

Inequality in Parental Transfers, Borrowing Constraints, and Optimal Higher Education Subsidies*

Preliminary and Incomplete

Youngmin Park

Bank of Canada

February 14, 2018

Abstract

This paper studies optimal education subsidies when parental transfers for college education are unequally distributed across students and cannot be publicly observed. After documenting substantial inequality in the amount of parental transfers among American college students with similar observed family resources, I examine the implications of unobservable heterogeneity in parental transfers for efficient design of education subsidy policy that minimizes the distortions generated by borrowing constraints. When inequality in educational attainment is driven by differences in parental transfers, providing larger subsidy for lower level of schooling is optimal, because additional resources given to constrained students choosing low schooling levels reduce distortions. This force is weakened if unobservable heterogeneity in returns to schooling also leads to different schooling choices. To quantify these mechanisms, I build a model with endogenous parental transfers where inequality in parental transfers among students with similar parental economic resources and returns to education is determined by heterogeneity in parental altruism. The quantitative model is calibrated to the U.S. economy and used to solve for the optimal subsidy that may assign different amounts for each year of college and parental income quartile. The optimal policy subsidizes the first two years of college much more heavily than later years. The shift of public spending towards early years of college is more pronounced for higher parental income groups, generating little variation of subsidy amounts for the first two years of college across parental income.

*This paper is a heavily revised version of the second chapter of my Ph.D. dissertation at the University of Western Ontario that previously circulated under the title, "Heterogeneity in Parental Altruism and Optimal Higher Education Subsidies". I am very grateful to my thesis advisors, Lance Lochner, Igor Livshits, and Betsy Caucutt, for their guidance and support. I thank seminar/conference participants at the Korea Development Institute, Korea Institute of Public Finance, University of New Brunswick, Bank of Canada, Queen's University, Federal Reserve Bank of Atlanta, 2014 Canadian Economic Association Annual Conference, 2015 Summer School on Socioeconomic Inequality (Chicago), Conference on Economics of Education (Penn State), Workshop on Intergenerational Transmission of Economic Status (Carlos III), 2016 Canadian Public Economics Group Meeting, HCEO Workshop on Determinants of Human Capital Formation (Chicago) for helpful comments and suggestions. This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET) and Compute/Calcul Canada. The views expressed in this paper are those of the author and do not necessarily represent those of the Bank of Canada.

1 Introduction

In the U.S., the cost of a college education is high and many students receive financial help from their parents to pay for it. During the 2016–2017 academic year, the average total cost of college for American undergraduate students was \$23,757, of which 35% was covered by parents and other relatives while another 35% was covered by scholarships and grants ([Sallie Mae, 2017](#)). However, the amount of parental contributions differs greatly across students,¹ reflecting differences in various factors relevant for parental transfer decisions, such as preferences and the amount of economic resources available for the family.²

Growing evidence suggests that money received from parents during youth is an important determinant of educational attainment, and that inequality in parental transfers can explain sizable differences in schooling outcomes by family background ([Keane and Wolpin, 2001](#); [Johnson, 2013](#)). If a college financial aid policy is to reduce inequality in educational achievement originating from unequal family support, it would be effective to explicitly target students who lack substantial support from parents. However, such policies are not feasible as, unlike income and wealth, private transfers between individuals are hard to be observed and verified. This paper studies efficient design of financial aid (or subsidy) policies when parental transfers are unequally distributed across students and cannot be directly observed by the policymaker.

I begin by documenting new facts about the amount of transfers American undergraduate students received from their parents using data from the 2011–2012 National Post-Secondary Student Aid Survey (NPSAS:12). First, the amount of parental transfers is very unequally distributed across students, even among those who have similar family resources and pay similar total college expenses. This implies that current need-based³ financial aid system in the U.S., which awards larger amount of financial aid to students with lower family resources, cannot precisely target those in need, because parental resources are imperfectly correlated with the amount of resources actually available to students. Second, conditional on family resources and net college costs (total cost less grant aid), students who receive lower parental transfers are more likely to exhaust their government loans, borrow from private lenders, and work longer hours while enrolled, all of which indicating they might have greater difficulty in making ends meet. Given fairly limited ability of youth to borrow against their future earnings,⁴ those without financial support from parents might be unable to afford college.

Motivated by this evidence, I analytically examine optimal education subsidies in a two-period model of discrete schooling choice where otherwise identical students are endowed with different amount of initial wealth (or parental transfer), which is not publicly observable, and face limited borrowing opportunities to pay for direct schooling costs as well as consumption expenditure while in school. As discussed by [Becker \(1994\)](#), students with sufficiently high parental transfer are not credit constrained and choose the schooling level that maximizes lifetime earnings net of schooling costs, while those with lower parental transfers may choose a lower cost option because they must suffer from inefficiently low consumption during schooling when the credit constraint binds. The social planner aims to minimize inefficiencies caused by borrowing constraints—distorted schooling

¹For example, [Haider and McGarry \(2012\)](#) documented that, among the respondents of the Health and Retirement Survey (HRS), the two most common answers for the fraction of tuition they covered for their children were 100% and 0%.

²[Ellwood and Kane \(2000\)](#) and [Hotz et al. \(2017\)](#) estimated that parental transfers are positively correlated with parental income/wealth.

³Although the importance of merit aid has increased recently ([McPherson and Schapiro, 1998](#)), most of the non-repayable financial aid is still need-based: in 2011–2012, only 35% of the grant aid received by public four-year college students was non-need-based ([College Board, 2016](#)).

⁴The U.S. government student loans have tight limits: in 2016–2017, the maximum amount of Stafford loans for first-year dependent undergraduate students was \$5,500, covering only 23% of the average cost in the same year. Most private student loans require evidence of creditworthiness, which many undergraduate students would fail to provide, and charge higher interest rates than the interest rates offered by the government student loans ([Consumer Financial Protection Bureau, 2012](#)).

investment and intertemporal consumption allocation—by distributing an exogenously given amount of public funds through an education subsidy that conditions on schooling levels.

The main theoretical result is that the unobservable inequality in parental transfers creates a force to heavily subsidize lower levels of schooling. In particular, I provide conditions under which the optimal marginal subsidy is negative in the sense that the amount of subsidy decreases in schooling levels. If initial wealth were publicly observable, giving larger lump-sum transfers to constrained students with lower initial wealth would reduce aggregate distortions. Although such individual-specific lump-sum transfers cannot be implemented due to private information, the distribution targeted to students with lower initial wealth can be achieved by raising the amount of subsidy for lower schooling levels chosen by them, exploiting the variation in schooling choice induced by inequality in initial wealth and borrowing constraints.

I also show that the existence of fewer students with low initial wealth leads to higher optimal subsidies, especially for low schooling levels. This provides a mechanism that weakens the dependence of subsidy amounts on family resources at low levels of schooling, when the subsidy schedule is set optimally for multiple groups of students with different observable family resources: although it is efficient to spend less public funds on students from richer families, who receive higher parental transfers on average, a high fraction of public funds available for one group can be directed towards constrained students when there are only a few of them. This highlights that the property of the subsidy schedule, set optimally for a group of students with identical family resources (but with heterogeneous parental transfers), is also important for how the amount of optimal subsidy varies across groups for each schooling level.

The optimality of negative marginal subsidy contrasts with the U.S. need-based financial aid policy, under which students receive higher total financial aid over the lifetime than others with similar family resources if they stay in college for longer periods of time or attend more expensive institutions; it critically hinges on the condition that the allocations of students choosing lower schooling levels are more distorted by credit constraints, which in turn results from the assumption that students differ only in the amount of transfers received from their parents. However, the returns to schooling vary substantially across individuals and they are also important for schooling decisions.⁵ I show that, when differences in educational attainment are driven by unobservable differences in returns to schooling (rather than parental transfers), it is optimal to provide a positive marginal subsidy, as those choosing more costly schooling options are more likely to be constrained, a mechanism first introduced by [De Fraja \(2002\)](#). More generally, when students differ along both dimensions, the policymaker must weigh the relative importance of the two types of heterogeneity when designing an education subsidy.

To quantify these mechanisms, I extend the model to a life cycle setting, where years of college education represent schooling choices, and calibrate it to the U.S. economy. The quantitative model incorporates heterogeneity in wage returns to schooling, which I call ‘ability’, as well as psychic returns. Importantly, I follow [Becker and Tomes \(1986\)](#) in explicitly modeling transfer decisions of altruistic parents who care about themselves as well as their children in order to capture endogeneity of parental transfers with respect to education policies. The degree of parental altruism is allowed to differ across families, which is the key to generating inequality in parental transfers conditional on income and wealth. The model parameters are chosen so that the model replicates the joint distribution of educational attainment, parental income, parental transfer, and measured ability of American youth from the National Longitudinal Surveys of Youth 1997 (NLSY97) data under current financial aid policy.

The calibrated model is used to solve for budget-equivalent optimal subsidy schedules separately by parental income quartile. Compared to the current policy, the optimal policy gives larger amount of total subsidy to youth

⁵See, for example, [Carneiro, Hansen, and Heckman \(2003\)](#), [Cunha, Heckman, and Navarro \(2005\)](#), and [Carneiro, Heckman, and Vytlačil \(2011\)](#).

completing the first two years of college, while those completing the fourth year receive lower total subsidies. For the first two years of college, the amount of optimal subsidy varies less between the lowest and the third parental income quartile, providing somewhat similar public support to all youth regardless of their family background. Although the average years of college education is lower under the optimal policy, efficiency in schooling investment is improved due to lower dispersion in educational attainment: more youth attend college but less of them attain four years of college education. In addition to more efficient schooling, larger efficiency gain comes from better intertemporal consumption smoothing, despite the fact that parental transfers crowd out as much as the amount of aggregate distortions reduced by the optimal policy.

While there is a large literature that quantitatively evaluates education policies in the presence of inequality in intergenerational transfers,⁶ little attention has been paid to the implications of such inequality among families with similar economic resources, which I call ‘within-group inequality’. By explicitly modeling transfer decisions of heterogeneously altruistic parents and calibrating based on parental transfers data, I shed light on the consequences of the within-group inequality in intergenerational transfers. Specifically, I show that introducing heterogeneity in parental altruism, among families with identical parental income/wealth and ability, leads to a drastic change in the nature of optimal subsidy policy: a substantial reallocation of public funds towards lower years of college. As discussed earlier, the within-group inequality in parental transfers also affects the distribution of optimal subsidy across groups.

The implications of the inequality in parental transfers, driven by heterogeneity in parental altruism, have been studied in the context of optimal estate taxation by [Weinzierl \(2008\)](#) and [Farhi and Werning \(2013\)](#), who show how the optimal consumption allocation is determined by social preferences for equity within as well as across generations. Although reducing consumption inequality is one of the important objectives for tax and transfer policies, it may not be the direct goal of college financial aid policies, which are often said to aim at increasing college access for students who would not otherwise have the financial resources to attend. Therefore, this paper focuses on credit constraints as the fundamental market imperfection to be addressed, abstracting from equity concerns. Building on [Bénabou \(2002\)](#), distortions in schooling and intertemporal consumption allocation are measured in monetary terms by individuals’ maximum willingness to pay to eliminate them, and then aggregated across individuals to construct a social objective function to be minimized.⁷ This approach clarifies what a policy intends to achieve, and it transparently captures the key trade-off faced by individuals as well as the social planner in the economy: schooling vs. intertemporal consumption smoothing.

The rest of this paper is organized as follows. Section 2 reports U.S. evidence on the relationship between parental transfers, family resources, and other student outcomes. Section 3 uses a two-period model to characterize the optimal education subsidy under alternative assumptions about underlying unobservable heterogeneity. Section 4 extends the two-period model to a multiperiod life cycle and also explicitly model parental transfer decisions. Section 5 presents calibration and Section 6 uses the calibrated model to quantitatively characterizes the optimal subsidies. Section 7 concludes.

⁶Examples include [Keane and Wolpin \(2001\)](#); [Caucutt and Kumar \(2003\)](#); [Restuccia and Urrutia \(2004\)](#); [Johnson \(2013\)](#); [Hanushek, Leung, and Yilmaz \(2014\)](#); [Abbott et al. \(2016\)](#); [Krueger and Ludwig \(2016\)](#); [Findeisen and Sachs \(2017\)](#).

⁷The idea of measuring economic efficiency by a sum of individuals’ willingness to pay is closely related to the classic work of [Kaldor \(1939\)](#) and [Hicks \(1940\)](#), which is further developed by [Hendren \(2017\)](#).

2 Parental Transfers, Family Resources, and Student Outcomes

2.1 U.S. Financial Aid System for College Education

In the U.S., most financial aid is need-based: federal rules determine federal grants and subsidized loans based on student need, and similar need calculation is used for state and institutional aid. A student's financial "need" equals the difference between the total estimated cost, which includes tuition and living expenses, of attending a given college and the Expected Family Contribution (EFC). The EFC is a number that determines students' eligibility for federal student aid, and it is intended to measure what a family can reasonably be expected to contribute to a child's education for the year. The EFC is calculated according to a formula specified in the law, using the financial information the student provides on the Free Application for Federal Student Aid (FAFSA), which must be filled out by all students seeking federal aid. For dependent students, EFCs depend primarily on parental income, while assets play a minor role.⁸ Generally, the amount of subsidized aid (such as grants and subsidized loans) awarded is increasing in student need (Dick and Edlin, 1997; Belley, Frenette, and Lochner, 2014), generating substantial differences in effective costs of college between students with different parental income.

2.2 Relationship between Expected and Actual Parental Contribution

The need-based financial aid system is built on the explicit expectation about how much parents will contribute to their children's college education, but not much is known about the actual amount of parental contribution relative to their EFC. In this subsection, I document the empirical patterns of expected and actual amount of parental transfers using the NPSAS:12 data.

The NPSAS:12 is a nationally representative sample survey of students enrolled during the 2011–2012 academic year in post-secondary institutions in the United States. It provides financial aid and family background data from administrative records and student surveys. Importantly, it contains both EFC and actual amount of parental contribution. The EFC comes from financial aid administrative records and the amount of parental transfers is based on the student survey question "through the end of the 2011–12 school year, about how much will your parents have helped you pay for any of your education and living expenses while you are enrolled in school?". However, the amount of parental transfer is only available in 12 categories on the National Center for Education Statistics (NCES) DataLab website⁹, where all NPSAS results in this paper were obtained using the table creation tool.

To ensure access to administrative data on parental income and financial aid packages, I only consider dependent students who had submitted a federal financial aid application. Because annual costs of attending college can substantially vary due to a few factors such as institution type and enrollment status, and they may affect the amount of financial aid as well as parental transfers among students with similar EFC, I apply additional sample restrictions to focus on a group of students facing relatively similar annual costs. I select U.S. citizens who were enrolled in four-year public institutions more than nine or more months full time during the 2011–2012 academic year at a single institution, paying the regular "in-jurisdiction" tuition fees. I also exclude students

⁸For some families with low parental income (e.g., \$25,000 or less in 2016–2017), the EFC is set to zero regardless of net worth. Moreover, the parents' net worth excludes the family's home, and it is subtracted by some exemption amounts, and then multiplied by the conversion rate of 12% before being added to the parents' total income (beyond some exemption levels) to compute the total parent's contribution.

⁹This website can be found at the following address: <http://nces.ed.gov/datalab>.

who lived at home while enrolled, since they had substantially lower costs due to savings in room and board charges. There is still remaining variation in costs, reflecting differences in tuition fees across states as well as differences in cost of living across regions. To address this concern, Appendix A presents some analysis conducted separately by (net) cost of college.

Table 1: EFC and Actual Parental Contribution

	% with Amount Parents Helped Pay for Expenses						Parental Income (\$)	Grants (\$)
	\$0 to \$2,000	\$1 to \$2,000	\$2,001 to \$5,000	\$5,001 to \$10,000	\$10,001 to \$20,000	\$20,001 or More		
EFC:								
\$0	40.3	35.5	8.6	7.9	4.9	2.8	17,782	10,631
\$1 to \$2,000	28.6	37.8	11.0	10.5	9.5	2.6	33,565	9,690
\$2,001 to \$5,000	30.8	35.1	12.9	11.6	6.3	3.3	51,361	6,633
\$5,001 to \$10,000	20.8	32.2	16.4	14.5	11.7	4.4	72,400	3,582
\$10,001 to \$20,000	14.8	26.3	17.0	14.6	18.5	8.7	103,130	2,634
\$20,001 or More	11.9	16.7	11.4	18.2	22.7	19.1	164,500	2,193
All	24.5	29.5	12.6	13.0	12.8	7.7	78,493	5,771

Table 1 reports the average amount of parental transfers and grants received in 2011–2012 and parental income in 2010, separately by 6 EFC categories. All monetary amounts in this subsection are current U.S. dollars. As the table shows, EFC is positively correlated with parental income and negatively correlated with grants: between EFC levels \$2,000 and \$10,000, around \$20,000 increase in parental income is associated with \$3,000 reduction in grants through larger EFC. However, actual parental contributions are generally not in line with EFCs, as the EFC categories are different from the actual transfer categories for substantial fraction of students. Although the actual amount of parental contribution is positively correlated with EFC, there is substantial heterogeneity in parental transfers among families with similar EFC. Even among those with more than \$20,000 EFC, more than 10% of students do not receive anything from their parents. Similarly, more than 10% of parents with zero EFC give more than \$5,000 to their children.

2.3 Effects of Parental Transfers on Student Borrowing and Labor Supply

The difference between the EFC and actual parental contribution might be entirely explained by measurement errors, as the latter data are collected through survey, and students may provide accurate answers. If this is the case, then the amount of actual parental contribution is likely to be uncorrelated with other student outcomes.

Table 2 shows the estimated effects of parental contributions on how students finance their college costs. These estimates demonstrate that those without help from parents are more likely to exhaust government student loans, take out private loans, and work longer hours while enrolled.

Summarizing the evidence from the NPSAS:12, I find that EFCs are in general different from actual amount of parental transfers. Those who do not receive help from parents are likely to borrow more and work longer hours. Overall, this evidence suggests that the need-based financial aid system may not be able to successfully eliminate the differences in education opportunities across students that result from differences in parental transfers.

Table 2: Estimated Effects on Borrowing and Labor Supply

	Took out		Hours Worked Per Week While Enrolled
	Maximum Federal Loans	Private Loans	
	(1)	(2)	(3)
Parental Contribution:			
\$1 to \$2,000	0.007 (0.019)	-0.012 (0.014)	-3.165* (0.538)
\$2,001 to \$5,000	-0.033 (0.024)	-0.055* (0.016)	-3.863* (0.629)
\$5,001 to \$10,000	-0.075* (0.030)	-0.040* (0.018)	-4.327* (0.708)
\$10,001 to \$20,000	-0.107* (0.027)	-0.087* (0.017)	-5.992* (0.717)
\$20,001 or More	-0.124* (0.035)	-0.076* (0.020)	-4.843* (0.881)
EFC:			
\$1 to \$2,000	-0.046 (0.034)	0.011 (0.016)	-0.761 (0.838)
\$2,001 to \$5,000	-0.006 (0.028)	0.051* (0.014)	-1.246 (0.638)
\$5,001 to \$10,000	-0.013 (0.029)	0.056* (0.016)	0.216 (0.726)
\$10,001 to \$20,000	-0.025 (0.034)	0.059* (0.017)	-0.078 (0.776)
\$20,001 or More	-0.132* (0.032)	0.037* (0.015)	-0.240 (0.673)
Cost Less Grant:			
\$10,001 to \$15,000	0.117* (0.024)	0.033* (0.011)	1.843* (0.684)
\$15,001 to \$20,000	0.204* (0.027)	0.065* (0.013)	2.402* (0.683)
\$20,001 or More	0.245* (0.029)	0.110* (0.017)	1.657* (0.729)
Constant	0.346* (0.023)	0.033* (0.009)	11.843* (0.599)

3 Theory of Optimal Education Subsidies

Motivated by the empirical evidence that there is substantial inequality in the amount of parental transfers among families with similar financial resources, I theoretically study how unobservable heterogeneity in initial resources among students affects the optimal design of education subsidy that minimizes aggregate distortions due to borrowing constraints. I first show that, eliminating inequality in parental transfers reduces aggregate distortions, even though there is no explicit social preference for inequality aversion. When parental transfers are not observable, however, the planner can achieve this goal by awarding different amount of subsidies across different education levels.

3.1 Modeling Schooling Choice

Youth live and consume for two periods, $t \in \{1, 2\}$, and their preferences over consumption are represented by the following log utility function:

$$\ln c_1 + \ln c_2, \quad (1)$$

which implies that time preference rate is zero and the elasticity of intertemporal substitution is 1. I make these assumptions for clean analytical results, although these assumptions are relaxed in the quantitative analysis.

Youth are endowed with initial wealth, or parental transfer, $b \geq 0$, and they face the choice set of schooling $\mathcal{J} \equiv \{0, 1, \dots, J\}$. The choice $j = 0$ represents not attending college, in which case youth pay nothing, i.e., $k_0 = 0$, start working and earn $y_0 > 0$. Let $\mathcal{J}_+ \equiv \mathcal{J} \setminus \{0\}$ be the choice set excluding the option of not attending college. The optimal consumption allocation is $c_1 = c_2 = (b + y_0)/2$, so the indirect utility of not attending college is

$$U_0(b) \equiv \max_{c_1 \geq 0, c_2 \geq 0} \left\{ \ln c_1 + \ln c_2 \mid c_1 + c_2 \leq b + y_0 \right\} = 2 \ln \left(\frac{b + y_0}{2} \right).$$

The choices $j \geq 1$, on the other hand, represents attending college, in which case, the youth attend college during period $t = 1$, need to pay tuition k_j , and earn $y_j > 0$ after schooling period $t = 2$. Youth can save at zero interest rate, but they cannot borrow. Therefore, when the youth attend college, they need to finance tuition cost and consumption during schooling out of their initial wealth.

$$\begin{aligned} U_j(b) &\equiv \max_{c_1 \geq 0, c_2 \geq 0} \left\{ \ln c_1 + \ln c_2 \right\}, \\ \text{subject to } c_1 + c_2 &\leq b - k_j + y_j, \\ c_1 &\leq b - k_j. \end{aligned} \quad (\text{BC})$$

When $b < k_j$, the consumption during schooling period cannot be positive, so the choice j is infeasible. In this case, I define the indirect utility to be $-\infty$. Therefore, the indirect utility can be written as follows:

$$U_j(b) = \begin{cases} -\infty & \text{for } b \leq k_j, \\ \ln(b - k_j) + \ln(y_j) & \text{for } b \in (k_j, k_j + y_j), \\ 2 \ln \left(\frac{b - k_j + y_j}{2} \right) & \text{for } b \geq k_j + y_j. \end{cases}$$

When the initial wealth is not enough, $b < k_j + y_j$, the borrowing constraint (BC) is binding. Then intertemporal consumption allocation is distorted: $c_1 = b - k_j < c_2 = y_j$.

The schooling choice problem is

$$U(b) \equiv \max_{j \in \mathcal{J}} \{U_j(b)\}. \quad (2)$$

The following lemma shows the relationship between parental transfers and schooling choice. All proofs are in Appendix B.

Lemma 1. *If $y_{j'} \geq y_j$ for all $j \in \mathcal{J}$ and $j' \in \mathcal{J}$ such that $j' > j$, then $\arg\max_{j \in \mathcal{J}} \{U_j(b)\}$ is increasing¹⁰ in b .*

With low initial wealth, some education choices are not feasible, and even when they are feasible, the utility of choosing that option is low, because students suffer from low consumption during schooling. Additional parental transfer affects the schooling choice by making more schooling options feasible, and also by making feasible options more valuable.

3.2 Measuring Distortions

There are two types of inefficiencies generated by borrowing constraints. First, it reduces aggregate consumption as constrained students may not choose the income-maximizing schooling levels. Second, it also distorts intertemporal consumption allocation for constrained students. Therefore the same level of consumption is less valued. To operationalize this idea, I construct a measure of the distortions expressed in consumption units. The money value of distortions are measured by individuals' willingness to pay to eliminate the distortions. It is analogous to risk premium that measures maximum willingness to pay to eliminate uncertainty. Later I will assume that the social objective is to minimize aggregate distortions. This implies that inequality per se does not enter the social objective, therefore separating efficiency from inequality. This idea is similar to Bénabou (2002).

Since the distortions are measured in dollars, I first define a transformation that turns utils into dollars. I first define the money-metric utility function, a monotonic transformation of the utility function (1) that expresses the utility levels in monetary units. The money metric utility function $\underline{C}(U)$ is defined as the minimum cost to achieve the utility level U :

$$\underline{C}(U) \equiv \min_{\tilde{c}_1 \geq 0, \tilde{c}_2 \geq 0} \left\{ \tilde{c}_1 + \tilde{c}_2 \mid \ln \tilde{c}_1 + \ln \tilde{c}_2 \geq U \right\} = 2 \exp \left(\frac{U}{2} \right). \quad (3)$$

Consider the transformed utility function $\underline{C}(\ln c_1 + \ln c_2)$. Since $\underline{C}(U)$ is a monotonic transformation of U , $\underline{C}(\ln c_1 + \ln c_2)$ represents the same preferences as $\ln c_1 + \ln c_2$. The transformed utility function $\underline{C}(\ln c_1 + \ln c_2) = 2\sqrt{c_1 c_2}$ is the same as the total consumption $c_1 + c_2$ if $c_1 = c_2$ and strictly smaller than $c_1 + c_2$ if $c_1 \neq c_2$, reflecting the welfare loss due to the intertemporal consumption distortion. Because individuals are indifferent between a potentially distorted consumption profile (c_1, c_2) and a undistorted consumption profile $(\sqrt{c_1 c_2}, \sqrt{c_1 c_2})$, they are willing to pay as much as $c_1 + c_2 - 2\sqrt{c_1 c_2}$ to eliminate the consumption fluctuation.

Let $V_j(b) \equiv \underline{C}(U_j(b))$ be the value of schooling j , measured in consumption units, and let $V_j^*(b) \equiv y_j - k_j + b$ be the actual total consumption conditional on j . Then $V_j^*(b) - V_j(b)$ is the maximum willingness to pay to eliminate the intertemporal consumption distortion, therefore it measures the magnitude of intertemporal consumption distortion associated with schooling j . The distortion in intertemporal consumption allocation is

¹⁰Increasing/decreasing implies weakly increasing/decreasing throughout the paper.

defined as the average of $V_j^*(b) - V_j(b)$ over all j :

$$\tau_c(b) \equiv \sum_{j \in \mathcal{J}} p_j(b) [V_j^*(b) - V_j(b)],$$

where $\mathbf{p}(b) \in \arg\max_{\mathbf{p} \in \Pi} \sum_{j \in \mathcal{J}} p_j V_j(b) = \arg\max_{\mathbf{p} \in \Pi} \sum_{j \in \mathcal{J}} p_j U_j(b)$ is the utility-maximizing lottery over $\mathcal{J}$ that belongs to the probability simplex $\Pi \equiv \{\mathbf{p} \in \mathbb{R}_+^{J+1} \mid \sum_{j \in \mathcal{J}} p_j = 1, p_j \geq 0, \forall j \in \mathcal{J}\}$. When (BC) does not bind (i.e., $b \geq k_j + y_j$) for all $j \in \mathcal{J}_+$ such that $p_j(b) > 0$, then $V_j^*(b) = V_j(b)$ hold and $\tau_c(b) = 0$. On the other hand, when $b < k_j + y_j$ for some $j \in \mathcal{J}_+$ with $p_j(b) > 0$, then (BC) binds and $\tau_c(b) > 0$. Moreover, the next lemma shows that, holding schooling choice $\mathbf{p}(b)$ constant, the intertemporal distortion can be reduced by additional wealth.

Lemma 2. (i) $dV_j(b)/db \geq dV_j^*(b)/db = 1$, where the inequality is strict if and only if (BC) binds; (ii) $dV_j(b)/db$ is decreasing in b ; (iii) $dV_j(b)/db$ is increasing in j if y_j and k_j are increasing in j .

Similarly, the distortion in schooling is defined as

$$\tau_s(b) \equiv \sum_{j \in \mathcal{J}} [p_j^* - p_j(b)] (y_j - k_j),$$

where $\mathbf{p}^* \in \arg\max_{\mathbf{p} \in \Pi} \sum_{j \in \mathcal{J}} p_j (y_j - k_j)$ is the choice probability vector without the borrowing constraints (BC). Notice that the schooling distortion $\tau_s(b)$ measures the lost total income due to deviation from the income-maximizing choice $\mathbf{p}^*$. When $b > k_j + y_j$ for all $j \geq 1$ with $p_j(b) > 0$, then the schooling choice does not depend on parental transfer, $p_j(b) = p_j^*$, and $\tau_s(b) = 0$.

By adding these two distortions, we get

$$\tau(b) \equiv \tau_c(b) + \tau_s(b) = \sum_{j \in \mathcal{J}} [p_j^* V_j^*(b) - p_j(b) V_j(b)],$$

which can be rewritten as

$$V(b) = \max_{j \in \mathcal{J}} \{y_j - k_j\} + b - \tau(b), \quad (4)$$

where $V(b) \equiv \underline{C}(U(b))$. According to this formula, the money-metric indirect utility can be decomposed into three components: (i) the amount of parental transfer, (ii) maximum income net of education cost, and (iii) distortions due to borrowing constraints. Additional parental transfer directly increases youth's welfare, but as the next lemma shows, it also reduces the distortion for constrained youth, which makes them even better off.

Lemma 3. $\tau(b)$ is decreasing in b .

The implication of this lemma is that, when the average amount of parental transfer is large enough to make everyone unconstrained, then fully equalizing the amount of parental transfers across students will lead to the maximum social welfare (as measured by aggregate money metric indirect utility), as the distortions can be completely eliminated. This argument provides a case for the need-based financial aid, which gives larger amount of subsidy to students who are likely to receive smaller amount of parental transfers, even when there is no explicit social preferences for equality of economic resources across individuals. Next, I characterize education subsidy policies that minimize aggregate distortions.

3.3 Planning Problem

There exists a measure 1 of youth. They are endowed with different amount of parental transfers, but they are identical in all other aspects. Let $\mathcal{I}$ be the set of youth types. Youth of type $i \in \mathcal{I}$ are endowed with parental transfer $b_i \geq 0$. Let $f_i > 0$ be the measure of type i youth. The planner is endowed with budget $E > 0$ that is designated to be used as education subsidy. The planner distributes the budget by assigning different amount of subsidy $g_j \geq 0$ for $j \in \mathcal{J}_+$. Since the budget cannot be spent on those who do not attend college, $g_0 = 0$ is imposed. Let $\mathbf{g} = (g_0, g_1, \dots, g_J)$ be the menu of subsidies, which I call a ‘policy’. Let $\Gamma \equiv \{\mathbf{g} \in \mathbb{R}_+^{J+1} | g_j \geq 0, \forall j \in \mathcal{J}, g_0 = 0\}$ be the set of policies satisfying the above constraints.

The optimal allocation solves the following planning problem.

Problem 1.

$$\min_{\mathbf{g} \in \Gamma, \{\mathbf{p}(b_i)\}_{i \in \mathcal{I}}} \sum_{i \in \mathcal{I}} f_i \tau(b_i, \mathbf{g}) \quad (5)$$

$$\text{subject to } \sum_{i \in \mathcal{I}} f_i \sum_{j \in \mathcal{J}} p_j(b_i) g_j \leq E, \quad (6)$$

$$\mathbf{p}(b_i) \in \operatorname{argmax}_{\mathbf{p} \in \Pi} \sum_{j \in \mathcal{J}} p_j V_j(b_i + g_j), \quad (7)$$

where $\tilde{\mathcal{J}} \subseteq \mathcal{J}$, $\tau(b, \mathbf{g}) \equiv \tau_c(b, \mathbf{g}) + \tau_s(b, \mathbf{g})$, $\tau_c(b, \mathbf{g}) \equiv \sum_{j \in \mathcal{J}} p_j(b) [V_j^*(b + g_j) - V_j(b + g_j)]$, and $\tau_s(b, \mathbf{g}) \equiv \sum_{j \in \mathcal{J}} [p_j^* - p_j(b)](y_j - k_j)$.

The optimal allocation is a policy $\mathbf{g}$ and a set of choice probabilities $\{\mathbf{p}(b_i)\}_{i \in \mathcal{I}}$, which is consistent with individual choices (7), that minimizes aggregate distortions (5) subject to the planner’s budget constraint (6). The planner directly chooses the individual decision rule because some schooling options may lead to different level of distortions or spending even when individuals are indifferent between them. Also notice that (7) allows for the possibility that individuals’ choice set is restricted, as I consider the case where everyone attends college in Section 3.4.

Problem 1 is equivalent to maximizing aggregate money metric welfare $\sum_{i \in \mathcal{I}} f_i V(b_i)$ if the planner’s budget constraint (6) is binding. This can be shown by rewriting the aggregate welfare as follows:

$$\sum_{i \in \mathcal{I}} f_i V(b_i) = \max_{j \in \mathcal{J}} \{y_j - k_j\} + \sum_{i \in \mathcal{I}} f_i b_i + \sum_{i \in \mathcal{I}} f_i \sum_{j \in \mathcal{J}} p_j(b_i) g_j - \sum_{i \in \mathcal{I}} f_i \tau(b_i, \mathbf{g}), \quad (8)$$

which, compared to (4), additionally includes the amount of total education subsidy received. When the planner’s budget constraint (6) does not bind, however, Problem 1 is different from the welfare maximization problem: if two allocations generate the same welfare but different level of aggregate distortions and spending, then the allocation with lower aggregate spending also has lower distortions, thus it is better according to the social objective. I discuss this in more detail in Section 3.5.

3.4 Negative Optimal Marginal Subsidy

I first show that the amount of optimal subsidy is decreasing in schooling level, that is, the marginal subsidy rate is negative. I prove this in a simple environment where there are two types of youth and two levels of schooling.

I first make an assumption on the costs and benefits of education.

Assumption 1. $\mathcal{J} = \{0, 1, 2\}$ and $0 < k_2 - k_1 < (y_2 - y_1)y_1/y_2$.

This assumption implies $k_1 < k_2$, $y_1 < y_2$, and $y_1 - k_1 < y_2 - k_2$, meaning higher level of schooling is more profitable and more expensive. Therefore, in the absence of subsidy, everyone would choose schooling 2 if they could borrow enough, but the ability to choose the option may be limited by the borrowing constraints and the amount of parental transfers.

Lemma 4. If Assumption 1 holds, then $\underline{b}_2 = (y_2 k_2 - y_1 k_1)/(y_2 - y_1) \in (k_2, k_1 + y_1)$ uniquely solves $V_1(\underline{b}_2) = V_2(\underline{b}_2)$.

Due to borrowing constraints, the amount of parental transfer affect schooling choices. Those with $b < \underline{b}_2$, who are constrained at both $j = 1$ and $j = 2$, prefer $j = 1$ while those with $b > \underline{b}_2$ prefer $j = 2$. Assumption 1 implies that the marginal return to $j = 2$ is large enough to ensure some individuals choose $j = 2$ even though they are constrained at both $j = 1$ and $j = 2$.

Assumption 2. $\mathcal{I} = \{L, H\}$ with $b_L < \underline{b}_2$ and $b_H \geq k_2 + y_2$.

This assumption implies that, without a subsidy policy (i.e., $\mathbf{g} = \mathbf{0}$), those with lower parental transfer b_L are constrained at both $j = 1$ and $j = 2$, and choose $j = 1$, while those with higher parental transfer b_H are unconstrained at both choices and choose $j = 2$. The different education choices made by those with different amount of parental transfers makes the differential education subsidy by schooling level a useful policy instrument. By giving different amount of subsidy for different schooling levels, the planner can effectively target the subpopulation for whom subsidy is most useful. The next assumption ensures that both types still make different educational choices under the optimal policy because of limited budget.

Assumption 3. $V_1(b_L + E) > V_2(b_L + E)$.

When there is enough budget, this assumption is not satisfied (i.e., $V_1(b_L + E) \leq V_2(b_L + E)$), and it is efficient to make everyone attain $j = 2$ by setting, for example, $g_1 = g_2 = E$. When Assumption 3 holds, we can avoid the trivial case where both types make identical schooling choice. Given Assumptions 1 and 2, Assumption 3 holds if and only if $k_1 < b_L + E < \underline{b}_2$.

Proposition 1. Suppose that $(\hat{\mathbf{g}}, \{\hat{\mathbf{p}}(b_i)\}_{i \in \mathcal{I}})$ solves Problem 1 with $\tilde{\mathcal{J}} = \mathcal{J}_+$ and that Assumptions 1-3 hold. Then: (i) $\hat{p}_1(b_L) = \hat{p}_2(b_H) = 1$ and $\hat{g}_1 > \hat{g}_2$; (ii) If $y_1 + k_1 - b_L \geq \min\{E/f_L, E + f_H[y_2 - y_1 - (k_2 - k_1)]\}$, then $\hat{g}_1 = \min\{E/f_L, E + f_H[y_2 - y_1 - (k_2 - k_1)]\}$ and $\hat{g}_2 = \max\{0, E - f_L[y_2 - y_1 - (k_2 - k_1)]\}$.

The optimal policy does not lead to different education choices than the ones without subsidy policy, because, due to Assumption 3, it is inefficient to let the type L to choose schooling 2 even when it is possible to do so.¹¹ However, the distortion in intertemporal consumption allocation can be reduced by subsidizing lower schooling level, which is chosen by constrained students with lower parental transfers. To see this, suppose that $\mathbf{g} = \mathbf{0}$ and consider the effects of raising g_1 and g_2 on aggregate distortions. Because the allocation of those with higher parental transfers is undistorted even without subsidy (i.e., $\tau(b_H, \mathbf{g})|_{\mathbf{g}=\mathbf{0}} = 0$), subsidizing schooling 2 does not affect aggregate distortions. On the other hand, the distortion of those with lower parental transfers is reduce by additional subsidy on schooling 1:

$$\left. \frac{\partial \tau(b_L, \mathbf{g})}{\partial g_1} \right|_{\mathbf{g}=\mathbf{0}} = \frac{\partial}{\partial g_1} \left[V_1^*(b_L + g_1) - V_1(b_L + g_1) \right] \Big|_{\mathbf{g}=\mathbf{0}} = 1 - \left. \frac{\partial V_1(b_L + g_1)}{\partial g_1} \right|_{\mathbf{g}=\mathbf{0}} < 0,$$

¹¹Suppose that $V_1(b_L) \leq V_2(b_L + E)$. Then the social planner can make the type L to choose schooling 2 by setting $g_1 = 0$ and $g_2 = E$. However, this policy is dominated by $g_1 = g_2 = E$ because $V_2(b_L + E) < V_1(b_L + E)$.

where the inequality follows from Lemma 2. Therefore, it is optimal to raise g_1 to its limit, $g_1 = E/f_L$, as long as the type L remains constrained the type H still prefers schooling 2. When the latter constraint is binding, as result (ii) shows, an increase in g_1 need to be accompanied by an increase in g_2 in order to prevent the type H from choosing schooling 1.

Corollary 1. *Suppose that $(\hat{g}, \{\hat{p}(b_i)\}_{i \in I})$ solves Problem 1 with $\tilde{\mathcal{J}} = \mathcal{J}_+$ and that Assumptions 1-3 hold. If $y_1 + k_1 - b_L \geq \min \{E/f_L, E + f_H[y_2 - y_1 - (k_2 - k_1)]\}$, then $\hat{g}_1$, $\hat{g}_2$, and $\hat{g}_1 - \hat{g}_2$ are decreasing in f_L .*

Since the type L students are targeted for subsidy, higher fraction of the type L students lowers the amount of subsidy received per student. This implies that, the inequality of parental transfers within income group also affects how the optimal subsidy varies across income groups. Suppose that higher income parents are more likely to give higher transfers, that is, f_L is lower for students from higher income families. If the social planner spends the same amount of budget for low and high income groups, then this corollary implies that students from low income family receive larger amount of subsidies, and the difference of subsidy between low and high schooling level is also larger for higher income groups. This could lead to non-monotonic subsidy schedule across income groups as we see in the quantitative exercise of Section 6.

3.5 Incorporating Heterogeneous Returns to Schooling

In this subsection, I characterize the optimal policy when individuals choose different level of schooling due to differences in return to schooling. The analysis provides a clear comparison with the case when differences in parental transfers are the reason behind differences in education choice. [show that underlying source of heterogeneous schooling choice matters for optimal policy] Also, this bridges a link to the quantitative model that incorporates heterogeneous returns. In this subsection, I focus on the case where the differential return to schooling is non-pecuniary, or ‘psychic’.¹² In the quantitative model, I incorporate differential monetary return to schooling as well. They have similar implications as shown by old working paper.

Let ε_j be the psychic return for schooling $j \in \mathcal{J}$ and let $\varepsilon = (\varepsilon_0, \varepsilon_1, \dots, \varepsilon_J)$ be the vector of psychic returns. Then schooling choice for those with parental transfer b and psychic return ε is

$$\max_{j \in \mathcal{J}} \{V_j(b + g_j) + \varepsilon_j\}.$$

Note that, since $V_j(b + g_j)$ is measured in dollars, the additive psychic return ε_j is also expressed in dollars. Moreover, when (BC) does not bind for all $j \in \mathcal{J}_+$, $V_j(b + g_j) = b + V_j(g_j)$ holds, so the amount of parental transfers does not affect schooling decisions without binding borrowing constraints.¹³

The distortions can be defined similarly as before, but it is convenient to define average distortions over ε . First, let $F(\varepsilon)$ be the joint cumulative distribution function of ε and define $p_j(b, g)$ to be the fraction of

¹²Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005) find substantial psychic returns to schooling. See Heckman, Lochner, and Todd (2006) for a survey on this issue. The psychic returns are commonly assumed in models of schooling choices (e.g., Heckman, Lochner, and Taber, 1998; Bils and Klenow, 2000; Restuccia and Vandenbroucke, 2013).

¹³With the additive psychic returns, the money-metric transformation $\underline{C}(\cdot)$ matters for the schooling choice. That is, $\arg\max_{j \in \mathcal{J}} \{V_j(b + g_j) + \varepsilon_j\} = \arg\max_{j \in \mathcal{J}} \{U_j(b + g_j) + \varepsilon_j\}$ does not necessarily hold. Without the money-metric transformation (i.e., schooling choice is given by $\max_{j \in \mathcal{J}} \{U_j(b + g_j) + \varepsilon_j\}$), the curvature of the indirect utility function $U_j(\cdot)$ generates wealth effects even without binding borrowing constraints. See Caucutt, Lochner, and Park (2017) for discussion.

individuals with b choosing j :

$$p_j(b, \mathbf{g}) \equiv \int \mathbb{I}_{j \in \arg\max_{j' \in \mathcal{J}} \{V_{j'}(b + g_{j'}) + \varepsilon_{j'}\}} dF(\boldsymbol{\varepsilon}), \quad (9)$$

where $\mathbb{I}_x$ is the indicator function with $\mathbb{I}_x = 1$ if x is true and $\mathbb{I}_x = 0$ otherwise. Then, conditional on $(b, \mathbf{g})$, the average intertemporal consumption distortion is

$$\tau_c(b, \mathbf{g}) \equiv \sum_{j \in \mathcal{J}} p_j(b, \mathbf{g}) [V_j^*(b + g_j) - V_j(b + g_j)],$$

and the average schooling distortion is

$$\tau_s(b, \mathbf{g}) \equiv \sum_{j \in \mathcal{J}} \int \left(\mathbb{I}_{j \in \arg\max_{j' \in \mathcal{J}} \{y_{j'} - k_{j'} + \varepsilon_{j'}\}} - \mathbb{I}_{j \in \arg\max_{j' \in \mathcal{J}} \{V_{j'}(b + g_{j'}) + \varepsilon_{j'}\}} \right) (y_j - k_j + \varepsilon_j) dF(\boldsymbol{\varepsilon}).$$

Finally, by adding the two distortions, the average total distortion $\tau(b, \mathbf{g}) \equiv \tau_c(b, \mathbf{g}) + \tau_s(b, \mathbf{g})$ is given by

$$\tau(b, \mathbf{g}) = \int \max_{j \in \mathcal{J}} \{y_j - k_j + \varepsilon_j\} dF(\boldsymbol{\varepsilon}) + b + \sum_{j \in \mathcal{J}} p_j(b, \mathbf{g}) g_j - V(b, \mathbf{g}), \quad (10)$$

where $V(b, \mathbf{g}) \equiv \int \max_{j \in \mathcal{J}} \{V_j(b + g_j) + \varepsilon_j\} dF(\boldsymbol{\varepsilon})$ is the average money-metric utility conditional on $(b, \mathbf{g})$. Notice that, compared to the definition in Section 3.3, the distortion $\tau(b, \mathbf{g})$ additionally includes the lost psychic returns due to the schooling choice that does not maximize the sum of monetary and psychic returns $y_j - k_j + \varepsilon_j$.

The social planner cannot observe individual values of $\boldsymbol{\varepsilon}$, therefore can only condition subsidy amounts on observable education choice. The optimal policy solves the following problem:

Problem 2.

$$\begin{aligned} & \min_{\mathbf{g} \in \Gamma} \sum_{i \in \mathcal{I}} f_i \tau(b_i, \mathbf{g}) \\ & \text{subject to } \sum_{i \in \mathcal{I}} f_i \sum_{j \in \mathcal{J}} p_j(b_i, \mathbf{g}) g_j \leq E, \end{aligned} \quad (11)$$

where $p_j(b, \mathbf{g})$ and $\tau(b, \mathbf{g})$ are given by (9) and (10).

Problem 2 is similar to Problem 1, but it is more complicated due to the integrals over $\boldsymbol{\varepsilon}$. For analytical tractability, I assume that the psychic returns $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_J$ are identically and independently distributed with standard Gumbel (or type I extreme value) distribution, that is, $\ln F(\boldsymbol{\varepsilon}) = -\sum_{j \in \mathcal{J}} \exp(-\varepsilon_j)$. Then, as shown by McFadden (1978), $p_j(b, \mathbf{g}) = \exp(V_j(b + g_j)) / \sum_{j' \in \mathcal{J}} \exp(V_{j'}(b + g_{j'}))$ and $V(b, \mathbf{g}) = \ln(\sum_{j \in \mathcal{J}} \exp(V_j(b + g_j))) + \psi$ hold, where $\psi \approx 0.57721$ is the Euler-Mascheroni constant. These formulas are also useful for quantitative analysis in Section 4, since they make numerical integration over $\boldsymbol{\varepsilon}$ unnecessary.

Homogeneous Parental Transfers I begin with the case where everyone receive the same amount of parental transfers. Then individuals are heterogeneous only in the psychic returns to schooling. This case permits sharp characterization of the optimal policy.

Proposition 2. Suppose that $b_i = b$ for all $i \in \mathcal{I}$ and $\hat{\mathbf{g}}$ solves Problem 2. Then: (i) $\hat{g}_{j'} \geq \hat{g}_j$ for all $j \in \mathcal{J}_+$ and

$j' \in \mathcal{J}_+$ such that $y_{j'} \geq y_j$ and $k_{j'} \geq k_j$; (ii) For all $j \in \mathcal{J}_+$ such that $\hat{g}_j > 0$,

$$\hat{g}_j = \sum_{j' \in \mathcal{J}} p_{j'}(b, \hat{\mathbf{g}}) \hat{g}_{j'} + \frac{1}{1 + \lambda} - \frac{1}{V_j'(b + \hat{g}_j)}, \quad (12)$$

where λ is the Lagrangian multiplier on the constraint (11) and $V_j'(\cdot)$ is the derivative of $V_j(\cdot)$; (iii) If $b \geq y_j + k_j - \hat{g}_j$ for all $j \in \mathcal{J}_+$, then $\hat{\mathbf{g}} = \mathbf{0}$ and $\lambda = 0$.

The formula (12) shows that the optimal subsidy for schooling $j \in \mathcal{J}_+$ is determined by the level of aggregate spending, the social marginal value of public funds, captured by the Lagrangian multiplier on the planner's budget constraint (11), and the marginal value of wealth for schooling j . Importantly, (12) shows that the amount of optimal subsidy is larger for schooling level with higher marginal value of wealth, where additional subsidy has greater impact. As shown by Lemma 2, the marginal value of wealth for higher level of schooling is larger when y_j and k_j are increasing in j , implying that the subsidy amount for higher level of schooling is larger.

For a given level of parental transfers, borrowing constraints are more likely to bind for those attending higher schooling level with larger $y_j + k_j$, because they pay higher cost during schooling and they earn more after schooling. In contrast to the case with heterogeneous parental transfers, those attain more schooling are more likely to be constrained, thus giving larger subsidy to them reduces aggregate distortions. [consistent with current policy] As shown by De Fraja (2002), this result carries over to the case with heterogeneous monetary returns to schooling: those with higher monetary returns are more likely to attain more schooling and be constrained (Lochner and Monge-Naranjo, 2011), so it is efficient to more heavily subsidize higher level of schooling to help the constrained individuals with high returns.¹⁴ The quantitative model in Section 4 incorporates both heterogeneous monetary and psychic returns.

Result (iii) highlight the goal of the subsidy policy and also points out the inherently distortionary nature of the subsidy policy. It shows that, the optimal subsidy never makes everyone unconstrained even when there is enough budget to do so. When nobody is constrained even without subsidy, any subsidy policy would only distort schooling choices, thus it is efficient not to spend anything on education subsidy. Even when some individuals are constrained in the absence of subsidy, fully eliminating intertemporal consumption distortions through subsidy is never optimal because it creates larger schooling distortions. This implies that, as shown by the following corollary, there exists a maximum amount of spending that minimizes aggregate distortions, and thus the planner's budget constraint does not bind for the budget exceeding the maximum spending.

Corollary 2. Suppose that $b_i = b$ for all $i \in \mathcal{I}$. Then: (i) there exists $\bar{E} < \infty$ such that the planner's budget constraint (11) binds if and only if $E < \bar{E}$; (ii) $\bar{E} = 0$ if $b \geq \max_{j \in \mathcal{J}_+} \{y_j + k_j\}$; (iii) $0 < \bar{E} < \max_{j \in \mathcal{J}_+} \{y_j + k_j\} - b$ if $b < \max_{j \in \mathcal{J}_+} \{y_j + k_j\}$.

Distributing education subsidy reduces aggregate distortions if and only if there are some constrained individuals in the absence of the subsidy. Moreover, the upper bound for the maximum spending, which is the minimum spending that would make everyone unconstrained, is lower for larger parental transfers, reflecting larger need for budget for those with lower parental transfers. This could justify the need-based financial aid system when parental transfers are perfectly correlated with parental income.

¹⁴This force can be offset when there is a strong social preference for equity because those with higher returns earn more and are better off (Gottlieb and Moreira, 2012).

Heterogeneous Parental Transfers With heterogeneity in parental transfers as well as returns to schooling, those with different amount of parental transfers may choose the same schooling levels. In this case, we can derive a formula similar to (12), aggregated over parental transfers within each schooling level.

Corollary 3. Suppose that $\hat{\mathbf{g}}$ solves Problem 2. Then for all $j \in \mathcal{J}_+$ such that $\hat{g}_j > 0$,

$$\hat{g}_j = \sum_{i \in \mathcal{I}} \hat{q}_{i|j} \sum_{j' \in \mathcal{J}} p_{j'}(b_i, \hat{\mathbf{g}}) \hat{g}_{j'} + \frac{1}{1 + \lambda} - \frac{1}{\sum_{i \in \mathcal{I}} \hat{p}_{i|j} V_j'(b_i + \hat{g}_j)}, \quad (13)$$

where $\hat{p}_{i|j} \equiv f_i p_j(b_i, \hat{\mathbf{g}}) / \sum_{i' \in \mathcal{I}} f_{i'} p_j(b_{i'}, \hat{\mathbf{g}})$ and $\hat{q}_{i|j} \equiv f_i p_j(b_i, \hat{\mathbf{g}}) V_j'(b_i + \hat{g}_j) / \sum_{i' \in \mathcal{I}} f_{i'} p_j(b_{i'}, \hat{\mathbf{g}}) V_j'(b_{i'} + \hat{g}_j)$.

The formula (13) shows that the *conditional average* of the marginal value of wealth matters for the differential subsidy amounts across schooling levels. However, it is not clear how the average marginal value of wealth varies with schooling levels. For a given level of parental transfer, those choosing higher levels of schooling are more likely to be constrained and have higher marginal value of wealth. However, with binding borrowing constraints, those attaining higher level of schooling are more likely to have received larger parental transfers, thus have lower marginal value of wealth. Therefore, how the average marginal value of wealth, and therefore the subsidy amount, varies with schooling levels depend on which of these two forces dominates, which is quantified in the next section.

4 Quantitative Model

I now explore the quantitative implications of heterogeneous parental transfers for the design of education subsidies using a more general framework. I first extend the model to accommodate multi-period schooling and post-schooling period as well as heterogeneous monetary returns to schooling. Moreover, I endogenize parental transfers by explicitly modeling parent's behavior.

Then I calibrate preferences, costs and returns to college education by student ability and family income to match both the distribution of schooling and the relationship between schooling, ability, and family income in the U.S. I then characterize the optimal policy for different income groups.

Time is discrete and the a model period represents a calendar year. A family consists of a parent and a child (youth). In $t = 0$, parents give one-time transfer and, taking transfers as given, youth make schooling choice. The child lives until time $t = T_k$ and the parent lives until $t = T_p$.

4.1 Intertemporal Consumption Choice

Youth accrues a flow of utility from consumption c_t in each period t and they discount future utility flows at a subjective discount factor β . The lifetime utility from a consumption profile $\{c_t\}_{t=1}^{T_k}$ is

$$\sum_{t=1}^{T_k} \beta^t u(c_t), \quad (14)$$

where $u(c) = (c^{1-1/\gamma} - 1)/(1 - 1/\gamma)$ and $\gamma \in \mathbb{R}_+$ is the elasticity of intertemporal substitution (EIS).

Schooling options $j \in \mathcal{J}$ represent years of college education. Given annual gross interest rate $R = 1/\beta$, the

lifetime budget constraint for those who complete j years of college is

$$\sum_{t=1}^{T_k} R^{-t} c_t \leq b_j + y_j(a) - k_j + g_j, \quad (15)$$

where $b_j \geq 0$ is the amount of parental transfers conditional on schooling j and $y_j(a)$ is lifetime earnings for those with schooling j and ‘ability’ a . I assume that $y_j(a)$ is increasing in a . For those who do not attend college ($j = 0$), I keep assuming that there are no monetary costs, that is, $k_j = g_j = 0$ for $j = 0$, but b_j may not be zero for those.

Those who attend college ($j \in \mathcal{J}_+$) additionally face borrowing constraints:

$$\sum_{t=1}^j R^{-t} c_t + k_j - g_j - b_j \leq \bar{d}_j, \quad (16)$$

where $\bar{d}_j$ is the borrowing limit that may vary with schooling option.

Let $U_j(a, b_j + g_j)$ be the indirect utility function conditional on choice j and ability a , which is the lifetime utility (14) maximized subject to the lifetime budget constraint (15) as well as borrowing constraints (16) for $j \in \mathcal{J}_+$. As before, I define $U_j(a, b_j + g_j) = -\infty$ if the option j is not feasible, and let $V_j(a, b_j + g_j) \equiv \underline{C}(U_j(a, b_j + g_j))$ be the money metric indirect utility, where $\underline{C}(U)$ is defined similarly as (3):

$$\underline{C}(U) \equiv \min_{\{\tilde{c}_t\}_{t=1}^{T_k}} \left\{ \sum_{t=1}^{T_k} R^{-t} \tilde{c}_t \mid \sum_{t=1}^{T_k} \beta^t u(\tilde{c}_t) \geq U, \tilde{c}_t \geq 0 \forall t \right\} = \sum_{t=1}^{T_k} R^{-t} u^{-1} \left(\frac{U}{\sum_{t=1}^{T_k} \beta^t} \right).$$

I also define $V_j^*(a, b_j + g_j) \equiv y_j(a) - k_j + b_j + g_j$ to be the money metric indirect utility in the absence of the borrowing constraints. We can show a version of Lemma 2 for this general model.

Lemma 5. *For $j \in \mathcal{J}_+$, $\partial V_j(a, z)/\partial z \geq \partial V_j^*(a, z)/\partial z = 1$ holds and the inequality is strict if and only if (16) binds. Moreover, $\partial V_j(a, z)/\partial z$ is increasing in a .*

4.2 Education Choice

For a given policy $\mathbf{g}$, the education choice for youth with schooling returns $(a, \boldsymbol{\varepsilon})$ and parental transfer $\mathbf{b} = (b_0, b_1, \dots, b_J)$ is

$$\max_{j \in \mathcal{J}} \{V_j(a, b_j + g_j) + \mu_j + \sigma \varepsilon_j\},$$

where μ_j and σ are the location and scale parameters of the distribution of psychic returns for option $j \in \mathcal{J}$. Then, conditional on $(a, \mathbf{b}, \mathbf{g})$, the fraction of youth choosing option $j \in \mathcal{J}$ is

$$p_j(a, \mathbf{b}, \mathbf{g}) \equiv \int \mathbb{I}_{j \in \arg \max_{j' \in \mathcal{J}} \{V_{j'}(a, b_{j'} + g_{j'}) + \mu_{j'} + \sigma \varepsilon_{j'}\}} dF(\boldsymbol{\varepsilon}), \quad (17)$$

and the average utility is

$$V(a, \mathbf{b}, \mathbf{g}) \equiv \int \max_{j \in \mathcal{J}} \{V_j(a, b_j + g_j) + \mu_j + \sigma \varepsilon_j\} dF(\boldsymbol{\varepsilon}). \quad (18)$$

The location parameter μ_j captures the component of psychic returns common to all youth with different psychic returns ε . As I assume $\mu_0 = 0$, μ_j for $j \in \mathcal{J}_+$ measures the common psychic returns relative to not attending college, which could be positive or negative. The scale parameter $\sigma \geq 0$ determines the degree of heterogeneity in psychic returns relative to monetary returns in determining schooling choices. When $\sigma = 0$, all youth with identical $(a, \mathbf{b}, \mathbf{g})$ choose the same level of schooling, therefore, differences in schooling choice is entirely driven by differences in monetary returns. On the other hand, when $\sigma \rightarrow \infty$, the heterogeneity in psychic returns dominate monetary returns, and all options are equally likely to be chosen. The location and scale parameters will be important in fitting the distribution over schooling choices as shown in the next subsection.

4.3 Parental Transfer Decision

In $t = 0$, parents are endowed with lifetime wealth $w \in \mathbb{R}_{++}$, which they split between their own consumption and transfer to their children. Parents care about their children's welfare, that is, they are altruistic towards them, but the degree of altruism $\delta \in (0, 1)$ is heterogeneous across parents. When making transfer choice, parents do not know about their children's psychic returns ε , but otherwise they have the same information as their children. Not knowing their children's preferences for education, parents offer and commit to a menu of transfers $\mathbf{b}$.

Problem 3. Consider a family with $(w, \delta, a, \mathbf{g})$. Taking $\mathbf{p}(a, \mathbf{b}, \mathbf{g})$ and $V(a, \mathbf{b}, \mathbf{g})$ as given, the parent solves

$$\begin{aligned} \max_{\mathbf{b}} & \left\{ (1 - \delta) \sum_{t=1}^{T_p} \beta^t v(\underline{c}_p) + \delta \sum_{t=1}^{T_k} \beta^t v(\underline{c}_k) \right\} \\ \text{subject to} & \sum_{t=1}^{T_p} R^{-t} \underline{c}_p = w - \sum_{j \in \mathcal{J}} p_j(a, \mathbf{b}, \mathbf{g}) b_j \end{aligned} \quad (19)$$

$$\begin{aligned} & \sum_{t=1}^{T_k} R^{-t} \underline{c}_k = V(a, \mathbf{b}, \mathbf{g}) \\ & b_j \in [0, w] \quad \forall j \in \mathcal{J}, \end{aligned} \quad (20)$$

where $v(c) = (c^{1-1/\eta} - 1)/(1 - 1/\eta)$.

For a given menu of parental transfer, the schooling option $j \in \mathcal{J}$ is chosen with probability $p_j(a, \mathbf{b}, \mathbf{g})$, which leads to average amount of transfer $\sum_{j \in \mathcal{J}} p_j(a, \mathbf{b}, \mathbf{g}) b_j$ and average money metric utility of the child $V(a, \mathbf{b}, \mathbf{g})$. Although parents face uncertainty about their children's preferences, I assume that parents are risk neutral so that the degree of the uncertainty does not affect their decisions. Specifically, I assume that parents only care about their own average utility and their children's average utility, both measured in consumption units. However, the degree of parental altruism differ across parents, so parents with higher δ put higher weights on their children's utility. Parents choose transfers to smooth resources between themselves and their children, and the desire for smoothing is captured by the elasticity of intergenerational substitution (EGS) $\eta \in \mathbb{R}_+$.

The parents' behavior formulated in Problem 3 is more general than the one typically considered, and can be mapped to more familiar model of [Becker and Tomes \(1986\)](#). In particular, when $\mu_j = 0$ for all $j \in \mathcal{J}$, $\sigma = 0$, and $\gamma = \eta$, the parent's objective function becomes

$$(1 - \delta) \sum_{t=1}^{T_p} \beta^t u(c_{t,p}) + \delta \sum_{t=1}^{T_k} \beta^t u(c_{t,p}), \quad (21)$$

where $c_{t,p}$ and $c_{t,k}$ are actual consumption of the parent and the child in each period. Since (21) is weighted average of each family member's utility from own consumption, the resulting allocation can be seen as an outcome of any collective models commonly considered in family economics (Browning, Chiappori, and Weiss, 2014). The distinction between the attitudes towards intertemporal (γ) and intergenerational (η) consumption smoothing, first made by Córdoba and Ripoll (2014), allows the quantitative model to better capture the empirical relationship between education attainment, parental resources, and youth ability. The roles of the two smoothing parameters, together with borrowing constraints, in determining parental transfer is emphasized in the following proposition.

Proposition 3. *For a family with $(w, \delta, a, \mathbf{g})$, suppose that $\hat{\mathbf{b}}$ solves Problem 3. Then, for $\hat{b}_j > 0$,*

$$\hat{b}_j = \sum_{j' \in \mathcal{J}} \hat{b}_{j'} p_{j'}(a, \hat{\mathbf{b}}, \mathbf{g}) + \sigma \left\{ \left(\frac{\delta}{1 - \delta} \right) \frac{v'(\underline{c}_k)}{v'(\underline{c}_p)} - \frac{1}{V_{j,2}(a, \hat{b}_j + g_j)} \right\}, \quad (22)$$

where $V_{j,2}(\cdot, \cdot)$ is the partial derivative of $V_j(\cdot, \cdot)$ with respect to the second argument, and $\underline{c}_p$ and $\underline{c}_k$ are given by (19) and (20), evaluated at $\mathbf{b} = \hat{\mathbf{b}}$.

The formula is identical to the condition for the optimal policy (12), except for the term for the intergenerational smoothing $\delta v'(\underline{c}_k)/[(1 - \delta)v'(\underline{c}_p)]$ that replaces $1/(1 + \lambda)$. This reflects that, unlike the social planner who cares only about reducing distortions, parents care about welfare level of children, so they are willing to give money even when it does not reduce distortions. To see this, suppose that borrowing constraints (16) are not binding (i.e., $V_{j,2}(a, b_j + g_j) = 1$) for all $j \in \mathcal{J}_+$. Then (22) implies $\hat{b}_j = \hat{b}$ for all $j \in \mathcal{J}$. When $\hat{b} > 0$, it satisfies the first order condition

$$(1 - \delta)v'(\underline{c}_p) = \delta v'(\underline{c}_k),$$

where $\underline{c}_p = (w - \hat{b}) / \sum_{t=1}^{T_p} R^{-t}$ and $\underline{c}_k = (\hat{b} + V(a, \mathbf{0}, \mathbf{g})) / \sum_{t=1}^{T_k} R^{-t}$. This condition, derived by aggregating (22) over $j \in \mathcal{J}$, implies wealthier or more altruistic parents give larger transfer. Youth with higher ability get lower transfer because they earn more and therefore have higher welfare. The gradient of parental transfer with respect to these factors depends on the EGS (η).

When (16) binds (i.e., $V_{j,2}(a, \hat{b}_j + g_j) > 1$) for some $j \in \mathcal{J}_+$, the first order conditions can be rearranged to give:

$$(1 - \delta)v'(\underline{c}_p) = \delta v'(\underline{c}_k) \left(\sum_{j \in \mathcal{J}} \hat{p}_{j|b_{j'} > 0} \frac{1}{V_{j,2}(a, \hat{b}_j + g_j)} \right)^{-1} > \delta v'(\underline{c}_k),$$

where $\hat{p}_{j|b_{j'} > 0} \equiv p_j(a, \hat{\mathbf{b}}, \mathbf{g}) \mathbb{I}_{\hat{b}_j > 0} / \sum_{j' \in \mathcal{J}} p_{j'}(a, \hat{\mathbf{b}}, \mathbf{g}) \mathbb{I}_{\hat{b}_{j'} > 0}$. With binding borrowing constraints, the EIS (γ) also affects the gradient of parental transfers through its effect on the marginal value of wealth $V_{j,2}(a, \hat{b}_j + g_j)$. In contrast to the case where (16) does not bind, higher ability youth may receive higher parental transfers: although they are better off, their allocations are more likely to be distorted by borrowing constraints, as they borrow more for a given level of schooling and also are more likely to choose higher level of schooling. Therefore, when the EGS is large relative to the EIS, parents may give them more to help them pay for education expenses.

The concern about borrowing constraints also motivates parents to give different amount of transfers for

different level of schooling. From (22), the difference between parental transfer for schooling j' and j can be written as

$$\hat{b}_{j'} - \hat{b}_j = \sigma \left\{ \frac{1}{V_{j,2}(a, \hat{b}_j + g_j)} - \frac{1}{V_{j',2}(a, \hat{b}_{j'} + g_{j'})} \right\}.$$

If (16) do not bind for both choices, then $\hat{b}_{j'} = \hat{b}_j$ holds, reflecting the fact that there is no need for parents to distort children's schooling choices because parents are altruistic and not paternalistic.¹⁵ More generally, $\hat{b}_{j'} \geq \hat{b}_j$ holds if and only if $V_{j',2}(a, \hat{b}_{j'} + g_{j'}) \geq V_{j,2}(a, \hat{b}_j + g_j)$. That is, parents give larger transfers for schooling levels for which children's consumption is more intertemporally distorted. This provides a mechanism for marginal education subsidy from parents, which is often taken as exogenously given in empirical studies (Keane and Wolpin, 2001; Johnson, 2013) and interpreted as reflecting paternalistic preferences of parents.¹⁶

5 Calibration

I now discuss the calibration procedure to set the parameter values for quantitative analysis. I first exogenously set the parameters associated with monetary returns to and costs of schooling. Then the remaining preference parameters are calibrated by simulating the model and using data from the NLSY97. All monetary amounts in this and next section are denominated in 2004 U.S. dollars using the consumer price index (CPI-U-RS).

5.1 Parameters Set Externally

The choice set of schooling is $\mathcal{J} = \{0, 1, 2, 4\}$. Therefore, in the model, individuals may choose to attend college or attend for 1, 2, or 4 years. Since some individuals in the data choose schooling levels outside the choice set, each schooling choice is empirically mapped to 12, 13, 14-15, and 16-17 years of completed schooling.

Monetary Returns to Schooling Youth are age 17 in $t = 0$. All individuals work until age 65 and live until age 80. The present discounted value of lifetime after-tax earnings for those choosing $j \in \mathcal{J}$ years of college is

$$y_j(a) = \sum_{x=1}^{65-17-j} R^{-(x+j)} \left[\tilde{y}_j(a, x) - \mathcal{T}(\tilde{y}_j(a, x)) \right],$$

where $\tilde{y}_j(a, x)$ is the before-tax annual earnings for those with schooling j , ability a , and potential work experience x , and $\mathcal{T}(\tilde{y})$ is the amount of income taxes paid. I assume an annual interest rate 3%, which implies $R = 1/\beta = 1.03$, and use the tax function $\mathcal{T}(\tilde{y}) = 0.264[1 - (0.012\tilde{y}^{0.964} + 1)^{-1/0.964}]\tilde{y}$, estimated by Guner, Kaygusuz, and Ventura (2014).¹⁷

Since the NLSY97 cohorts, aged 12–17 in 1997, are still too young for the estimation of the lifecycle earnings profile, the parameters of the earnings function are estimated using data from the 1979 cohorts of the National Longitudinal Surveys of Youth (NLSY79) for years 1979–2012. As a measure of ability, I use quartiles of the

¹⁵For this reason, parents typically make lump-sum, or unconditional, transfers in dynastic models with complete information (e.g., Becker and Tomes, 1986; Caucutt and Lochner, 2012; Abbott et al., 2016; Krueger and Ludwig, 2016).

¹⁶Similar behavior can arise without paternalism. For example, when there exists a repeated interaction between a parent and a child, an opportunistic behavior of the child may lead parents to directly pay for educational expenses instead of giving cash (Brown, Scholz, and Seshadri, 2012).

¹⁷This functional form was first proposed by Gouveia and Strauss (1994).

Armed Forces Qualifying Test (AFQT) scores.¹⁸ I select all individuals from the random sample with at least 12 years of completed schooling and at most 17 years of completed schooling. Those who are enrolled in school are excluded. I regress log annual earnings on indicators for years of schooling and AFQT quartiles, along with a third-order polynomial in experience to estimate the annual earnings $\tilde{y}_j(a, x)$ as a function of years of schooling, ability, and experience. Appendix C shows the OLS estimates of the regression (Table A3) as well as the present discounted value of lifetime earnings $y_j(a)$ (Table A4).

Monetary Costs of Schooling The amounts of monetary cost (k_j) and subsidy (g_j) for each schooling level are computed based on the average annual amounts of tuition cost and grant aid, estimated using data from the 2003–2004 National Post-Secondary Student Aid Survey (NPSAS:04) for full-time, full-year dependent students who applied for federal financial aid.¹⁹ To parsimoniously capture the key features of the need-based financial aid system, I estimate the average amount of grant aid separately by family income quartiles²⁰ and also conduct the optimal subsidy exercise in Section 6 based on differential subsidy schedule across the four groups of parental income. Next, assuming students can only borrow from the government,²¹ I take the cumulative limit, implied by annual limits of the Stafford Loan Program in 2003–2004, as the borrowing limit ($\bar{d}_j$). Tables A5 and A6 in Appendix C show the estimated annual and total amounts.

5.2 Calibrating Preference Parameters

The remaining parameters include the parameters for consumption smoothing (γ, η), distribution of psychic returns to schooling ($\mu_1, \mu_2, \mu_4, \sigma$), and distribution of parental altruism, which is assumed to be a beta distribution with two parameters.²² These parameters are chosen so that the model replicates empirical relationships between education attainment, ability, parental income, and parental transfers from the NLSY97 data. Although it is not possible to provide an analytical proof that the parameters are identified, below I provide an informal argument about how each parameter affects certain features of the data.

The consumption smoothing parameters can be inferred from how education attainment and parental transfer amount vary with parental income and youth ability. As shown in Section 4.3, with binding borrowing constraints, both intertemporal and intergenerational elasticity of substitution affects how the amount of parental transfer varies with parental income and youth ability, but the ability gradient of parental transfer is particularly informative about the relative magnitude between them: a strong and positive ability gradient of parental transfer would suggest a large value of the EGS relative to the EIS. Next, for a given value of the EGS, the gradient of education attainment is informative about the EIS. With a small value of EIS, the intertemporal consumption distortion associated with further education attainment is large, and therefore education attainment depends more strongly on parental income. Moreover, the small EIS also implies that the negative wealth effect of having higher ability is large,²³ which substantially weakens the strong, positive effects of ability on education attainment that would

¹⁸The AFQT test scores are widely used as a measure of cognitive ability. Most respondents took the test as part of the NLSY79 and NLSY97 surveys.

¹⁹Note that the academic year 2003–2004 overlaps with the periods of college education for NLSY97 respondents.

²⁰Since the NPSAS:04 contains only those enrolled in college, its family income distribution is different from that of the NLSY97. To address this concern, I use the parental income quartiles of the NLSY97 as thresholds to divide the NPSAS:04 individuals into four groups.

²¹Despite increasing importance of private credit markets for students in the U.S., private student loan amounts account for less than 20 percent of all student loan dollars distributed in 2004–2005 (Lochner and Monge-Naranjo, 2011).

²²The beta distribution, defined on the unit interval, is very flexible and is commonly used to represent bounded distributions.

²³The negative wealth effect arises because, for a given level of schooling, the intertemporal consumption distortion is larger for higher ability students who have higher future income. As shown by Lochner and Monge-Naranjo (2011), this can generate a negative relationship

prevail in the absence of borrowing constraints.

For a given value of consumption smoothing parameters, the distributions of psychic returns and parental altruism determine the distributions of education attainment and parental transfers, but they also affect the estimated gradients of education attainment and parental transfers, which are attenuated toward zero because the psychic returns and parental altruism, relevant for schooling and parental transfer decisions, are unobservable to the econometrician. Therefore, the degree of heterogeneity in psychic returns and parental altruism, in addition to the consumption smoothing parameters, also affects the estimated gradients.

Based on this argument, I choose 3 sets of target statistics (total 18 targets) that would provide enough information to pin down the remaining 8 parameters: (i) marginal distribution of education attainment; (ii) OLS estimates of the regression of education attainment on parental income quartiles and youth ability quartiles; (iii) OLS estimates of the regression of the probability of receiving more than \$1,000 from parents on parental income quartiles and youth ability quartiles. Note that, of these statistics, there is no explicit link between education attainment and parental transfers, which can be used to evaluate the model's ability to replicate the key empirical correlation.

The target statistics are computed using the NLSY97 data. The NLSY97 is a longitudinal survey of 8,984 Americans from the cohort born between 1980 and 1984. The survey was conducted annually from 1997 through 2011 and biennially since. It contains extensive information on each youth's educational outcomes, together with detailed information about family background. Importantly, there are questions about the amount of parental transfers youth received, which have been recently used to study the role of parental transfers in education choices (e.g., [Johnson, 2013](#); [Abbott et al., 2016](#); [Findeisen and Sachs, 2017](#)). From these questions, I compute the amount of total parental transfers each youth received between ages 18 and 26²⁴, discounted at a 3% annual interest rate back to age 17 (See Appendix ?? for details). To be consistent with the measurement of parental transfers, education outcomes are also measured at age 26.

I exclude youth who are part of the minority and poor white oversamples, using only the full random samples. I also select those with 12–17 years of completed schooling who are never observed to attend graduate school. Individuals are also dropped if any of the variables used (including income and net wealth of parents, education attainment and AFQT score of youth, and amount of parental transfer) is missing.

The parameters are chosen to minimize the sum of squared differences between the statistics based on actual data and simulated data, which is generated by drawing psychic returns and parental altruism and solving the family's problem for each NLSY97 individual.²⁵ Solving the model and comparing it with actual data requires a number of additional assumptions. First, the model in Section 4 assumes that a family consists of a parent and a youth, which does not necessarily hold for all families in the data. To account for these differences, the regressions for the second and third sets of target statistics additionally control for the number of siblings (1, 2, 3+) and an indicator for two-parent families, assuming that these variables affect outcome levels but not gradients of outcomes. Second, I assume a constant stream of parental income to compute the remaining discounted value of lifetime income of parents as of the year when the child is age 17. Then the lifetime wealth of parents (w) is defined as the sum of the remaining lifetime income and initial net wealth of parents.

between education attainment and ability when the EIS is small.

²⁴The questions about annual amount of transfers are available only until 2010, when the youngest cohort was 26 years old.

²⁵To better fit the schooling distribution, the probabilities are multiplied by 10, which is equivalent to putting 100 times higher weights on them.

5.3 Model Fit and Parameter Values

Table 3 reports the calibrated parameter values and Table 4 shows the target statistics, along with the statistics based on the calibrated model. The calibrated value of the EIS is smaller than 1, consistent with most of the estimates reported in [Browning, Hansen, and Heckman \(1999\)](#). As Table 4b shows, the implied intertemporal consumption smoothing motive replicates the parental income and youth ability gradients of education attainment observed in the data reasonably well, with the exception of the coefficient on the second parental income quartile. On the other hand, the calibrated EGS is greater than 1, consistent with [Córdoba and Ripoll \(2014\)](#).²⁶ Since the EGS larger than the EIS, the desire to smooth consumption across generations is weaker than the smoothing motive across time (i.e., within generation), and the model is able to replicate positive relationship between the amount of parental transfers and youth ability, reported in Table 4c, although the relationship in the data is weaker.

The calibrated distribution of parental altruism implies that parents put around 34% of weights on their children's welfare on average, with the standard deviation lower than the mean value. The calibrated distribution of psychic returns to schooling suggests that there are substantial psychic costs of attending college, consistent with other studies in the literature (e.g., [Carneiro, Hansen, and Heckman, 2003](#); [Cunha, Heckman, and Navarro, 2005](#)). The location and scale parameters of the psychic returns are much larger than monetary costs of college (around \$8,000 each year), suggesting that psychic returns are important determinants of schooling choice. However, with the scale parameter much smaller than the location parameters, the degree of heterogeneity in psychic returns is small.

Table 3: Calibrated Preference Parameters

Parameter	Description	Value
Consumption smoothing:		
γ	Elasticity of intertemporal substitution	0.840
η	Elasticity of intergenerational substitution	1.167
Distribution of parental altruism:		
$E(\delta)$	Mean of parental altruism	0.342
$SD(\delta)$	Standard deviation of parental altruism	0.217
Distribution of psychic returns to schooling:		
μ_1	Location parameter of psychic returns for completing 1 year of college	-64,301
μ_2	Location parameter of psychic returns for completing 2 years of college	-65,619
μ_4	Location parameter of psychic returns for completing 4 years of college	-171,295
σ	Scale parameter of psychic returns	15,899

5.4 Assessment of the Calibrated Model

To assess the performance of the calibrated model, I compare some key statistics that were not calibration targets. In particular, I choose the statistics that capture the key economic forces in the model: within-group inequality in parental transfers and correlation between parental transfers and education attainment. I find that the model

²⁶[Córdoba and Ripoll \(2014\)](#) identify the EGS in a different context, using the parental income gradient of fertility in a dynastic model of [Becker and Barro \(1988\)](#).

Table 4: Target Statistics and Model Fit

	0	1	2	4
Data	0.358	0.103	0.185	0.354
Model	0.360	0.104	0.186	0.350

(a) Fraction with Completed Years of College

	Parental Income Quartile			AFQT Quartile			Constant
	2	3	4	2	3	4	
Data	0.027	0.406	0.763	0.539	1.173	1.984	0.581
Model	0.186	0.300	0.700	0.631	1.305	1.889	0.625

(b) Effects on Completed Years of College

	Parental Income Quartile			AFQT Quartile			Constant
	2	3	4	2	3	4	
Data	0.086	0.154	0.282	0.079	0.161	0.240	0.458
Model	0.151	0.211	0.311	0.035	0.059	0.101	0.521

(c) Effects on Probability of Receiving More Than \$1,000 from Parents

does a reasonable job reproducing these statistics from the data, and thus is a good framework for studying the optimal design of education subsidy.

The key heterogeneity in the model is the amount of parental transfers among families with similar economic resources, which is mainly generated by the heterogeneity in parental altruism. Table 5 shows the fraction of youth in each parental transfer quartile, conditional on each parental income quartile. In the data, those with higher parental income quartile are more likely to receive larger amount of parental transfers, but the correlation between parental income and transfer is imperfect, generating the dispersion of parental transfers within parental income group. The model replicates these patterns well.

In the model, the amount of parental transfers matters for schooling choices because of borrowing constraints. The strength of this force would be captured by the correlation between parental transfers and schooling, which is not explicitly targeted for calibration. Table 6 reports the fraction of youth choosing each schooling option as well as average years of schooling, separately for those receiving below and above median amount of parental transfers. In the data, parental transfers are positively correlated with schooling in the sense that those with above median transfers are more likely to attain 2 or 4 years of college and also attain more schooling on average. The model also reproduces this correlation but overstates the effects on the probability of completing 4 years of college.

Table 5: Distribution of Parental Transfers by Parental Income

		% with Parental Transfer			
		Quartile 1	Quartile 2	Quartile 3	Quartile 4
Parental Income:					
Quartile 1	Data	40.7	33.6	18.3	7.4
	Model	40.8	41.4	16.3	1.5
Quartile 2	Data	28.3	30.0	28.2	13.5
	Model	27.8	26.6	28.8	16.8
Quartile 3	Data	22.0	22.2	30.5	25.4
	Model	21.1	18.3	28.6	32.0
Quartile 4	Data	9.1	14.2	23.1	53.7
	Model	10.4	13.6	26.3	49.7

Table 6: Distribution of Years of College Education by Parental Transfers

		% with Years of College				Average Years
		0	1	2	4	
Parental Transfer:						
Below Median	Data	58.3	13.0	16.8	11.9	0.94
	Model	57.1	14.3	21.0	7.5	0.86
Above Median	Data	13.3	7.5	20.2	58.9	2.84
	Model	14.8	6.5	16.1	62.6	2.89

6 Quantitative Characterization of Optimal Education Subsidies

Now I use the quantitative model to numerically characterize optimal policies. The theoretical analysis in Section 3 demonstrates that the inequality in parental transfers conditional on parental resources tends to make it efficient to more heavily subsidize low schooling levels and weaken the dependence of subsidy amounts on parental resources. However, the property of optimal policy also depends on the degree of heterogeneity in returns to schooling as well as the response of college attendance with respect to a policy change. The quantitative model in Section 4 has a rich specification that captures these trade-offs.

Table 7: Comparing Budget-Equivalent Subsidy Policies (2004 U.S. Dollars)

	Subsidy for Years (g_j)			Average Subsidy
	1	2	4	
Current Subsidy for Lowest Income Quartile	6,042	12,466	26,531	
Optimal Subsidies:				
Heterogeneous Parental Altruism	12,665	16,018	19,157	13,570
Homogeneous Parental Altruism	0	5	19,545	13,570
Highest Income Quartile	14,026	17,285	13,159	13,570

Table 8: Subsidies (2004 U.S. Dollars)

	Subsidy for Years (g_j)			Average Subsidy
	1	2	4	
A. Current Subsidy				
Parental Income:				
Quartile 1	6,042	12,466	26,531	8,007
Quartile 2	4,809	9,293	18,985	7,999
Quartile 3	3,450	7,176	14,340	7,147
Quartile 4	3,234	6,582	12,998	8,388
All				7,885
B. Optimal Subsidy without Redistribution across Income				
Parental Income:				
Quartile 1	11,616	13,876	15,574	8,007
Quartile 2	10,156	11,390	13,017	7,999
Quartile 3	8,022	10,509	9,828	7,147
Quartile 4	7,083	8,684	11,524	8,388
All				7,885
C. Optimal Subsidy				
Parental Income:				
Quartile 1	13,961	17,466	21,790	11,636
Quartile 2	10,297	11,922	14,707	8,684
Quartile 3	9,984	11,344	11,019	8,304
Quartile 4	5,400	5,365	2,434	2,910
All				7,885

Table 10: Changes in Education Attainment, Relative to Current Subsidy

	% with Completed Years				Average Years
	0	1	2	4	
A. Optimal Subsidy without Redistribution					
Parental Income:					
Quartile 1	-9.5	14.0	4.4	-8.8	-0.13
Quartile 2	-9.9	12.4	5.1	-7.6	-0.08
Quartile 3	-8.3	8.2	6.8	-6.7	-0.05
Quartile 4	-3.4	4.4	2.5	-3.4	-0.04
All	-7.8	9.8	4.7	-6.6	-0.07
B. Optimal Subsidy					
Parental Income:					
Quartile 1	-17.7	12.7	12.0	-7.0	0.09
Quartile 2	-10.8	11.9	5.4	-6.6	-0.04
Quartile 3	-10.6	10.4	6.7	-6.5	-0.02
Quartile 4	0.8	5.9	3.4	-10.1	-0.28
All	-9.5	10.2	6.9	-7.5	-0.06

Table 9: Changes Relative to Current Subsidy (2004 U.S. Dollars)

	Distortions			Parental Transfer
	Consumption (τ_c)	Schooling (τ_s)	Total (τ)	
A. Optimal Subsidy without Redistribution across Income				
Parental Income:				
Quartile 1	-2,539	1,308	-1,231	-1,471
Quartile 2	-165	-577	-742	-1,323
Quartile 3	683	-1,265	-582	-854
Quartile 4	175	-418	-244	-503
All	-462	-237	-700	-1,038
B. Optimal Subsidy				
Parental Income:				
Quartile 1	-2,356	397	-1,958	-1,944
Quartile 2	0	-842	-842	-1,409
Quartile 3	560	-1,317	-758	-1,430
Quartile 4	-330	686	356	1,246
All	-532	-269	-801	-885

7 Conclusion

Appendix

A Additional Empirical Results from the NPSAS:12 Data

Table A1: EFC and Actual Parental Contribution by Cost

	% with Amount Parents Helped Pay for Expenses						Parental Income (\$)	Grants (\$)
	\$1 to \$2,000	\$2,001 to \$5,000	\$5,001 to \$10,000	\$10,001 to \$20,000	\$20,001 or More			
	\$0							
A. Annual Cost \$20,000 or Less								
EFC:								
\$0	42.7	35.3	8.3	6.4	4.2	3.1	17,254	8,488
\$1 to \$2,000	27.6	34.8	12.9	14.4	8.8	1.4	32,470	7,821
\$2,001 to \$5,000	33.2	35.6	13.1	11.1	5.1	1.9	50,788	5,464
\$5,001 to \$10,000	20.0	33.5	21.0	14.0	8.4	3.1	71,878	2,894
\$10,001 to \$20,000	15.6	27.7	20.9	15.7	13.6	6.4	99,864	1,794
\$20,001 or More	15.8	23.0	16.9	18.9	14.1	11.2	164,511	2,623
All	27.8	31.8	14.7	12.5	8.5	4.6	68,732	5,175
B. Annual Cost \$20,001 or More								
EFC:								
\$0	38.8	35.6	8.7	8.9	5.4	2.6	18,124	12,023
\$1 to \$2,000	29.1	39.5	10.0	8.5	9.8	3.1	34,150	10,689
\$2,001 to \$5,000	29.3	34.8	12.8	11.9	7.0	4.2	51,712	7,351
\$5,001 to \$10,000	21.2	31.6	14.3	14.7	13.2	5.0	72,645	3,906
\$10,001 to \$20,000	14.5	25.7	15.4	14.2	20.5	9.6	104,452	2,974
\$20,001 or More	10.6	14.6	9.6	17.9	25.5	21.7	164,496	2,050
All	22.8	28.3	11.6	13.2	14.9	9.1	83,223	6,059

Table A2: EFC and Actual Parental Contribution by Net Cost

	% with Amount Parents Helped Pay for Expenses						Parental Income (\$)	Grants (\$)
	\$1	\$2,001	\$5,001	\$10,001	\$20,001			
	\$0 to \$2,000	to \$5,000	to \$10,000	to \$20,000	or More			
C. Annual Cost Less Grant \$15,000 or Less								
EFC:								
\$0	41.6	35.1	8.7	7.3	4.4	2.8	17,284	11,310
\$1 to \$2,000	27.8	39.8	10.4	10.6	9.6	1.8	32,996	11,048
\$2,001 to \$5,000	32.9	35.2	12.9	10.9	5.2	2.9	50,909	8,985
\$5,001 to \$10,000	17.7	42.4	20.8	10.2	7.4	1.4	71,246	7,832
\$10,001 to \$20,000	14.7	27.4	25.3	20.0	9.6	2.9	95,573	6,867
\$20,001 or More	16.3	23.4	19.2	18.1	15.6	7.3	156,049	7,693
All	32.3	35.2	12.5	10.2	6.9	2.9	45,099	10,076
D. Annual Cost Less Grant \$15,001 or More								
EFC:								
\$0	33.1	36.9	8.6	11.3	7.4	2.8	19,739	6,620
\$1 to \$2,000	30.5	33.2	12.4	10.4	9.2	4.3	34,932	6,424
\$2,001 to \$5,000	28.9	35.1	12.9	12.2	7.2	3.7	51,767	4,518
\$5,001 to \$10,000	21.6	28.8	15.0	16.0	13.2	5.4	72,668	2,075
\$10,001 to \$20,000	14.9	25.9	15.7	13.8	20.0	9.7	104,302	1,856
\$20,001 or More	11.2	15.3	10.1	18.2	24.0	21.2	165,981	1,192
All	18.7	25.3	12.8	15.1	17.1	11.1	102,064	2,584

B Proofs and Analytical Details

B.1 Proof of Lemma 1

Due to [Milgrom and Shannon \(1994\)](#), it is sufficient to show that $U_j(b)$ obeys single crossing differences. That is, for all $j' > j$ and $b' > b$, $U_{j'}(b) \geq (>) U_j(b)$ implies $U_{j'}(b') \geq (>) U_j(b')$. For a pair of options (j, j') , there are 4 possibilities, depending on whether those with b are constrained or not for each option.

(i) Unconstrained for both (j, j') : In this case, it is easy to show that $U_{j'}(b) \geq (>) U_j(b)$ if and only if $y_{j'} - k_{j'} \geq (>) y_j - k_j$. Therefore, the choice is not affected by b . Since those with $b' > b$ are also unconstrained for both options, $U_{j'}(b') \geq (>) U_j(b')$ follows.

(ii) Constrained only for j' : In this case, (1) $b \geq k_j + y_j$ or $j = 0$, and (2) $b < k_{j'} + y_{j'}$ and $j' > j$. First, note that $U_{j'}(b) \geq U_j(b)$ implies $y_{j'} - k_{j'} > y_j - k_j$. To see this, let $U_j^*(b)$ be the indirect utility function without (BC). Then $U_{j'}^*(b) > U_{j'}(b) \geq U_j(b) = U_j^*(b)$ holds. Therefore, $y_{j'} - k_{j'} > y_j - k_j$ holds because it is equivalent to $U_{j'}^*(b) > U_j^*(b)$. This also implies that $U_{j'}(b') > U_j(b')$ holds for those with $b' \geq y_{j'} + k_{j'}$.

For $b < y_{j'} + k_{j'}$, the condition $U_{j'}(b) \geq (>)U_j(b)$ is

$$4(b - k_{j'})y_{j'} - (b - k_j + y_j)^2 \geq (>)0.$$

The derivative of the left hand side respect to b is positive if and only if $4y_{j'} - 2(b - k_j + y_j) \geq 0$, or $b \leq 2y_{j'} - (y_j - k_j)$. Since $y_{j'} - k_{j'} > y_j - k_j$ implies $k_{j'} + y_{j'} < 2y_{j'} - (y_j - k_j)$, $U_{j'}(b) - U_j(b)$ is strictly increasing for $b < y_{j'} + k_{j'}$. Therefore, $U_{j'}(b) \geq U_j(b)$ implies $U_{j'}(b') > U_j(b')$ for $b' \in (b, y_{j'} + k_{j'})$.

(iii) Constrained only for j : In this case, $0 < j < j'$ and $k_{j'} + y_{j'} \leq b < k_j + y_j$, and the condition $U_{j'}(b) \geq (>)U_j(b)$ is

$$(b - k_{j'} + y_{j'})^2 - 4(b - k_j)y_j \geq (>)0.$$

The derivative of the left hand side respect to b is positive if and only if $2(b - k_{j'} + y_{j'}) - 4y_j \geq 0$, or $b \geq 2y_j - (y_{j'} - k_{j'})$. Since $y_{j'} \geq y_j$ implies $2y_j - (y_{j'} - k_{j'}) \leq k_{j'} + y_{j'}$, $U_{j'}(b) - U_j(b)$ is increasing for $b \in [y_{j'} + k_{j'}, y_j + k_j]$. Therefore, $U_{j'}(b) \geq (>)U_j(b)$ implies $U_{j'}(b') \geq (>)U_j(b')$ for $b' \in (b, k_j + y_j)$. Moreover, since $U_{j'}(b') \geq (>)U_j(b')$ holds for $b' = k_j + y_j$, $U_{j'}(b'') \geq (>)U_j(b'')$ also holds for $b'' > b' = k_j + y_j$ by the same argument as case (i).

(iv) Constrained for both (j, j') : In this case, $0 < j < j'$ and $b < \min\{k_j + y_j, k_{j'} + y_{j'}\}$, and the condition $U_{j'}(b) \geq (>)U_j(b)$ is

$$(b - k_{j'})y_{j'} - (b - k_j)y_j \geq (>)0,$$

which is equivalent to $b \geq (>)(y_{j'}k_{j'} - y_jk_j)/(y_{j'} - y_j)$. Therefore, $U_{j'}(b) \geq (>)U_j(b)$ implies $U_{j'}(b') \geq (>)U_j(b')$ for $b' \in (b, \min\{k_j + y_j, k_{j'} + y_{j'}\})$. We can also show that $U_{j'}(b') \geq (>)U_j(b')$ holds for $b' \geq \min\{k_j + y_j, k_{j'} + y_{j'}\}$ by applying the arguments in cases (i), (ii), and (iii).

B.2 Proof of Lemma 2

If $j = 0$ or $b \geq k_j + y_j$, then $V_j(b) = V_j^*(b) = b - k_j + y_j$ and $V_j'(b) = 1$. If $j > 1$ and $b < k_j + y_j$,

$$V_j'(b) = \sqrt{\frac{y_j}{b - k_j}} > 1.$$

From this, it is easy to see that $V_j'(b)$ is decreasing in b and $V_j'(b)$ is increasing in j if y_j and k_j are increasing in j .

B.3 Proof of Lemma 3

First, note that $V(b)$ is continuous by the maximum theorem, and therefore $\tau(b)$ is continuous as well. Next, by the envelope theorem, $V'(b) = V_j'(b)$ for some $\hat{j} \in \operatorname{argmax}_{j \in \mathcal{J}} \{V_j(b)\}$ where $V(b)$ is differentiable. This implies $\tau(b)$ is decreasing in b where $V(b)$ is differentiable because $\tau'(b) = 1 - V'(b)$. When $V(b)$ is not differentiable at b , we can apply the same argument to show that the left- and right-derivative of $\tau(b)$ is negative. Since $\tau(b)$ is continuous, this completes the proof.

B.4 Single Crossing Differences

Corollary 4. *If $y_{j'} \geq y_j$ for all $j \in \mathcal{J}$ and $j' \in \mathcal{J}$ such that $j' > j$, then $V_{j'}(b + g_{j'}) \geq (>)V_j(b + g_j)$ implies $V_{j'}(b' + g_{j'}) \geq (>)V_j(b' + g_j)$ where $b' > b$.*

Proof. Since Lemma 1 does not require assumptions on k_j and also relies on ordinal properties of $U_j(b)$, we can apply Lemma 1 by letting $\tilde{k}_j \equiv k_j - g_j$ and $\tilde{V}_j(b) \equiv V_j(b + g_j)$. $\square$

By the contrapositive of this lemma, for $j' > j$ and $b' > b$, $V_{j'}(b' + g_{j'}) < (\leq)V_j(b' + g_j)$ also implies $V_{j'}(b + g_{j'}) < (\leq)V_j(b + g_j)$.

B.5 Proof of Lemma 4

I first show that the following inequality holds by Assumption 1:

$$k_2 < \frac{y_2 k_2 - y_1 k_1}{y_2 - y_1} < k_1 + y_1.$$

$k_2 < (y_2 k_2 - y_1 k_1)/(y_2 - y_1)$ holds due to $k_1 < k_2$ and $(y_2 k_2 - y_1 k_1)/(y_2 - y_1) < k_1 + y_1$ is equivalent to $k_2 - k_1 < y_1/y_2(y_2 - y_1)$.

Because $V_1(k_2) > V_2(k_2) = -\infty$ and $V_1(k_2 + y_2) = V_1^*(k_2 + y_2) < V_2^*(k_2 + y_2) = V_2(k_2 + y_2)$, there exists $\underline{b}_2 \in (k_2, k_2 + y_2)$ such that $V_1(\underline{b}_2) = V_2(\underline{b}_2)$ by the intermediate value theorem. For those with $b \in (k_2, k_1 + y_1)$, (BC) binds for both choices, so $V_1(b) \geq V_2(b)$ holds if and only if $(b - k_1)y_1 \geq (b - k_2)y_2$, which is equivalent to $b \leq (y_2 k_2 - y_1 k_1)/(y_2 - y_1)$. Therefore, $V_1(b) > V_2(b)$ for $b \in (k_2, \underline{b}_2)$, $V_1(b) = V_2(b)$ for $b = \underline{b}_2$ and $V_1(b) < V_2(b)$ for $b \in (\underline{b}_2, y_1 + k_1]$. Moreover, by the single crossing differences, $V_1(b) > V_2(b)$ for $b \in (k_1, k_2)$ and $V_1(b) < V_2(b)$ for $b > y_1 + k_1$ also hold. Therefore $\underline{b}_2$ uniquely solves $V_1(\underline{b}_1) = V_2(\underline{b}_2)$.

B.6 Proof of Proposition 1

To simplify notation, let $p_{j|i} \equiv p_j(b_i)$ throughout the proof of this proposition.

Lemma 6. $V_1(b_L + \hat{g}_1) \neq V_2(b_L + \hat{g}_2)$ or $V_1(b_H + \hat{g}_1) \neq V_2(b_H + \hat{g}_2)$.

Proof. Suppose that $V_1(b_L + \hat{g}_1) = V_2(b_L + \hat{g}_2)$ and $V_1(b_H + \hat{g}_1) = V_2(b_H + \hat{g}_2)$. Since $V_1(b_H + \hat{g}_1) = V_2(b_H + \hat{g}_2)$ implies $\hat{g}_1 > \hat{g}_2$, $\hat{p}_{1|L} > 0$ or $\hat{p}_{1|H} > 0$ cannot be optimum, as we can reduce aggregate spending by setting $p_{1|L} = p_{1|H} = 0$. Therefore, $\hat{p}_{1|L} = \hat{p}_{1|H} = 0$ must hold. Then, from the budget constraint, $\hat{g}_2 = E$. Also $V_1(b_H + \hat{g}_1) = V_2(b_H + \hat{g}_2)$ implies $\hat{g}_1 = E + y_2 - y_1 - (k_2 - k_1)$. Since $\hat{g}_2 = E < y_1 + k_1 - b_L < y_2 + k_2 - b_L$, the type L is constrained at $j = 2$. If the type L is unconstrained at $j = 1$, then $V_1(b_L + \hat{g}_1) = V_1^*(b_L + \hat{g}_1) = V_2(b_L + \hat{g}_2) < V_2^*(b_L + \hat{g}_2)$, which contradicts $V_1(b_H + \hat{g}_1) = V_2(b_H + \hat{g}_2)$. If the type L is constrained at $j = 1$, then $V_1(b_L + \hat{g}_1) = V_2(b_L + \hat{g}_2)$ implies $(b_L - k_1 + \hat{g}_1)y_1 = (b_L - k_2 + \hat{g}_2)y_2$, which is equivalent to $E = y_1 + k_2 - b_L$ due to $\hat{g}_1 = E + y_2 - y_1 - (k_2 - k_1)$ and $\hat{g}_2 = E$. This contradicts $E < y_1 + k_1 - b_L$ because $y_1 + k_1 < y_2 + k_2$. $\square$

Lemma 7. $\hat{p}_{1|H} = 0$.

Proof. Suppose that $\hat{p}_{1|H} > 0$. Then $V_1(b_H + \hat{g}_1) \geq V_2(b_H + \hat{g}_2)$ holds, which implies $\hat{g}_1 - \hat{g}_2 \geq y_2 - y_1 - (k_2 - k_1) > 0$ and $V_1(b_L + \hat{g}_1) \geq V_2(b_L + \hat{g}_2)$. Since $V_1(b_L + \hat{g}_1) = V_2(b_L + \hat{g}_2)$ and $V_1(b_H + \hat{g}_1) = V_2(b_H + \hat{g}_2)$ cannot hold at the same time due to Lemma 6, consider the following 2 cases:

(i) $V_1(b_H + \hat{g}_1) > V_2(b_H + \hat{g}_2)$: In this case, $V_1(b_L + \hat{g}_1) > V_2(b_L + \hat{g}_2)$ also holds. Therefore, $\hat{p}_{1|L} = \hat{p}_{1|H} = 1$ and $\hat{g}_1 = E$. However, this allocation is not optimal, because there exists an alternative allocation that leads to higher social welfare with the same level of spending. Consider $\tilde{g}_1 = \tilde{g}_2 = E$ and $\tilde{p}_{1|L} = \tilde{p}_{2|H} = 1$. Then $V_1(b_L + \hat{g}_1) = V_1(b_L + \tilde{g}_1) > V_2(b_L + \hat{g}_2)$ and $V_1(b_H + \hat{g}_1) = V_1(b_H + \tilde{g}_1) < V_2(b_H + \hat{g}_2)$.

(ii) $V_1(b_H + \hat{g}_1) = V_2(b_H + \hat{g}_2)$ and $V_1(b_L + \hat{g}_1) > V_2(b_L + \hat{g}_2)$: In this case, $\hat{p}_{1|L} = 1$. This allocation is not optimal because we can achieve the same level of social welfare with lower spending by letting the type H to switch from $j = 1$ to $j = 2$. Consider $\tilde{g} = \hat{g}$, $\tilde{p}_{1|L} = \tilde{p}_{2|H} = 1$. Since $V_1(b_H + \hat{g}_1) = V_2(b_H + \hat{g}_2)$, this allocation gives the same utility for both types. However, aggregate spending is reduced because $\hat{g}_1 > \hat{g}_2$. $\square$

Lemma 8. $\hat{p}_{2|L} = 0$.

Proof. Suppose that $\hat{p}_{2|L} > 0$. Then $V_1(b_L + \hat{g}_1) \leq V_2(b_L + \hat{g}_2)$ holds. From Lemma 6, we also have $\hat{p}_{1|H} = 0$, which implies $V_1(b_H + \hat{g}_1) \leq V_2(b_H + \hat{g}_2)$. Note that $V_1(b_L + \hat{g}_1) < V_2(b_L + \hat{g}_2)$ and $V_1(b_H + \hat{g}_1) = V_2(b_H + \hat{g}_2)$ cannot simultaneously hold due to the single crossing differences. Since $V_1(b_L + \hat{g}_1) = V_2(b_L + \hat{g}_2)$ and $V_1(b_H + \hat{g}_1) = V_2(b_H + \hat{g}_2)$ cannot hold at the same time due to Lemma 6, consider the following 2 cases:

(i) $V_1(b_L + \hat{g}_1) < V_2(b_L + \hat{g}_2)$ and $V_1(b_H + \hat{g}_1) < V_2(b_H + \hat{g}_2)$: In this case, $\hat{p}_{2|L} = \hat{p}_{2|H} = 1$ and, therefore, $\hat{g}_2 = E$ holds. However, in this case, there exists an alternative allocation that leads to higher social welfare with the same level of spending. Consider $\tilde{g}_1 = \tilde{g}_2 = E$ and $\tilde{p}_{1|L} = \tilde{p}_{2|H} = 1$. Note that $V_1(b_L + \tilde{g}_1) > V_2(b_L + \tilde{g}_2) = V_2(b_L + \hat{g}_2)$ is satisfied because $\tilde{g}_1 = \tilde{g}_2 = E < b_2 - b_L$. Moreover, $V_1(b_H + \tilde{g}_1) < V_2(b_H + \tilde{g}_2) = V_2(b_H + \hat{g}_2)$ is also satisfied. Since the type L is better off while the type H is indifferent, aggregate welfare is improved.

(ii) $V_1(b_L + \hat{g}_1) = V_2(b_L + \hat{g}_2)$ and $V_1(b_H + \hat{g}_1) < V_2(b_H + \hat{g}_2)$: For $\hat{p}_{2|L} > 0$ to be optimal, $\hat{g}_1 \geq \hat{g}_2$ need to be satisfied. If $\hat{g}_1 > \hat{g}_2$, then $\hat{p}_{2|L} = 1$, which implies $\hat{g}_2 = E$. Then $V_1(b_L + \hat{g}_1) = V_2(b_L + \hat{g}_2) < V_1(b_L + \hat{g}_2)$, which implies $\hat{g}_1 < \hat{g}_2$, leading to a contradiction. If $\hat{g}_1 = \hat{g}_2$, then $\hat{g}_1 = \hat{g}_2 = E$ holds. Since $V_1(b_L + E) > V_2(b_L + E)$ holds, the assumption $V_1(b_L + \hat{g}_1) = V_2(b_L + \hat{g}_2)$ is violated. $\square$

Lemma 9. $\hat{g}_1 > \hat{g}_2$.

Proof. With $\hat{p}_{1|L} = \hat{p}_{2|H} = 1$, $\tau(b_H, \hat{g}) = 0$ holds, so the optimal allocation minimizes the distortion of the type L :

$$\begin{aligned} & (\hat{g}_1, \hat{g}_2) \in \underset{g_1, g_2}{\operatorname{argmin}} \left\{ V_1^*(b_L + g_1) - V_1(b_L + g_1) \right\} \\ & \text{subject to } f_L g_1 + f_H g_2 \leq E \tag{23} \\ & V_1(b_L + g_1) \geq V_2(b_L + g_2) \tag{IC_L} \\ & V_1(b_H + g_1) \leq V_2(b_H + g_2) \tag{IC_H} \\ & g_1 \geq 0, \quad g_2 \geq 0 \end{aligned}$$

There are 2 cases to consider:

(i) $E < f_L(y_1 + k_1 - b_L)$: In this case, g_1 cannot exceed $y_1 + k_1 - b_L$, implying that, for all feasible g_1 and g_2 values, the type L is constrained and $\partial V_1(b_L + g_1)/\partial g_1 > 1$. Therefore, it is optimal to increase g_1 as much as possible, leading to $\hat{g}_1 = E/f_L$ and $\hat{g}_2 = 0$. (IC_L) is satisfied because $V_1(b_L + \hat{g}_1) > V_1(b_L + E) > V_2(b_L + E) > V_2(b_L + \hat{g}_2)$, where the second inequality is due to Assumption 3. (IC_H) is satisfied as long as $E \leq f_L[y_2 - y_1 - (k_2 - k_1)]$. On the other hand, if $E > f_L[y_2 - y_1 - (k_2 - k_1)]$, (IC_H) is binding and $(\hat{g}_1, \hat{g}_2)$ is determined to satisfy both (IC_H) and the budget constraint (23). Therefore, $\hat{g}_1 = E + f_H[y_2 - y_1 - (k_2 - k_1)]$

and $\hat{g}_2 = E - f_L[y_2 - y_1 - (k_2 - k_1)]$. **(IC_L)** is also satisfied because $V_1(b_L + \hat{g}_1) > V_1(b_L + E) > V_2(b_L + E) > V_2(b_L + \hat{g}_2)$, where Assumption 3 and $\hat{g}_1 > E > \hat{g}_2$ are used for inequalities.

(ii) $E \geq f_L(y_1 + k_1 - b_L)$: In this case, it is possible to raise g_1 beyond $y_1 + k_1 - b_L$, which, however, does not reduce aggregate distortion, as $\partial V_1(b_L + g_1)/\partial g_1 = 1$ for $g_1 \geq y_1 + k_1 - b_L$. Therefore, any $(\hat{g}_1, \hat{g}_2)$ satisfying $\hat{g}_1 \geq y_1 + k_1 - b_L$ and (23) is optimal, as long as both **(IC_L)** and **(IC_H)** are satisfied. $\hat{g}_1 \geq y_1 + k_1 - b_L$ implies $\hat{g}_1 > \hat{g}_2$ because $\hat{g}_1 - \hat{g}_2 \geq (\hat{g}_1 - E)/f_H \geq (y_1 + k_1 - b_L - E)/f_H > 0$, where the last inequality follows from Assumptions 2 and 3.

Next, I check whether this allocation satisfies **(IC_L)** and **(IC_H)**. **(IC_L)** is satisfied because $V_1(b_L + \hat{g}_1) > V_1(b_L + E) > V_2(b_L + E) > V_2(b_L + \hat{g}_2)$, which follows from $\hat{g}_1 > E > \hat{g}_2$. If $y_1 + k_1 - b_L \leq E + f_H[y_2 - y_1 - (k_2 - k_1)]$, then $g_1 = y_1 + k_1 - E$ and $g_2 = [E - f_L(y_1 + k_1 - b_L)]/f_H$ satisfies **(IC_H)**, thus there exists $\hat{g}$ satisfying $\hat{g}_1 \geq y_1 + k_1 - b_L$, **(IC_L)**, and **(IC_H)**. On the other hand, if $y_1 + k_1 - b_L > E + f_H[y_2 - y_1 - (k_2 - k_1)]$, **(IC_H)** is binding, so $\hat{g}_1 = E + f_H[y_2 - y_1 - (k_2 - k_1)]$ and $\hat{g}_2 = E - f_L[y_2 - y_1 - (k_2 - k_1)]$. $\square$

B.7 Proof of Proposition 2

Part (i): I first calculate several derivatives that are useful for writing down the first order conditions.

By differentiating (9) with respect to g_j , we get

$$\begin{aligned} \frac{\partial p_j(b, \mathbf{g})}{\partial g_j} &= \frac{\exp(V_j(b + g_j))V_j'(b + g_j) \left[\sum_{j' \in \mathcal{J}} \exp(V_{j'}(b + g_{j'})) \right] - \exp(V_j(b + g_j))^2 V_j'(b + g_j)}{\left[\sum_{j' \in \mathcal{J}} \exp(V_{j'}(b + g_{j'})) \right]^2} \\ &= p_j(b, \mathbf{g})(1 - p_j(b, \mathbf{g}))V_j'(b + g_j). \end{aligned}$$

The derivative of (9) with respect to $g_{j'}$ for $j' \neq j$ is

$$\begin{aligned} \frac{\partial p_j(b, \mathbf{g})}{\partial g_{j'}} &= \frac{-\exp(V_j(b + g_j)) \exp(V_{j'}(b + g_{j'}))V_{j'}'(b + g_{j'})}{\left[\sum_{j'' \in \mathcal{J}} \exp(V_{j''}(b + g_{j''})) \right]^2} \\ &= -p_j(b, \mathbf{g})p_{j'}(b, \mathbf{g})V_{j'}'(b + g_{j'}). \end{aligned}$$

The derivative of the aggregate spending with respect to g_j is

$$\begin{aligned} \frac{\partial}{\partial g_j} \left(\sum_{j' \in \mathcal{J}} p_{j'}(b, \mathbf{g})g_{j'} \right) &= p_j(b, \mathbf{g}) + \sum_{j' \in \mathcal{J}} \frac{\partial p_{j'}(b, \mathbf{g})}{\partial g_j} g_{j'} \\ &= p_j(b, \mathbf{g}) + p_j(b, \mathbf{g})(1 - p_j(b, \mathbf{g}))V_j'(b + g_j)g_j - \sum_{j' \in \mathcal{J} \setminus \{j\}} p_j(b, \mathbf{g})p_{j'}(b, \mathbf{g})V_{j'}'(b + g_{j'})g_{j'} \\ &= p_j(b, \mathbf{g}) + p_j(b, \mathbf{g})V_j'(b + g_j)g_j - p_j(b, \mathbf{g})V_j'(b + g_j) \sum_{j' \in \mathcal{J}} p_{j'}(b, \mathbf{g})g_{j'} \\ &= p_j(b, \mathbf{g}) \left\{ 1 + \left(g_j - \sum_{j' \in \mathcal{J}} p_{j'}(b, \mathbf{g})g_{j'} \right) V_j'(b + g_j) \right\}. \end{aligned}$$

Finally, the derivative of $V(b, \mathbf{g})$ is

$$\frac{\partial V(b, \mathbf{g})}{\partial g_j} = \frac{\exp(V_j(b + g_j))}{\sum_{j' \in \mathcal{J}} \exp(V_{j'}(b + g_{j'}))} V_j'(b + g_j) = p_j(b, \mathbf{g})V_j'(b + g_j).$$

Then first order condition with respect to g_j is

$$p_j(b, \hat{\mathbf{g}}) V_j'(b + \hat{g}_j) \leq (1 + \lambda) p_j(b, \hat{\mathbf{g}}) \left[1 + \left(\hat{g}_j - \sum_{j' \in \mathcal{J}} p_{j'}(b, \hat{\mathbf{g}}) \hat{g}_{j'} \right) V_j'(b + \hat{g}_j) \right],$$

where the inequality is strict if and only if $g_j \geq 0$ is binding. By rearranging the above condition, we get (12) when $\hat{g}_j > 0$. Since $g_j + 1/V_j'(b + g_j)$ is strictly increasing in g_j , g_j that solves (12) is less than or equal to zero if and only if (??) holds.

Part (ii): Consider $j \in \mathcal{J}_+$ and $j' \in \mathcal{J}_+$ such that $y_{j'} \geq y_j$ and $k_{j'} \geq k_j$. Excluding obvious cases $\hat{g}_{j'} > \hat{g}_j = 0$ and $\hat{g}_{j'} = \hat{g}_j = 0$, we have 2 cases to consider. If $\hat{g}_j > 0$ and $\hat{g}_{j'} > 0$, then, from (12),

$$\hat{g}_j + \sqrt{\frac{b - k_j + \hat{g}_j}{y_j}} = \hat{g}_{j'} + \sqrt{\frac{b - k_{j'} + \hat{g}_{j'}}{y_{j'}}} \leq \hat{g}_{j'} + \sqrt{\frac{b - k_j + \hat{g}_{j'}}{y_j}},$$

which implies $g_{j'} \geq g_j$ because $g_j + \sqrt{(b - k_j + g_j)/y_j}$ is increasing in g_j . If $\hat{g}_j > \hat{g}_{j'} = 0$, then, from (??), the following holds:

$$\frac{1}{V_j'(b)} < \sum_{j' \in \mathcal{J}} p_{j'}(b, \hat{\mathbf{g}}) \hat{g}_{j'} + \frac{1}{1 + \lambda} \leq \frac{1}{V_{j'}'(b)}.$$

However, $1/V_j'(b) < 1/V_{j'}'(b)$ contradicts the assumption $y_{j'} \geq y_j$ and $k_j \geq k_{j'}$. Therefore, $\hat{g}_j > \hat{g}_{j'} = 0$ cannot hold.

Part (iii): Suppose that $b \geq y_j + k_j - \hat{g}_j$ for all $j \in \mathcal{J}_+$. Then $V'(b + \hat{g}_j) = 1$ for all $j \in \mathcal{J}_+$. If $\sum_{j' \in \mathcal{J}} p_{j'}(b, \hat{\mathbf{g}}) \hat{g}_{j'} \leq \lambda/(1 + \lambda)$, then $\hat{g}_j = 0$ for all $j \in \mathcal{J}$. Because aggregate spending is zero, which is less than the budget, the budget constraint (11) is not binding and $\lambda = 0$. If $\sum_{j' \in \mathcal{J}} p_{j'}(b, \hat{\mathbf{g}}) \hat{g}_{j'} > \lambda/(1 + \lambda)$, then $\hat{g}_j = \sum_{j' \in \mathcal{J}} p_{j'}(b, \hat{\mathbf{g}}) \hat{g}_{j'} - \lambda/(1 + \lambda) > 0$ holds for all $j \in \mathcal{J}_+$. By multiplying $\hat{g}_j$ with p_j and summing over $j \in \mathcal{J}_+$, we have

$$\hat{p}_0(b, \hat{\mathbf{g}}) \sum_{j \in \mathcal{J}} p_j(b, \hat{\mathbf{g}}) \hat{g}_j = -\frac{\lambda}{1 + \lambda} [1 - \hat{p}_0(b, \hat{\mathbf{g}})].$$

Since $\hat{g}_j > 0$ for all $j \in \mathcal{J}_+$ and $\lambda \geq 0$, this can hold only if $\hat{\mathbf{g}} = \mathbf{0}$ and $\lambda = 0$.

B.8 Proof of Corollary 2

Let $\bar{\mathbf{g}}$ be the solution to Problem 2 without the planner's budget constraint (11), and let $\bar{E}$ be the implied aggregate spending:

$$\bar{E} \equiv \sum_{j \in \mathcal{J}} p_j(b, \bar{\mathbf{g}}) \bar{g}_j.$$

Part (i): If $\bar{E} < \infty$ exists, then the inequality constraint (11) binds if and only if $E < \bar{E}$. Let $\hat{\mathbf{g}}$ be the solution to Problem 2 where we replace the inequality constraint (11) in Problem 2 with an equality constraint. Then the

first order condition when $\hat{g}_j > 0$ can be written as

$$\hat{g}_j + \frac{1}{V_j'(b + \hat{g}_j)} = E + \frac{1}{1 + \lambda}, \quad (24)$$

and $\hat{g}_j = 0$ if and only if

$$\frac{1}{V_j'(b)} \geq E + \frac{1}{1 + \lambda}. \quad (25)$$

The next lemma shows that everyone is unconstrained when E is large enough.

Lemma 10. *If $E \geq \max_{j \in \mathcal{J}_+} \{y_j + k_j - b\} > 0$, then $b \geq \max_{j \in \mathcal{J}_+} \{y_j + k_j - \hat{g}_j\}$.*

Proof. I first show that $\hat{g}_j \geq \max_{j' \in \mathcal{J}_+} \{y_{j'} + k_{j'} - b\}$ for some $j \in \mathcal{J}_+$. Suppose not, i.e., $\hat{g}_j < \max_{j' \in \mathcal{J}_+} \{y_{j'} + k_{j'} - b\}$ for all $j \in \mathcal{J}_+$. Then

$$\sum_{j \in \mathcal{J}} p_j(b, \hat{\mathbf{g}}) \hat{g}_j < \left(\max_{j' \in \mathcal{J}_+} \{y_{j'} + k_{j'} - b\} \right) \sum_{j \in \mathcal{J}} p_j(b, \hat{\mathbf{g}}) = \max_{j' \in \mathcal{J}_+} \{y_{j'} + k_{j'} - b\} \leq E,$$

which violates the equality budget constraint. Thus, there exists $\hat{j} \in \mathcal{J}_+$ such that $\hat{g}_{\hat{j}} \geq \max_{j' \in \mathcal{J}_+} \{y_{j'} + k_{j'} - b\} - b$, which implies

$$\hat{g}_{\hat{j}} = E - \frac{\lambda}{1 + \lambda} \geq \max_{j' \in \mathcal{J}_+} \{y_{j'} + k_{j'} - b\}.$$

Then, for all $j \in \mathcal{J}_+$,

$$\left(g_j + \frac{1}{V_j'(b + g_j)} \right) \Big|_{g_j = y_j + k_j - b} = y_j + k_j - b + 1 \leq \max_{j' \in \mathcal{J}_+} \{y_{j'} + k_{j'} - b\} + 1 \leq E + \frac{1}{1 + \lambda},$$

which implies that $\hat{g}_j \geq y_j + k_j - b$ holds for all $j \in \mathcal{J}_+$. $\square$

Therefore, for all $E \geq \max_{j \in \mathcal{J}_+} \{y_j + k_j - b\}$, by aggregating (24) over $j \in \mathcal{J}_+$, we get

$$p_0(b, \hat{\mathbf{g}})E = -(1 - p_0(b, \hat{\mathbf{g}})) \frac{\lambda}{1 + \lambda}.$$

This equation can hold only if $\lambda \in (-1, 0)$. The negative Lagrangian multiplier on the equality constraint implies that reducing aggregate spending decreases aggregate distortions, therefore the inequality budget constraint does not bind for all $E \geq \max_{j \in \mathcal{J}_+} \{y_j + k_j - b\}$. Thus, the aggregate spending in the absence of equality constraint, $\bar{E}$, must be finite and strictly less than $\max_{j \in \mathcal{J}_+} \{y_j + k_j - b\}$.

Part (ii): If $b \geq \max_{j \in \mathcal{J}_+} \{y_j + k_j\}$, then $b \geq \max_{j \in \mathcal{J}_+} \{y_j + k_j - \hat{g}_j\}$ also holds due to $\hat{g}_j \geq 0$. Therefore, by Lemma 2, $\hat{\mathbf{g}} = \mathbf{0}$ and the inequality budget constraint never binds for any $E \geq 0$, implying $\bar{E} = 0$.

Part (iii): Since $\bar{E} < \max_{j \in \mathcal{J}_+} \{y_j + k_j\}$ is shown in part (i), I only prove $\bar{E} > 0$. Note that $\bar{g}_j = 0$ if and only if

$$\frac{1}{V_j'(b)} \geq \bar{E} + 1. \quad (26)$$

Suppose that $\bar{E} = 0$. Since there exists some $\hat{j} \in \mathcal{J}_+$ such that $V_{\hat{j}}'(b) > 1$, $\bar{g}_{\hat{j}} > 0$ should hold because the condition (26) cannot hold. This contradicts the assumption $\bar{E} = 0$ because $p_j(b, \bar{g}) > 0$.

B.9 Proof of Corollary 3

The first order condition for $j \in \mathcal{J}_+$ is

$$\sum_{i \in \mathcal{I}} f_i p_j(b_i, \hat{\mathbf{g}}) V_j'(b_i + \hat{g}_j) \leq (1 + \lambda) \sum_{i \in \mathcal{I}} f_i p_j(b_i, \hat{\mathbf{g}}) \left[1 + \left(\hat{g}_j - \sum_{j' \in \mathcal{J}} p_{j'}(b_i, \hat{\mathbf{g}}) \hat{g}_{j'} \right) V_j'(b_i + \hat{g}_j) \right].$$

Rearranging terms gives the result.

B.10 Proof of Lemma 5

When (16) does not bind, $V_j(a, z) = V_j^*(a, z) = y_j(a) - k_j + z$. Therefore, $\partial V_j(a, z)/\partial z = \partial V_j^*(a, z)/\partial z = 1$ holds. When (16) binds,

$$\begin{aligned} \frac{\partial V_j(a, z)}{\partial z} &= \sum_{t=1}^{T_k} R^{-t} \left[u' \left(u^{-1} \left(\frac{U_j(a, z)}{\sum_{t=1}^{T_k} \beta^t} \right) \right) \sum_{t=1}^{T_k} \beta^t \right]^{-1} \frac{\partial U_j(a, z)}{\partial z} \\ &= \left[u' \left(u^{-1} \left(\frac{U_j(a, z)}{\sum_{t=1}^{T_k} \beta^t} \right) \right) \right]^{-1} u' \left(\frac{z - k_j + \bar{d}_j}{\sum_{t=1}^j R^{-t}} \right) \\ &> 1, \end{aligned}$$

where the inequality holds if and only if

$$u^{-1} \left(\frac{U_j(a, z)}{\sum_{t=1}^{T_k} \beta^t} \right) > \frac{z - k_j + \bar{d}_j}{\sum_{t=1}^j R^{-t}} \Leftrightarrow U_j(a, z) > \sum_{t=1}^{T_k} \beta^t u \left(\frac{z - k_j + \bar{d}_j}{\sum_{t=1}^j R^{-t}} \right).$$

This holds because the annual consumption during schooling is lower than the annual consumption after schooling when (16) binds.

Since $u(c)$ is increasing and $u'(c)$ is decreasing in c , $\partial V_j(a, z)/\partial z$ is increasing in a if and only if $U_j(a, z)$ is increasing in a :

$$\frac{\partial U_j(a, z)}{\partial a} = u' \left(\frac{y_j(a) - \bar{d}_j}{\sum_{t=j+1}^{T_k} \beta^t} \right) \frac{dy_j(a)}{da} \geq 0.$$

Finally, the amount of borrowing has the same sign as $dy_j(a)/da$, because it can be written as:

$$\sum_{t=1}^j R^{-t} c_t + k_j - z = \sum_{t=1}^j R^{-t} \frac{y_j(a) - k_j + z}{\sum_{t=1}^{T_k} R^{-t}} + k_j - z. \quad (27)$$

This completes the proof that $\partial V_j(a, z)/\partial a$ has the same sign as $dy_j(a)/da$.

B.11 Proof of Proposition 3

First, note that (17) and (18) are given by

$$p_j(a, \mathbf{b}, \mathbf{g}) = \frac{\exp([V_j(a, b_j + g_j) + \mu_j]/\sigma)}{\sum_{j' \in \mathcal{J}} \exp([V_{j'}(a, b_j + g_j) + \mu_j]/\sigma)},$$

$$V(a, \mathbf{b}, \mathbf{g}) = \sigma \ln \left(\sum_{j \in \mathcal{J}} \exp([V_j(a, b_j + g_j) + \mu_j]/\sigma) \right) + \sigma \psi.$$

By differentiating $p_j(a, \mathbf{b}, \mathbf{g})$ with respect to b_j , we get

$$\frac{\partial p_j(a, \mathbf{b}, \mathbf{g})}{\partial b_j} = p_j(a, \mathbf{b}, \mathbf{g})(1 - p_j(a, \mathbf{b}, \mathbf{g})) \frac{1}{\sigma} \frac{\partial V_j(a, b_j + g_j)}{\partial b_j}.$$

The derivative of $p_j(a, \mathbf{b}, \mathbf{g})$ with respect to $b_{j'}$ for $j' \neq j$ is

$$\frac{\partial p_j(a, \mathbf{b}, \mathbf{g})}{\partial b_{j'}} = -p_j(a, \mathbf{b}, \mathbf{g})p_{j'}(a, \mathbf{b}, \mathbf{g}) \frac{1}{\sigma} \frac{\partial V_{j'}(a, b_{j'} + g_{j'})}{\partial b_{j'}}.$$

The derivative of average transfer with respect to b_j is

$$\frac{\partial}{\partial b_j} \left(\sum_{j' \in \mathcal{J}} p_{j'}(a, \mathbf{b}, \mathbf{g}) b_{j'} \right) = p_j(a, \mathbf{b}, \mathbf{g}) \left\{ 1 + \left(b_j - \sum_{j' \in \mathcal{J}} p_{j'}(a, \mathbf{b}, \mathbf{g}) b_{j'} \right) \frac{1}{\sigma} \frac{\partial V_j(a, b_j + g_j)}{\partial b_j} \right\}.$$

Finally, the derivative of $V(\mathbf{b}, \mathbf{g}, a)$ with respect to b_j is

$$\frac{\partial V(a, \mathbf{b}, \mathbf{g})}{\partial b_j} = p_j(a, \mathbf{b}, \mathbf{g}) \frac{\partial V_j(a, b_j + g_j)}{\partial b_j}.$$

The first order condition for b_j is

$$(1 - \delta)v'(\underline{c}_p) \left\{ 1 + \left(b_j - \sum_{j' \in \mathcal{J}} b_{j'} p_{j'}(a, \mathbf{b}, \mathbf{g}) \right) \frac{1}{\sigma} \frac{\partial V_j(a, b_j + g_j)}{\partial b_j} \right\} \leq \delta v'(\underline{c}_k) \frac{\partial V_j(a, b_j + g_j)}{\partial b_j},$$

where the inequality is strict if and only if $b_j \geq 0$ is binding. By rearranging the above condition, we get (12) when $\hat{b}_j > 0$.

C Details on Calibration

Table A3: OLS Estimates from NLSY79

	Log Annual Earnings
Completed Years of College:	
1	0.231* (0.0416)
2-3	0.337* (0.0375)
4-5	0.717* (0.0386)
AFQT Quartiles:	
2	0.204* (0.0463)
3	0.364* (0.0452)
4	0.522* (0.0477)
Experience	0.146* (0.00890)
Experience ² /100	-0.546* (0.0566)
Experience ³ /10,000	0.700* (0.106)
Constant	8.344* (0.0545)
Observations	27,694

Standard errors in parentheses

* $p < 0.05$

Table A4: Present Discounted Value of Lifetime Earnings at Age 17 (2004 U.S. dollars)

Completed Years of College (j)	$y_j(a)$ for AFQT Quartiles			
	1	2	3	4
0	248,175	304,200	356,762	418,005
1	296,053	362,919	425,651	498,746
2	312,324	382,878	449,070	526,197
4	411,402	504,391	591,632	693,283

Table A5: Average Annual Amounts for Each Year of College (2004 U.S. Dollars)

Years of College	Tuition & Fees	Grant Aid for Income Quartiles				Stafford Loan Limit
		1	2	3	4	
1	8,119	6,223	4,953	3,554	3,331	2,625
2	8,621	6,815	4,757	3,952	3,552	3,500
3	9,674	7,495	5,379	3,903	3,410	5,500
4+	9,975	8,111	5,368	4,044	3,709	5,500

Table A6: Present Discounted Values at Age 17 (2004 U.S. dollars)

Completed Years of College (j)	Cost (k_j)	Subsidy (g_j) for Income Quartiles				Borrowing Limit ($\bar{d}_j$)
		1	2	3	4	
1	7,883	6,042	4,809	3,450	3,234	2,549
2	16,009	12,466	9,293	7,176	6,582	5,848
4	33,724	26,531	18,985	14,340	12,998	15,768

References

- Abbott, Brant, Giovanni Gallipoli, Costas Meghir, and Giovanni L Violante. 2016. "Education Policy and Intergenerational Transfers in Equilibrium."
- Becker, Gary S. 1994. *Human capital*. The University of Chicago Press.
- Becker, Gary S. and Robert J. Barro. 1988. "A Reformulation of the Economic Theory of Fertility." *The Quarterly Journal of Economics* 103 (1):1–25.
- Becker, Gary S. and Nigel Tomes. 1986. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics* 4 (3):pp. S1–S39.

- Belley, Philippe, Marc Frenette, and Lance Lochner. 2014. "Post-secondary attendance by parental income in the U.S. and Canada: Do financial aid policies explain the differences?" *Canadian Journal of Economics* 47 (2):664–696.
- Bénabou, Roland. 2002. "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?" *Econometrica* 70 (2):481–517.
- Bils, Mark and Peter J. Klenow. 2000. "Does Schooling Cause Growth?" *The American Economic Review* 90 (5):1160–1183.
- Brown, Meta, John Karl Scholz, and Ananth Seshadri. 2012. "A New Test of Borrowing Constraints for Education." *The Review of Economic Studies* 79 (2):511–538.
- Browning, Martin, Pierre-André Chiappori, and Yoram Weiss. 2014. *Economics of the Family*. Cambridge University Press.
- Browning, Martin, Lars Peter Hansen, and James J. Heckman. 1999. "Chapter 8 Micro data and general equilibrium models." Elsevier, 543 – 633.
- Carneiro, Pedro, Karsten T. Hansen, and James J. Heckman. 2003. "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice." *International Economic Review* 44 (2):361–422.
- Carneiro, Pedro, James J. Heckman, and Edward J. Vytlačil. 2011. "Estimating Marginal Returns to Education." *American Economic Review* 101 (6):2754–81.
- Caucutt, Elizabeth M and Krishna B Kumar. 2003. "Higher Education Subsidies and Heterogeneity: a Dynamic Analysis." *Journal of Economic Dynamics and Control* 27 (8):1459–1502.
- Caucutt, Elizabeth M. and Lance Lochner. 2012. "Early and Late Human Capital Investments, Borrowing Constraints, and the Family." Working Paper 18493, National Bureau of Economic Research.
- Caucutt, Elizabeth M., Lance Lochner, and Youngmin Park. 2017. "Correlation, Consumption, Confusion, or Constraints: Why Do Poor Children Perform so Poorly?" *The Scandinavian Journal of Economics* 119 (1):102–147.
- College Board. 2016. "Trends in Student Aid." .
- Consumer Financial Protection Bureau. 2012. "Private Student Loans." .
- Córdoba, Juan Carlos and Marla Ripoll. 2014. "The Elasticity of Intergenerational Substitution, Parental Altruism, and Fertility Choice." Working Paper 14013, Iowa State University.
- Cunha, Flavio, James Heckman, and Salvador Navarro. 2005. "Separating Uncertainty from Heterogeneity in Life Cycle Earnings." *Oxford Economic Papers* 57 (2):191–261.
- De Fraja, Gianni. 2002. "The Design of Optimal Education Policies." *The Review of Economic Studies* 69 (2):437–466.

- Dick, Andrew W. and Aaron S. Edlin. 1997. "The Implicit Taxes from College Financial Aid." *Journal of Public Economics* 65 (3):295–322.
- Ellwood, David T. and Thomas J. Kane. 2000. "Who is Getting a College Education? Family Background and the Growing Gaps in Enrollment." In *Securing the Future: Investing in Children from Birth to College*, edited by Sheldon Danziger and Jane Waldfogel. New York: Russell Sage Foundation, –.
- Farhi, Emmanuel and Iván Werning. 2013. "Estate Taxation with Altruism Heterogeneity." *The American Economic Review* 103 (3):489–95.
- Findeisen, Sebastian and Dominik Sachs. 2017. "Optimal Need-Based Financial Aid."
- Gottlieb, Daniel and Humberto Moreira. 2012. "Should Educational Policies Be Regressive?" *Journal of Public Economic Theory* 14 (4):601–623.
- Gouveia, Miguel and Robert P. Strauss. 1994. "Effective Federal Individual Income Tax Functions: An Exploratory Empirical Analysis." *National Tax Journal* 47 (2):317–339.
- Guner, Nezih, Remzi Kaygusuz, and Gustavo Ventura. 2014. "Income Taxation of U.S. Households: Facts and Parametric Estimates." *Review of Economic Dynamics* 17 (4):559 – 581.
- Haider, Steven J. and Kathleen M. McGarry. 2012. "Parental Investments in College and Later Cash Transfers." Working Paper 18485, National Bureau of Economic Research.
- Hanushek, Eric A., Charles Ka Yui Leung, and Kuzey Yilmaz. 2014. "Borrowing Constraints, College Aid, and Intergenerational Mobility." *Journal of Human Capital* 8 (1):pp. 1–41.
- Heckman, James J, Lance Lochner, and Christopher Taber. 1998. "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents." *Review of Economic Dynamics* 1 (1):1 – 58.
- Heckman, James J., Lance J. Lochner, and Petra E. Todd. 2006. "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond." In *Handbook of the Economics of Education*, vol. 1, edited by E. Hanushek and F. Welch. Elsevier, 307 – 458.
- Hendren, Nathaniel. 2017. "Efficient Welfare Weights."
- Hicks, J. R. 1940. "The Valuation of the Social Income." *Economica* 7 (26):105–124.
- Hotz, V. Joseph, Joshua Rasmussen, Emily Wiemers, and Kate Maxwell Koegel. 2017. "The Role of Parental Wealth and Income in Financing Children's College Attendance and Its Consequences."
- Johnson, Matthew T. 2013. "Borrowing Constraints, College Enrollment, and Delayed Entry." *Journal of Labor Economics* 31 (4):669–725.
- Kaldor, Nicholas. 1939. "Welfare Propositions of Economics and Interpersonal Comparisons of Utility." *The Economic Journal* 49 (195):549–552.
- Keane, Michael P and Kenneth I Wolpin. 2001. "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment." *International Economic Review* 42 (4):1051–1103.

- Krueger, Dirk and Alexander Ludwig. 2016. "On the Optimal Provision of Social Insurance: Progressive Taxation versus Education Subsidies in General Equilibrium." *Journal of Monetary Economics* 77:72 – 98.
- Lochner, Lance J. and Alexander Monge-Naranjo. 2011. "The Nature of Credit Constraints and Human Capital." *American Economic Review* 101 (6):2487–2529.
- McFadden, Daniel. 1978. "Modeling the Choice of Residential Location." In *Spatial Interaction Theory and Planning Models*, edited by A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull. Amsterdam: North Holland, 75–96.
- McPherson, Michael S. and Morton Owen Schapiro. 1998. *The Student Aid Game: Meeting Need and Rewarding Talent in American Higher Education*. Princeton University Press.
- Milgrom, Paul and Chris Shannon. 1994. "Monotone Comparative Statics." *Econometrica* 62 (1):pp. 157–180.
- Restuccia, Diego and Carlos Urrutia. 2004. "Intergenerational Persistence of Earnings: The Role of Early and College Education." *The American Economic Review* 94 (5):1354–1378.
- Restuccia, Diego and Guillaume Vandenbroucke. 2013. "The Evolution of Education: A Macroeconomic Analysis." *International Economic Review* 54 (3):915–936.
- Sallie Mae. 2017. "How America Pays for College 2017: Sallie Mae's 10th National Study of College Students and Parents." .
- Weinzierl, Matthew. 2008. "Optimal Policy with Heterogeneous Parental Altruism." In *Essays in Optimal Taxation*, Ph.D. thesis, chap. 3. Harvard University, 166–204.