurance against earnings volatility as other borrowers, a limit is placed on the overall repayment burden they carry relative to their peers.

5.2.3 Comparing Public Cost

The modified REPAYE, graduate tax, and capped repayment ICL policy scenarios all meet the self-financing requirement, with the present value of repayments exceeding the dollar amount of loans disbursed to the cohort upon enrollment in higher education. As I point out above, simply stating that future revenues will exceed total loan disbursement may not be adequate to evaluate true public cost. Table 5 shows internal rates of return for each policy scenario. Cash flows projected in the modified REPAYE plan carry the highest IRR of 4.76%. However, a completely revenue neutral policy should carry an IRR of 0%, indicating that the present value of loan disbursements is precisely equal to the discounted value of future repayments. In this sense, I evaluate the fiscal robustness of the three proposed ICL policies in terms of a near-zero IRR (approaching true revenue neutrality) with a buffer to account for unexpected events, such as recessions or structural change in the labor market, that may interrupt projected repayments.

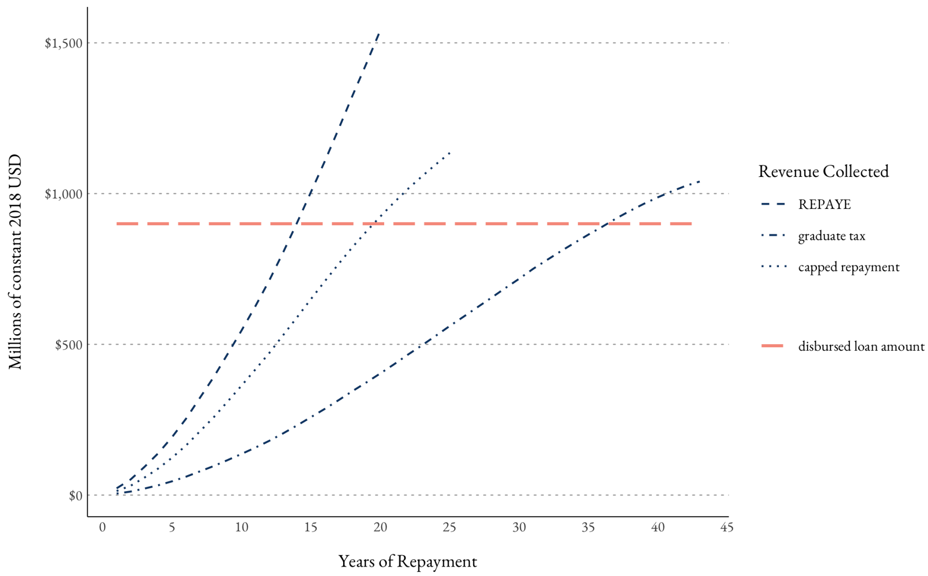
Table 5: Fiscal Characteristics

	IRR	Loans Disbursed ($ millions)	Revenues Collected ($ millions)
Modified REPAYE	4.76%	900	1,542
Graduate Tax	0.63%	900	1,040
Capped Repayment	1.77%	900	1,134

Note: Revenue totals are not adjusted to present value

As table 5 shows, the graduate tax scenario carries the both the lowest IRR, 0.63%, and the lowest repayment revenue, USD 1,040 million, among proposed ICL policies. In addition, the 43-year repayment period carries a larger degree of risk and uncertainty relative to the comparatively short repayment periods in the modified REPAYE and capped repayment scenarios (20 and 25 years respectively). While projected revenues for the graduate tax policy most closely approach true revenue neutrality, this policy also carries the highest uncertainty over how earnings dynamics may evolve away from projections over the more than four decade repayment period. As figure 11 shows, revenue in the graduate tax plan exceeds the quantity of loans disbursed after 35 years of repayment, relative to less than 20 years under the modified REPAYE and capped repayment programs. Alternatively, the capped repayment scenario carries an intermediate IRR relative to the modified REPAYE and graduate tax scenarios. With an IRR of 1.77%, the capped repayment scenario's 25 year repayment period significantly reduces uncertainty about revenue projections compared to the graduate tax scenario.

Figure 11: Cumulative Revenue



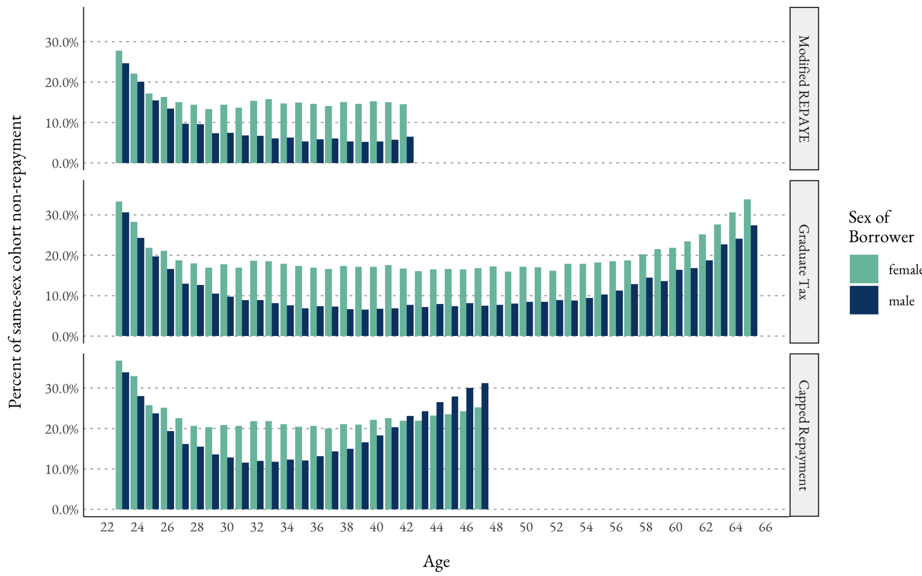
5.2.4 Distributional Characteristics

Establishing distributional equity in an ICL program is challenging, requiring a both insurance against financial hardship for low-earners and avoidance of excessive reliance on high-earners to pay down the majority of the pooled loan balance. I evaluate distributional outcomes as a balance between which individuals benefit from repayment insurance and which repay relatively high quantities over the full repayment period.

Distributional considerations fracture largely along the lines of sex, raising questions over repayment parity as male and female incomes diverge over evolving career paths and life events. As figure 12 shows, a higher proportion of women than men are in non-payment at every age in the modified REPAYE and graduate tax scenarios. The disproportionate number of female borrowers in non-payment during child-bearing years suggests that ICL policies extend relief to women engaged in child-rearing. Slightly elevated rates of female non-payment among women aged 31 to 34 in all policy scenarios matches a temporary reduction copula estimates of earnings correlation at these ages (figure 6).

Age also plays an important role determining the proportion of individuals falling below the repayment threshold. Non-payment rates are especially high for both men and women younger than age 25 under all policies, highlighting the need for flexible repayment during the period of uncertain earnings following graduation. Similarly, low earnings in late career are reflected in the proportion of individuals in non-payment under the graduate tax scenario, which rises substantially for both men and women after age 60. This draws into question the consequences of claiming a portion of borrowers' income as they approach and prepare for retirement, especially since the graduate tax scenario places no constraint on lifetime repayment relative to the initial loan amount.

Figure 12: Fraction of Borrowers in Non-Repayment by Age

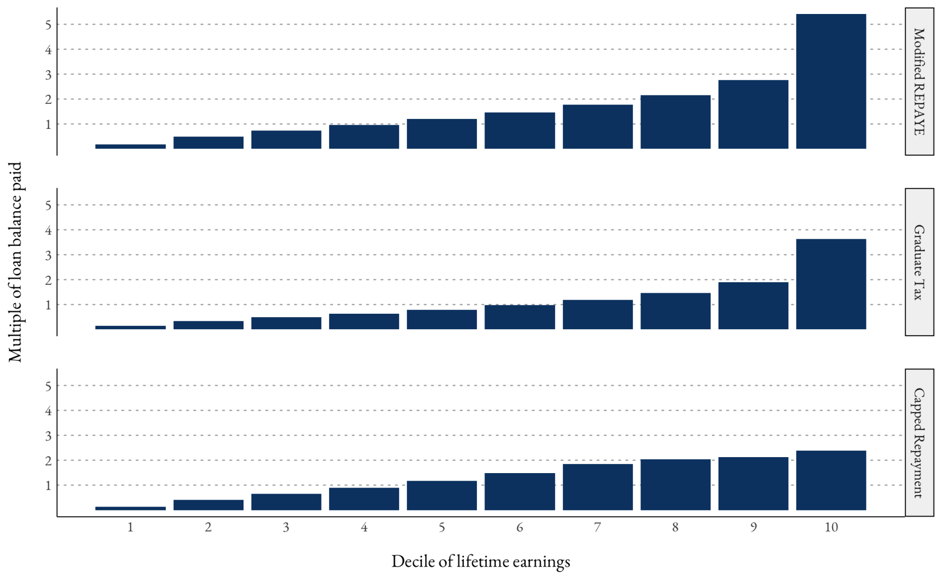


By contrast, the proportion of individuals in non-payment under the capped repayment plan also rises towards the end of the 25-year repayment period, although for different reasons. Individuals whose lifetime earnings exceed two times the original loan amount ($90,000) are exempted from additional repayment. Among the simulated cohort of 20,000 borrowers, a total of 6,357 individuals meet the repayment cap over the course of the repayment period. However, just 23% of individuals meeting the repayment cap are female. As figure 12 shows, non-payment rates for men exceed those of women after age 40 under the capped repayment plan, reflecting the larger proportion of men than women who meet the repayment cap and are exempted from repayment. However, the capped repayment plan's mid-career non-payment rates among both male and female borrowers is the highest among all of the ICL policies, demonstrating the effects of a higher repayment threshold in exempting low-income borrowers from the financial stress in making repayments. In this respect, the capped repayment plan is the only policy that offers both the insurance characteristics of a high repayment threshold and a mechanism to cap repayments among high-earners below a level of undue burden. Under the capped repayment plan earners in the top deciles of lifetime earnings (regardless of sex) pay just two times the amount borrowed (figure 13).[5] By contrast, since the modified REPAYE and

[5]Borrowers are exempted from additional payments after reaching two times the original loan amounts. Because final monthly payments calculated as 10% of discretionary earnings may slightly exceed the $90,000 repayment cap, earners in the top decile pay slightly more than two times the loan amount on average.

graduate tax scenarios do not include repayment caps, the highest earning borrowers pay 5.4 and 3.6 times the loan amount borrowed respectively, indicating a disproportionate reliance on high earners to repay the cohort's pooled loan balance.

Figure 13: Multiples of Loan Amount Repaid by Lifetime Earnings Decile



6 Discussion

6.1 ICL Policy Recommendations

Among the income-contingent student loan policies proposed, the capped repayment plan offers both equitable outcomes for borrowers of all income levels and sufficient revenue streams to robustly satisfy the self-financing constraint. With the highest repayment threshold of any plan, individuals are exempted from repayment during periods of annual earnings less than $30,000. Despite the generous insurance for low earners, total lifetime repayments the capped repayment plan distributes well across lifetime earnings deciles, with the median borrower paying $58,959 relative to the $45,000 amount borrowed. While repayments among high earners exceed the loan amount borrowed, they too benefit from insurance against repayment under periods of low-income, as well as exemption from payment after repaying two times the initial loan amount.

Capping repayment among high earning borrowers may also be a consideration for the sustainability and longevity of the ICL program. Under the modified REPAYE and graduate tax scenarios substantial reliance on high earners for revenue, skewed repayments may fairly reflect equally skewed lifetime earnings between high- and low-earners. However, two issues emerge to temper enthusiasm for such a skewed repayment structure. First, such a skewed level of repayment may render mandatory ICL policies unpopular among high-earners. High-earners are also likely to possess disproportionate resources to lobby against an unpopular policy, perhaps threatening the longevity of an ICL program and the insurance benefits it offers to low-earning college graduates. Additionally, reliance on a small subset of borrowers to pay a large portion of the collective loan balance may encourage strategic behavior among individuals to misreport earnings or minimize income streams eligible for ICL repayment. If such strategic behavior were substantial enough among top-earners, the financial performance of the cohort's loan pool could be threatened.

With these considerations in mind, the capped repayment plan alone links lifetime repayment to the cost of college attendance while substantially reducing risk to low-earning borrowers. As a self-financing policy that also moderates distributional considerations among borrowers, the capped repayment plan is the best option from the perspective of both public cost and distributional equity among borrowers.

6.2 Opportunities for future research

Federal student financing in the United States is highly complex, representing an amalgamation of complex grant and loan programs targeting borrowers from diverse backgrounds. In large part, the complexity of federal student financing rises from the need to accommodate borrowers from vulnerable groups with affordable access to higher education. While income-contingent loans can do much to address the needs of low-earning and risk-averse borrowers, further research is needed to understand how dynamic income projections interact with diverse borrower characteristics, with profound implications for projections of both public cost and distributional considerations. Opportunities for future research include dynamic earnings models that account for demographics known to be driving forces of lifetime earnings, such as race, as well as transitional periods of unemployment. Additionally, the ICL policies I propose assume maximum federal loan amounts at the current status quo. Assessing the viability of ICLs with more generous lending limits would do much to address funding constraints experienced by many student borrowers. Finally, projections of public cost may be sensitive to opportunistic behavior among borrowers, who may shift earnings behavior around the repayment threshold. Creating a model that is sufficiently robust to such behavior is a key consideration in designing an affordable ICL policy.

7 Conclusion

Policy simulations using earnings dynamics from the National Longitudinal Survey of Youth demonstrate that income-contingent loans are not only affordable from the perspective of public cost but also adaptable to differing levels of progressivity in overall repayment contributions within the borrowing cohort. As the cost of attendance at public and private colleges continues to climb, ICLs have the potential to reshape student borrowers' evaluation of risks and rewards. In turn, insurance against repayment stress may encourage a larger number of students towards fields of study leading towards high-prestige, low-earnings careers, such as teaching and other forms of public service. The proposed capped repayment ICL program has the potential to isolate cost of attendance among college graduates, while both removing risk as borrowers move through an uncertain labor market and constraining overall repayment contribution to equitable levels.

References

S. Armstrong, L. Dearden, M. Kobayashi, and N. Nagase. Student loans in japan: Current problems and possible solutions. *Economics of Education Review*, 71:120 – 134, 2019. ISSN 0272-7757. doi: https://doi.org/10.1016/j.econedurev.2018.10.012. URL http://www.sciencedirect.com/science/article/pii/S027277571730691X. Higher Education Financing: Student Loans.

M. J. Bailey and S. M. Dynarski. Gains and gaps: Changing inequality in u.s. college entry and completion. Working Paper 17633, National Bureau of Economic Research, December 2011. URL http://www.nber.org/papers/w17633.

N. Barr, B. Chapman, L. Dearden, and S. Dynarski. The us college loans system: Lessons from australia and england. *Economics of Education Review*, 71:32 – 48, 2019. ISSN 0272-7757. doi: https://doi.org/10.1016/j.econedurev.2018.07.007. URL http://www.sciencedirect.com/science/article/pii/S0272775718301778. Higher Education Financing: Student Loans.

S. Bonhomme and J.-M. Robin. Assessing the equalizing force of mobility using short panels: France, 1990-2000. Université paris1 panthéon-sorbonne (post-print and working papers), HAL, 2009. URL https://EconPapers.repec.org/RePEc:hal:cesptp:hal-00308801.

J. Britton, L. van der Erve, and T. Higgins. Income contingent student loan design: Lessons from around the world. *Economics of Education Review*, 71:65 – 82, 2019. ISSN 0272-7757. doi: https://doi.org/10.1016/j.econedurev.2018.06.001. URL http://www.sciencedirect.com/science/article/pii/S0272775717301668. Higher Education Financing: Student Loans.

Y. Cai, B. Chapman, and Q. Wang. Repayment burdens of mortgage-style student loans in china and steps toward income-contingent loans. *Economics of Education Review*, 71:95 – 108, 2019. ISSN 0272-7757. doi: https://doi.org/10.1016/j.econedurev.2018.10.006. URL http://www.sciencedirect.com/science/article/pii/S0272775717302078. Higher Education Financing: Student Loans.

B. Chapman. Income Contingent Loans for Higher Education: International Reform. CEPR Discussion Papers 491, Centre for Economic Policy Research, Research School of Economics, Australian National University, June 2005. URL https://ideas.repec.org/p/auu/dpaper/491.html.

R. Chetty, J. N. Friedman, E. Saez, N. Turner, and D. Yagan. Mobility report cards: The role of colleges in intergenerational mobility. Working Paper 23618, National Bureau of Economic Research, July 2017. URL http://www.nber.org/papers/w23618.

L. Dearden. Evaluating and designing student loan systems: An overview of empirical approaches. *Economics of Education Review*, 71:49 – 64, 2019. ISSN 0272-7757. doi: https://doi.org/10.1016/j.econedurev.2018.11.003. URL http://www.sciencedirect.com/science/article/pii/S0272775717304995. Higher Education Financing: Student Loans.

S. Demarta and A. J. McNeil. The t copula and related copulas. *International Statistical Review / Revue Internationale de Statistique*, 73(1):111–129, 2005. ISSN 03067734, 17515823. URL http://www.jstor.org/stable/25472643.

R. Diris and E. Ooghe. The economics of financing higher education. *Economic Policy*, 33(94):265–314, 03 2018. ISSN 0266-4658. doi: 10.1093/epolic/eiy003. URL https://doi.org/10.1093/epolic/eiy003.

S. Flood and J. Pacas. Using the annual social and economic supplement as part of a current population survey panel. *Journal of Economic and Social Measurement*, 42(3-4):225–248, 2016. doi: 10.3233/JEM-180447.

M. Friedman. The role of government in education, 1955.

J. Geweke and M. Keane. An empirical analysis of earnings dynamics among men in the psid: 1968-1989. *Journal of Econometrics*, 96(2):293–356, 2000. URL https://EconPapers.repec.org/RePEc:eee:econom:v:96:y:2000:i:2:p:293-356.

V. Gonzalez, L. Ahlman, and A. Fung. Student debt and the class of 2018,. Technical report, Washington D.C., 2019.

E. A. Hanushek, C. K. Y. Leung, and K. Yilmaz. Borrowing constraints, college aid, and intergenerational mobility. Working Paper 10711, National Bureau of Economic Research, August 2004. URL http://www.nber.org/papers/w10711.

T. Higgins and M. Sinning. Modeling income dynamics for public policy design: An application to income contingent student loans. *Economics of Education Review*, 37:273 – 285, 2013. ISSN 0272-7757. doi: https://doi.org/10.1016/j.econedurev.2013.08.009. URL http://www.sciencedirect.com/science/article/pii/S0272775713001118.

L. Lochner and A. Monge-Naranjo. Chapter 8 - student loans and repayment: Theory, evidence, and policy. volume 5 of *Handbook of the Economics of Education*, pages 397 – 478. Elsevier, 2016. doi: https://doi.org/10.1016/B978-0-444-63459-7.00008-7. URL http://www.sciencedirect.com/science/article/pii/B9780444634597000087.

R. Murphy, J. Scott-Clayton, and G. Wyness. The end of free college in england: Implications for quality, enrolments, and equity. Working Paper 23888, National Bureau of Economic Research, September 2017. URL http://www.nber.org/papers/w23888.

M. Nerlove. Some problems in the use of income-contingent loans for the finance of higher education. *Journal of Political Economy*, 83(1):157–183, 1975. ISSN 00223808, 1537534X. URL http://www.jstor.org/stable/1833278.

A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris*, 8:229–231, 1959.