HCEO WORKING PAPER SERIES

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

Youngmin Park

Working Paper

2019-004

02/2019

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

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

Youngmin Park

Bank of Canada

February 6, 2019

Abstract

This paper studies optimal education subsidies when parental transfers are unequally distributed across students and cannot be publicly observed. After documenting substantial inequality in parental transfers among US college students with similar family resources, I examine its implications for how the education subsidy should vary with schooling level and family resources to minimize inefficiencies generated by borrowing constraints. Unobservable heterogeneity in parental transfers creates a force to heavily subsidize low schooling levels chosen by borrowing-constrained students with low parental transfers. This force is stronger for rich families, but it is weakened if heterogeneity in returns to schooling also leads to different schooling choices. These mechanisms are quantified using a calibrated model. Quantitative analysis suggests a reform that reallocates public spending toward the first two years of college. The reform also reduces the gap in subsidy amounts by parental income during early years of college.

*This paper is based on the second chapter of my PhD dissertation at the University of Western Ontario that was 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. For their comments, I thank seminar participants at the University of Western Ontario, Korea Development Institute, Korea Institute of Public Finance, University of New Brunswick, Bank of Canada, Queen's University, Federal Reserve Bank of Atlanta, and University of Connecticut and participants at the 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), 2018 Asian Meeting of the Econometric Society, and 2018 Annual Meeting of the Society for Economic Dynamics. 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

College education is costly in the US, and many students receive financial help from their parents to pay for it. During the 2016–2017 academic year, the average cost of college for American undergraduate students was $23,757, of which 31% was covered by parents, while 35% was paid by grants and scholarships and 19% by student borrowing (Sallie Mae, 2017). However, parental contributions toward college expenses differ greatly across students (Haider and McGarry, 2012).[1] While some of the differences in parental support across families reflect disparities in available economic resources (e.g., Ellwood and Kane, 2000; Johnson, 2013), little is known about the extent to which parental transfers during school vary among families with similar resources.

The current need-based financial aid system in the US recognizes the differences in families' financial ability to contribute, and more financial aid is awarded to students with low family resources.[2] Yet, there might still be variation in parental support conditional on family resources, as parental transfer decisions could be also driven by other factors, such as parents' preferences for giving. As noted by Brown, Scholz, and Seshadri (2012), this raises a concern that the current need-based financial aid system may not effectively target students with low financial support from parents. In this paper, I provide empirical evidence that there exists substantial inequality in parental transfers among students with similar family resources and show that recognizing that parental support may differ across families with the same resources has important implications for financial aid policy.

Using data from the 2011–2012 National Post-Secondary Student Aid Survey (NPSAS:12), I first document that parental transfers are very unequally distributed across American undergraduate students, even among those considered to have similar family resources for financial aid purposes. Controlling for family resources, I also demonstrate that 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 indicate greater difficulties in making ends meet. These outcomes also imply a higher utility cost of college, which is likely to discourage college attendance for youth expecting low financial support from their parents.

Motivated by this evidence, I next analyze optimal financial aid when there exists inequality in parental transfers conditional on family resources and students face limited borrowing opportunities. Inefficiency due to borrowing constraints naturally justifies the existence of financial aid in the form of student loans and non-repayable "subsidies" such as grants and scholarships.[3] The borrowing constraints can reflect underlying frictions in the credit market that restrict intertemporal trade between individuals, such as limited commitment to loan repayment (e.g., Kehoe and Levine, 1993; Lochner and Monge-Naranjo, 2011). Borrowing to invest in human capital can be particularly difficult because

[1]They document 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]While 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).

[3]Although there is some disagreement about the importance of borrowing constraints for higher education given current financial aid policy (see Lochner and Monge-Naranjo (2012) for a recent review), few argue that such constraints would be irrelevant if all financial aid were removed.

such capital cannot be acquired and sold by the creditor in case of non-repayment, and thus serves as poor collateral on loans (Becker, 1975). Because of these repayment issues related to lending, it might be challenging to further expand the government student loan program to directly relax borrowing constraints.[4] Appealing to such frictions, this paper takes existing student loans as given and considers restructuring the subsidy component of the current financial aid policy in a budget-neutral way.

For analytical characterization of optimal policy, I begin with a simple two-period model of schooling choice, in which otherwise identical students are endowed with different amounts of parental transfers that are imperfectly correlated with family resources. The differences in parental transfers affect schooling choice because students cannot borrow to pay for direct schooling costs and consumption while in school. As discussed by Becker (1975) and Lochner and Monge-Naranjo (2011), students with sufficiently high parental transfers are not borrowing constrained and choose the schooling level that maximizes lifetime earnings net of schooling costs. Those with lower parental transfers may choose lower schooling levels that are less costly because they must suffer from inefficiently low consumption during school when the borrowing constraint binds.

The social planner aims to minimize inefficiencies caused by borrowing constraints—distorted schooling investment and intertemporal consumption allocations. Since these inefficiencies are reflected in a standard utilitarian social welfare function, it could serve as a social objective function to be maximized. However, such a criterion also rewards equity in lifetime consumption, thus it would justify redistribution even without any market imperfections. Although reducing consumption inequality is an important objective for tax and transfer policies, it may not be a direct goal of college financial aid policies, which could increase earnings inequality by helping able students attend college.[5] To separate efficiency from equity concerns, I construct a monetary measure of distortions that reflects an individual's maximum willingness to pay to eliminate them; the measure is then linearly aggregated across individuals to give a social objective function to be minimized.[6]

Notice that the social objective is not to maximize schooling or even lifetime earnings, but it is to both efficiently invest in schooling and smooth consumption over the life cycle. In some cases, a greater distortion in schooling may be worthwhile if it provides enough improvement in consumption smoothing. For example, promoting college attendance is not necessarily desirable if it means that many college students have to suffer from exceptionally low consumption while in school. This is often ignored in policy discussions, where the primary focus is on improving education outcomes. The social objective proposed in this paper transparently captures the key trade-off faced by individuals as well as the social planner in the economy: education investment vs. consumption smoothing.

The social planner is endowed with a budget for education subsidies. The key challenge for the

[4]Although student loans cannot generally be expunged through bankruptcy in the US, many borrowers default by stopping making payments. According to the official cohort default rate for federal student loans produced by the US Department of Education, 10.8% of borrowers who entered repayment in 2015 defaulted within the following three years.

[5]For example, Hanushek, Leung, and Yilmaz (2003) find that education subsidies are generally inferior to other transfer schemes in achieving distributional objectives.

[6]Bénabou (2002) develops a similar measure of aggregate efficiency that abstracts from equity concerns. While it reflects the level of aggregate resources and uninsurable idiosyncratic risks, the efficiency of intertemporal smoothing is not taken into account (See Appendix B.2 for comparison). His measure forms the basis for the welfare decomposition method of Flodén (2001) that is widely used in macroeconomics. See, for example, Heathcote, Storesletten, and Violante (2008, 2017) and Abbott et al. (forthcoming) for applications.

subsidy design is that the planner can only observe individuals' schooling choices and their family resources; neither parental transfers nor consumption/savings are observable. This assumption reflects that, compared with income or assets, it is more difficult for a third party to observe and verify transfers between family members. Therefore, subsidy amounts depend solely on schooling levels and family resources. This feature of the policy is consistent with the current need-based financial aid system, where financial need of a student for a given year is determined by schooling cost and family resources.

I begin characterizing how the (total) amount of subsidy students receive should depend on their schooling choices when everyone has identical family resources and returns to schooling. The main theoretical result is that unobservable heterogeneity in parental transfers encourages subsidies at low levels of schooling. In particular, I show that the *optimal amount of subsidy decreases in schooling levels*. If parental transfers were publicly observable, then giving large lump-sum transfers to borrowing-constrained students with low parental transfers would reduce aggregate distortions. Although this is not feasible because information about parental transfers is private, those with low parental transfers can be effectively targeted with high subsidies at low schooling levels, exploiting the variation in schooling choice induced by inequality in parental transfers and borrowing constraints. The optimal policy redistributes toward those with low parental transfers, but the redistribution is not driven by social concerns about equity. The high optimal subsidy at low schooling levels does not depend on the magnitude of the causal effects of schooling on earnings either; it is based on the robust result that exogenously set borrowing limits lead to under-investment in schooling.[7]

The force to reallocate public spending toward lower schooling levels is stronger for "rich" families with high family resources. With two schooling levels, I show that the *optimal subsidy for the low schooling level increases with family resources* if the budget for education subsidy is held constant across family resource levels. Compared with poor families, there exists a lower fraction of students among rich families who are borrowing constrained due to low parental transfers and thus choose the low level of schooling. Therefore, for rich families, it is less costly to award high subsidies per student at the low schooling level. This provides a mechanism that weakens the dependence of subsidy amounts on family resources at low levels of schooling when the budget is set optimally across family resources: although it might be efficient to spend less on students from rich families who receive high parental transfers on average, a higher fraction of the budget for rich families should be directed toward low schooling levels. This illustrates that inequality in parental transfers conditional on family resources also affects how optimal subsidies vary across families with different resources.

These properties of the optimal policy contrast with current US need-based financial aid policy, which offers less aid to students with high family resources and more aid (over the lifetime) for staying in school longer or attending costlier institutions. The theoretical result critically hinges on the condition that the allocations of students choosing lower schooling levels are more distorted by borrowing constraints, which in turn relies on the assumption that students have identical returns to schooling and differ only in the amount of transfers received from their parents. However, considerable evidence suggests that the returns to schooling vary substantially across individuals, which also

[7]As Lochner and Monge-Naranjo (2011) show, this does not necessarily hold when borrowing limits increase one-for-one with the cost of investment.

impacts schooling decisions (e.g., Carneiro, Hansen, and Heckman, 2003; Cunha, Heckman, and Navarro, 2005; Carneiro, Heckman, and Vytlacil, 2011). Individual differences in returns to schooling can be another source of private information. I demonstrate that the features of current policy can be justified as optimal if parental transfers are perfectly correlated with family resources and differences in educational attainment conditional on family resources are driven only by unobservable heterogeneity in returns to schooling. In this case, the *optimal subsidy increases in schooling levels* because, among students with identical parental transfers, those choosing higher schooling levels due to higher schooling returns are more likely to be borrowing constrained.

More generally, when students with identical family resources can differ along both dimensions (parental transfers and returns to schooling), the social planner must balance the relative importance of the two types of unobservable heterogeneity in designing subsidies because they have opposite implications for how subsidy amounts should vary with schooling levels. The overall structure of optimal subsidies will depend on where in the joint distribution of schooling and family resources most students are borrowing constrained, which is shaped by the nature of unobservable heterogeneity leading to the differences in schooling choices. This suggests that understanding what drives differences in educational attainment across individuals is crucial for designing efficient education policy.

To quantify the theoretical insights, I extend the two-period model to a multi-period life-cycle setting, where schooling choices are represented by the highest year of college completed, and calibrate it to the US economy. The quantitative model incorporates two types of heterogeneity in returns to schooling that are found to be important in the literature: heterogeneity in wage (or monetary) returns, which I call "ability," and heterogeneity in psychic returns (i.e., tastes). 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. As in Brown, Scholz, and Seshadri (2012), the degree of parental altruism is allowed to differ across families, which is key to generating inequality in parental transfers conditional on observed family characteristics.[8] Model parameters are chosen to replicate the joint distribution of educational attainment, parental income, parental transfers, and measured ability of American youth from the National Longitudinal Surveys of Youth 1997 (NLSY97) under current financial aid policy. To parsimoniously capture the need-based feature, subsidy amounts under the current financial aid policy are estimated separately by each year of college and quartile of parental income using financial aid records from the 2003–2004 National Post-Secondary Student Aid Survey (NPSAS:04).

Based on the calibrated model, I solve for the optimal subsidy that varies with years of college and parental income quartiles given the same budget as current policy. Compared with current policy, the optimal policy provides larger subsidies to those who attain one or two years of college education, while providing lower subsidies for those completing four years. The optimal subsidy for the first two years of college education varies less among low and middle-income families, providing similar public support for some college education across parental income. As a result, more youth attend college but fewer of them complete four years, and the dispersion in educational attainment is reduced. These

[8]Appendix C.2 summarizes evidence consistent with heterogeneity in parental altruism. Alternative modeling assumptions on parental behavior and their policy implications are also discussed.

results highlight the role of heterogeneous parental support conditional on income that encourages high subsidies at low schooling levels.

The optimal policy improves efficiency in both schooling investment and intertemporal consumption allocation, with a sizable gain in overall efficiency: the efficiency gain from the budget-neutral restructuring of the current policy is half as large as the efficiency loss from removing the current policy. Most of the efficiency gain comes from better consumption smoothing, suggesting that evaluating policies solely based on education outcomes can be misleading from the efficiency point of view.

While there is extensive literature that quantitatively evaluates higher education policies in the presence of parental transfers (e.g., Keane and Wolpin, 2001; Caucutt and Kumar, 2003; Restuccia and Urrutia, 2004; Johnson, 2013; Hanushek, Leung, and Yilmaz, 2014; Abbott et al., forthcoming; Caucutt and Lochner, forthcoming), few studies explore the optimal policy design. On the other hand, the literature on optimal education and human capital policies typically abstracts from intergenerational linkages (e.g., Bohacek and Kapicka, 2008; Findeisen and Sachs, 2016; Stantcheva, 2017) or only considers policies that do not depend on family background (Krueger and Ludwig, 2016).

One exception is Colas, Findeisen, and Sachs (2018), who both theoretically and quantitatively characterize how optimal subsidy for college education varies with parental income. Although their work is most closely related to this paper, it differs in three important respects. First, they do not incorporate unobserved heterogeneity in parents' preferences for giving. Second, subsidy amounts do not vary with years of college. Third, the main rationale for the need-based financial aid system is to eliminate the adverse impact of redistributive taxes on education investment. This paper complements their work by motivating the need-based financial aid system based on borrowing constraints and by exploring how it is affected when those in need are not easily identified due to heterogeneity in parental preferences. As discussed earlier, awarding differential financial aid by year of college among students with identical family resources is a crucial policy instrument for effectively targeting those who are borrowing constrained because of low parental support.

Providing larger subsidies for higher levels of education investment—a feature of current US financial aid policy—is often justified on the grounds that it would help restore efficiency of human capital investment for borrowing-constrained individuals by encouraging them to invest more than they otherwise would. However, existing theories of optimal education subsidy that derive this feature rely on factors other than borrowing constraints, such as positive external effects of human capital (De Fraja, 2002) and equity concerns that give rise to redistributive taxes (Bovenberg and Jacobs, 2005). The results presented in this paper imply that borrowing constraints can indeed deliver this feature when they bind more severely for those who invest more. Yet, borrowing constraints need not justify a policy that encourages investment at the margin when they are more likely to bind for those with low investment levels. This is the case when unobservable heterogeneity in available resources for investment, rather than returns to investment, drives differences in investment decisions.

The rest of this paper is organized as follows. Section 2 reports US evidence on parental transfers and family resources. Section 3 uses a two-period model to analyze the optimal education subsidy. Section 4 extends the model to a multi-period life-cycle setting with endogenous parental transfers and quantitatively characterizes the optimal subsidies. Section 5 concludes.

2 Evidence on Parental Transfers and Family Resources

In this section, I examine inequality in parental transfers across families with similar resources using the NPSAS:12 data. The NPSAS:12 is a nationally representative survey of students enrolled during the 2011–2012 academic year in US post-secondary institutions. An important feature of this data is that it contains information about family resources, as measured by the "expected family contribution" (EFC), as well as how much parents helped pay for college expenses.

The EFC is a number that determines a student's eligibility for federal financial aid, and it is intended to measure what "the student's family may be reasonably expected to contribute toward the student's postsecondary education for the academic year."[9] It is calculated according to a formula specified by law, using the information—such as income, assets, family size—the student provides on the Free Application for Federal Student Aid that must be filled out by all students seeking federal aid. For dependent students, a category under which most traditional students fall, the EFC is primarily determined by previous year's parental income, with the net worth of parents playing only a minor role.[10] For example, home equity for primary residence and retirement assets are excluded from the EFC calculation. The EFC is an important determinant of financial aid in the US, where most financial aid is need-based: federal rules determine federal aid based on student need, and similar need calculation is used for state and institutional aid. For federal financial aid purposes, a student's financial "need" equals the difference between the "cost of attendance" and the EFC. The cost of attendance is the total cost of attending a given college that includes tuition and living expenses, and it is determined by institutions. Generally, the amount of financial aid awarded is increasing in student need (Dick and Edlin, 1997).

In the NPSAS:12 data, the EFC and other variables related to financial aid come from administrative records. Parental transfer amounts, available in 12 categories, are 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?" To ensure access to administrative data on financial aid variables, I only consider dependent students who had submitted a federal financial aid application. The cost of attendance can vary substantially because of factors such as institution type and enrollment status, which may affect the amount of financial aid as well as parental transfers among students with similar EFC. To control for the variation in parental transfers driven by differences in cost, I apply additional sample restrictions and focus on a group of students facing relatively similar costs of attendance. I select US citizens who were enrolled in a single four-year public institution more than nine months full time during the 2011–2012 academic year and paying the regular "in-jurisdiction" tuition fees. I also exclude students who lived at home while enrolled because they had substantially lower cost of attendance from savings in room and board charges.

[9]See Part F of Title IV of the Higher Education Act for further details on the needs analysis formulas used to award federal financial aid. The rules for the 2011–2012 academic year are explained in Federal Student Aid (2011).

[10]Students are considered to be financially independent of their parents for federal financial aid purposes if they meet any of the following criteria: are age 24 or older; are married; have legal dependents; are veterans of the US armed forces or on active duty; or are orphans or wards of the court. All other students are considered to be dependent unless they can prove that they are receiving no parental support and are determined to be independent by a financial aid officer using professional judgement.

Table 1: Sample Descriptive Statistics

Variables	Mean
Tuition and fees ($)	8,198
Cost of attendance ($)	22,336
Grant ($)	5,771
Net cost of attendance ($)	16,565
EFC ($)	12,552
Parents' income in 2010 ($)	78,493
Own income in 2010 ($)	3,132
Hours worked per week while enrolled	10.1
Federal student loan amount ($)	4,647
Took out federal student loans (%)	74.6
Took out maximum amount of federal student loans (%)	42.5
Private student loan amount ($)	524
Took out private student loans (%)	9.4

Descriptive statistics for this sample are provided in Table 1.[11] For the 2011–2012 academic year, the average cost of attendance is $22,336, of which tuition and fees account for 37%. After excluding all grants (and scholarships), which cover 26% of the cost, the average "net cost" is $16,565. The average amount of EFC, or what students and their parents are expected to contribute, is 76% of the average net cost, but most of the EFC reflects parents' expected contribution, as students' income is very low on average. Around three-quarters of students take out federal student loans, and students borrow $4,647 on average, which is enough to cover the difference between the average net cost and the average EFC. Although students can borrow up to the net cost through the government student loan program, it has fixed maximum borrowing limits: for dependent students in 2011–2012, the annual limit is $5,500 for first-year students, $6,500 for second-year students, and $7,500 for third-year students and beyond. Forty-three percent of students borrow as much as possible from the government. Markets for private student loans exist, but it is generally more difficult and costly to borrow from private lenders. Most private student loans require evidence of creditworthiness, which many undergraduate students would fail to provide, and charge higher interest rates than those offered by government student loans (Consumer Financial Protection Bureau, 2012). As a result, only 9% of students take out private student loans and the average amount borrowed is fairly small.

[11] All results based on the NPSAS:12 data use the sample weights to account for the sampling scheme of the survey and are obtained using the table creation tool on the National Center for Education Statistics (NCES) DataLab website. Direct access to the data is granted only to US researchers.

2.1 Parental Contribution Conditional on Family Resources

Table 2 reports the distribution of parental contribution conditional on EFC, where both variables are categorized into six groups based on the same thresholds. It also shows the average parental income and grant aid (including scholarships) received conditional on EFC. Students with higher EFC have higher parental income and receive lower grants on average. They also tend to receive higher parental transfers on average, as they are more likely to receive high amounts while less likely to receive low amounts. However, actual parental contributions are generally not in line with EFCs for the majority of students. For example, 12% of students with more than $20,000 EFC do not receive anything from their parents, while 16% of parents with zero EFC give more than $5,000 to their children.[12]

Table 2: EFC and Actual Parental Contribution

EFC	% with Amount Parents Paid toward Expenses						Parental Income ($)	Grant ($)
	$0	$1 to $2,000	$2,001 to $5,000	$5,001 to $10,000	$10,001 to $20,000	$20,001 or More		
$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

The inequality in parental transfers conditional on EFC is likely to reflect differences across families that cannot be easily captured by the information used to calculate the EFC. The analysis in Appendix A.2 further demonstrates that it is difficult to predict parental contribution even when additional variables that are not directly related to financial factors, such as demographic characteristics and academic performance of students are used, although there are some variables such as parental education that have significant effects on parental contributions.

Of course, it is also possible that the variation in parental transfers conditional on EFC reflects pure measurement error because students may not provide accurate answers to the survey question. Measurement error is unlikely to be correlated with other variables, especially financial aid variables from administrative records. In contrast, actual parental contributions are likely to be correlated with how much students borrow or work in order to pay for college. To see whether the variation in reported parental contributions reflects measurement error, I therefore examine how they are correlated with student borrowing and labor supply.

[12] Appendix A.1 shows that the same patterns arise when stricter sample selection criteria are applied in order to control for other differences related to college financing that might drive differences in parental transfers conditional on EFC.

2.2 Effects of Parental Contribution on College Financing Decisions

Among students facing the same net cost, those who receive lower parental transfers must find other ways to pay for college. One way is to borrow more through the government student loan program. However, the borrowing limits might be too tight, especially for those with a high EFC facing high net costs. For example, students in the highest EFC category receive $2,193 in grant aid on average and face an average net cost of $20,143. This is more than triple what first-year students can borrow ($5,500). Government loan limits are more likely to be binding for students with little parental help, causing some students to seek additional loans from private lenders or to work more.

Table 3 reports ordinary least squares (OLS) estimates for the effects of parental contributions on how students finance their college costs, controlling for EFC and net costs. Columns 1 and 2 show results for taking out the maximum amount of federal student loans and private student loans, respectively. Column 3 reports results for hours worked per week while enrolled. These estimates show that, conditional on EFC and net costs, those with less help from parents are more likely to exhaust government student loans, turn to private lenders, and work longer hours while enrolled.

These outcomes are also more likely for students facing higher net costs, who need to finance greater amounts. However, once parental contributions and net costs are accounted for, the EFC has no discernible impacts on labor supply and the likelihood of "maxing out" federal student loans (except for the effect of $20,001 or more EFC). The EFC matters for private loan usage, which might reflect the fact that rich parents, who can easily demonstrate their creditworthiness, can help obtain a private loan with a low interest rate by co-signing with their children.[13]

2.3 Challenge in Designing Financial Aid Policies

This evidence poses a challenge for the design of financial aid policies. Given the substantial differences in parental contributions among parents with similar resources, the current need-based financial aid system may not be able to effectively address the inequality in education opportunities originating from unequal parental support. Since the analysis, thus far, has focused on youth attending four-year public universities, the inequality among all youth—including those not attending college—is likely to be even greater, if parental transfers affect college attendance.

Given limited borrowing opportunities, college students whose parents contribute less than what is expected by financial aid authorities suffer from low consumption. Providing financial aid based on actual, rather than expected, parental contributions would be more effective. However, this is challenging because parental transfers cannot be directly observed by financial aid authorities and families may not report transfer amounts truthfully. This paper explores how the amount of subsidy, distributed through financial aid, should vary with family resources and schooling choices in the presence of the unobservable differences in parental transfers.

[13] Information about the existence of a co-signer for private student loans is not available in the NPSAS:12 data.

Table 3: 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	-0.012	-3.165*
	(0.019)	(0.014)	(0.538)
$2,001 to $5,000	-0.033	-0.055*	-3.863*
	(0.024)	(0.016)	(0.629)
$5,001 to $10,000	-0.075*	-0.040*	-4.327*
	(0.030)	(0.018)	(0.708)
$10,001 to $20,000	-0.107*	-0.087*	-5.992*
	(0.027)	(0.017)	(0.717)
$20,001 or more	-0.124*	-0.076*	-4.843*
	(0.035)	(0.020)	(0.881)
EFC:			
$1 to $2,000	-0.046	0.011	-0.761
	(0.034)	(0.016)	(0.838)
$2,001 to $5,000	-0.006	0.051*	-1.246
	(0.028)	(0.014)	(0.638)
$5,001 to $10,000	-0.013	0.056*	0.216
	(0.029)	(0.016)	(0.726)
$10,001 to $20,000	-0.025	0.059*	-0.078
	(0.034)	(0.017)	(0.776)
$20,001 or more	-0.132*	0.037*	-0.240
	(0.032)	(0.015)	(0.673)
Net cost of attendance:			
$10,001 to $15,000	0.117*	0.033*	1.843*
	(0.024)	(0.011)	(0.684)
$15,001 to $20,000	0.204*	0.065*	2.402*
	(0.027)	(0.013)	(0.683)
$20,001 or more	0.245*	0.110*	1.657*
	(0.029)	(0.017)	(0.729)
Constant	0.346*	0.033*	11.843*
	(0.023)	(0.009)	(0.599)

Notes: Estimated using OLS. Columns (1) and (2) are linear probability models. Sample weights are used. Standard errors in parentheses.

* Significant at the 5% level.

3 Theory of Optimal Education Subsidies

In this section, I consider a simple two-period model to analytically characterize education subsidies that minimize inefficiencies caused by borrowing constraints in the presence of unobservable heterogeneity in parental transfers. To attain clean analytical results, this section makes a number of simplifying assumptions, such as exogeneity of parental transfers. The quantitative analysis in Section 4 shows that the insights developed here hold more generally.

I first show how the optimal subsidy varies with schooling level among families with identical resources. Next, I examine how it varies across family resources. Finally, I introduce unobservable heterogeneity in returns to schooling as another determinant of schooling and contrast its policy implications with those of the unobservable heterogeneity in parental transfers.

3.1 Modeling Schooling Choice

Consider individuals who invest in schooling in the first period and work in the second period. Their preferences are represented by the following utility function:

$$U = \ln c_1 + \ln c_2, \tag{1}$$

where c_t is consumption in periods $t \in \{1, 2\}$. Each individual is endowed with a parental transfer $b \in \mathbb{R}_+$, which is positively, but imperfectly, correlated with family resources. Parental transfers are exogenously given and thus can be thought of as initial wealth.

For now, assume that individuals have identical monetary returns to schooling. They face the choice set of schooling $\mathcal{J}$, which is finite and totally ordered. The cardinality of the choice set, $|\mathcal{J}|$, represents the number of choices available. Each schooling choice $j \in \mathcal{J}$ is associated with earnings at $t = 2$, $y_j \in \mathbb{R}_+$, as well as a monetary cost, $k_j \in \mathbb{R}$, that needs to be paid at $t = 1$. Schooling choices are ordered such that $y_j < y_{j'}$ for all $j < j'$. Moreover, schooling costs are related to earnings through $k_j = k(y_j)$ for some strictly increasing function $k : \mathbb{R}_+ \to \mathbb{R}$. Therefore, higher schooling levels lead to higher earnings, although they are more costly.

Individuals can save at a zero interest rate, but they cannot borrow.[14] Therefore, the schooling cost and consumption during schooling must be financed out of parental transfers. Conditional on schooling j, the indirect utility of an individual with parental transfer b is

$$U_j(b) \equiv \max_{(c_1,c_2)\in\mathbb{R}^2_+} \Big\{ \ln c_1 + \ln c_2 \Big\},$$

$$\text{subject to } c_1 + c_2 \leq b - k_j + y_j, \tag{2}$$

$$c_1 \leq b - k_j, \tag{BC}$$

where (2) is the lifetime budget constraint and (BC) is the borrowing constraint.

When $b < k_j$, consumption during school cannot be positive, so the choice j is not feasible. Let $\mathcal{J}(b) \equiv \{j \in \mathcal{J} | k_j \leq b\}$ be the set of feasible choices for those with parental transfer b. For a feasible

[14]Zero interest rate, time preference rate, and borrowing limit are not crucial; they are assumed to simplify algebra.

choice $j \in \mathcal{J}(b)$, the indirect utility is

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

When $b \geq k_j + y_j$, parental transfers are large enough to ensure full consumption smoothing over the life cycle: $c_1 = c_2 = (b - k_j + y_j)/2$. On the other hand, when $b < k_j + y_j$, the borrowing constraint (BC) binds, and consumption during school is too low: $c_1 = b - k_j < c_2 = y_j$.

The schooling choice problem is

$$U(b) \equiv \max_{j \in \mathcal{J}(b)} U_j(b). \tag{3}$$

The following lemma shows the relationship between parental transfers and schooling choice. See Appendix B for all proofs and other analytical details.

Lemma 1. $\operatorname{argmax}_{j \in \mathcal{J}(b)} U_j(b)$ *is increasing in b.*[15]

This is a discrete choice version of the well-known result established by Becker (1975) and Lochner and Monge-Naranjo (2011). In the absence of borrowing constraints, individuals would choose the schooling option that maximizes their lifetime consumption. With borrowing constraints, however, the consumption-maximizing schooling option may not be feasible for those with low parental transfers. Moreover, even when it is feasible, individuals may suffer from low consumption during school because of the binding borrowing constraint, which lowers the option's utility value. Increasing parental transfers affects the schooling choice by making more schooling options feasible, and by making feasible options more valuable by allowing for better consumption smoothing.

The negative effect of borrowing constraints on educational attainment is central to the optimal design of education subsidies in this paper.[16] Because otherwise identical individuals make different schooling choices owing to differences in parental transfers, the variation in schooling choices enables one to infer who receives less and who receives more parental transfers. Therefore, awarding potentially different subsidy amounts to those attaining different levels of schooling could be an effective policy instrument for screening individuals with different parental transfers.

3.2 Social Objectives

Now I discuss the goal of education subsidies. Notice that there are two sources of inefficiencies for borrowing-constrained individuals. First, lifetime consumption is reduced if they do not choose the schooling level that maximizes it. Second, a given level of lifetime consumption yields lower lifetime

[15]The solution to the schooling choice problem may not be unique, in which case $\operatorname{argmax}_{j \in \mathcal{J}(b)} U_j(b)$ is a set containing multiple elements. To compare sets, I follow the standard definition in the literature, called the "strong set order." A set X' is defined to be greater than or equal to X in the strong set order if for any $x' \in X'$ and $x \in X$, we have $\max\{x', x\} \in X'$ and $\min\{x', x\} \in X$. Throughout the paper, "increasing" and "decreasing" mean "non-decreasing" and "non-increasing," respectively.

[16]See Stinebrickner and Stinebrickner (2008) and Brown, Scholz, and Seshadri (2012) for empirical evidence.

utility if they are unable to fully smooth consumption over time. Since these inefficiencies are already reflected in individual utilities, the utilitarian social welfare function could serve as a measure of aggregate efficiency to be maximized. However, the utilitarian social welfare function also depends on the level of consumption inequality when the utility function is strictly concave: as a result of Jensen's inequality, any redistribution will lead to higher utilitarian social welfare, even when there are no market imperfections that would justify such reallocation.[17] Therefore, the social planner who maximizes the utilitarian social welfare function is "inequality averse" and aims to not only improve aggregate efficiency but also reduce inequality.

To separate efficiency from equity concerns, Bénabou (2002) constructs a measure of aggregate efficiency that puts no value on equity of consumption per se. His measure of aggregate efficiency is the utility of a representative agent whose consumption equals aggregate consumption in each period. While this approach takes into account the loss in aggregate consumption due to distorted schooling choice, the welfare loss from imperfect consumption smoothing over time is understated because it is evaluated from the perspective of the representative agent.[18]

Instead, I measure distortions in monetary units by individuals' maximum willingness to pay to eliminate them and aggregate this across individuals to form an aggregate measure of inefficiency. Since my measure of aggregate inefficiency does not directly depend on the degree of consumption inequality, it can be improved by redistribution only to the extent that distortions in schooling investment or intertemporal consumption smoothing are reduced. Because the measure of aggregate distortions is the sum of individuals' willingness to pay, an allocation that minimizes it satisfies the Kaldor-Hicks criterion of efficiency, a situation where there is no reallocation that generates winners who could compensate losers to achieve a Pareto-improving outcome (Kaldor, 1939; Hicks, 1940).

In order to construct a monetary measure of distortions, I first turn utils into dollars through the expenditure function $e : \mathbb{R} \cup \{-\infty\} \to \mathbb{R}_+$ that gives the minimum cost to achieve a given utility level U:

$$e(U) \equiv \inf_{(\tilde{c}_1, \tilde{c}_2) \in \mathbb{R}_+^2} \left\{ \tilde{c}_1 + \tilde{c}_2 \middle| \ln \tilde{c}_1 + \ln \tilde{c}_2 \geq U \right\} = 2 \exp\left(\frac{U}{2}\right). \tag{4}$$

Then $e(\ln c_1 + \ln c_2) = 2(c_1 c_2)^{1/2}$ is the "money-metric" utility function (McKenzie, 1957; Samuelson, 1974) that measures individual welfare in monetary terms. Notice that $e(\ln c_1 + \ln c_2)$ represents the same preference as $\ln c_1 + \ln c_2$, since $e(\cdot)$ is a monotonic transformation. Moreover, $e(\ln c_1 + \ln c_2)$ is equal to lifetime consumption $c_1 + c_2$ if $c_1 = c_2$, and strictly less than $c_1 + c_2$ if $c_1 \neq c_2$, reflecting the welfare loss from the intertemporal consumption distortion.[19] Because individuals are indifferent between a potentially distorted consumption profile (c_1, c_2) and an undistorted consumption profile $((c_1 c_2)^{1/2}, (c_1 c_2)^{1/2})$, they are willing to pay as much as $c_1 + c_2 - 2(c_1 c_2)^{1/2}$ to eliminate the intertemporal consumption distortion.

[17]Notice that the strict concavity of a utility function is a cardinal property that depends on a particular representation of the underlying preference. Therefore, the curvature of the utility function represents social preferences for equity rather than individual preferences on intertemporal consumption smoothing.

[18]See Appendix B.2 for details.

[19]In other words, the standard identity $e(\ln c_1 + \ln c_2) = c_1 + c_2$ holds if and only if the borrowing constraint does not bind.

Formally, I define distortions for a particular individual by using a lottery $\boldsymbol{p} \equiv (p_j)_{j \in \mathcal{J}}$ that characterizes her schooling choice in terms of choice probabilities.[20] Each lottery lies in the $|\mathcal{J}|$-dimensional probability simplex $\Pi \equiv \{\boldsymbol{p} \in \mathbb{R}_+^{|\mathcal{J}|} | \sum_{j \in \mathcal{J}} p_j = 1\}$. Consider an individual with a parental transfer b, and let $\boldsymbol{p}(b) \in \Pi$ be a lottery that is consistent with her schooling choice. For all feasible choices $j \in \mathcal{J}(b)$, let $V_j(b) \equiv e(U_j(b))$ be the money-metric indirect utility. Then the *distortion in intertemporal consumption allocation* for the individual is defined as follows:

$$\tau_c(b) \equiv \sum_{j \in \mathcal{J}(b)} p_j(b)\left[(y_j - k_j + b) - V_j(b)\right].$$

When the individual chooses a feasible schooling level j, she consumes $y_j - k_j + b$ and attains utility $U_j(b)$ over the lifetime. However, she could have attained the same level of utility if she consumed $V_j(b)/2$ in each period. Therefore, $(y_j - k_j + b) - V_j(b)$ is her maximum willingness to pay to eliminate the intertemporal consumption distortion. Taking the average across feasible schooling options, weighted by choice probabilities, yields $\tau_c(b)$.

Next, the *distortion in schooling allocation* measures the lost lifetime consumption from not choosing the consumption-maximizing schooling:

$$\tau_s(b) \equiv \max_{j \in \mathcal{J}}\{y_j - k_j\} - \sum_{j \in \mathcal{J}} p_j(b)(y_j - k_j).$$

Finally, the *total distortion* is the sum of the two distortions:

$$\tau(b) \equiv \tau_c(b) + \tau_s(b) = \max_{j \in \mathcal{J}}\{y_j - k_j\} + b - \sum_{j \in \mathcal{J}(b)} p_j(b)V_j(b).$$

It can be rearranged to give the following formula for welfare decomposition:

$$V(b) = \max_{j \in \mathcal{J}}\{y_j - k_j\} + b - \tau(b), \tag{5}$$

which shows that the monetary measure of welfare, $V(b) \equiv e(U(b))$, consists of the maximum lifetime consumption that can be attained in the absence of borrowing constraints, $\max_{j \in \mathcal{J}}\{y_j - k_j\} + b$, net of the distortions due to borrowing constraints, $\tau(b)$.

As will be shown later, how the "marginal value of wealth" conditional on schooling, $V_j'(b) \equiv dV_j(b)/db$, varies across schooling levels plays a central role in determining the optimal subsidy schedule. The next lemma establishes its properties.

Lemma 2. *For all $(j, b) \in \mathcal{J} \times \mathbb{R}_+$ such that $j \in \mathcal{J}(b)$: (i) $V_j'(b) > (=)1$ if and only if* (BC) *does (not) bind; and (ii) $V_j'(b)$ is strictly decreasing in b and strictly increasing in j if and only if* (BC) *binds.*

Consider a feasible choice $j \in \mathcal{J}(b)$. If (BC) does not bind (i.e., $b \geq k_j + y_j$) so that consumption

[20] Because of the way the schooling choice problem is formulated in (3), the choice probabilities are either zero or one. However, as will be shown Section 3.5, this approach becomes convenient for a more general model with additional unobservable heterogeneity that gives non-trivial choice probabilities conditional on b.

is fully smoothed across time, then $V_j(b) = y_j - k_j + b$ and $V_j'(b) = 1$. That is, an additional dollar of wealth raises welfare only by a dollar. On the other hand, when (BC) binds (i.e., $b < k_j + y_j$), then an additional dollar of wealth raises welfare by more than a dollar (i.e., $V_j'(b) > 1$), as it improves intertemporal consumption smoothing. Moreover, the marginal value of wealth is higher for lower initial wealth or higher schooling levels, as the intertemporal consumption allocation is more severely distorted.[21]

The next lemma shows that additional wealth also reduces the total distortion.

Lemma 3. *$\tau(b)$ is decreasing in b. Moreover, there exists $\bar{b} \in \mathbb{R}_+$ such that $\tau(b) = 0$ for all $b \geq \bar{b}$.*

Let $j^* \in \mathcal{J}$ be the schooling option with the lowest value of $k_j + y_j$ among those that give the highest lifetime earnings net of schooling cost, $y_j - k_j$, and define $\bar{b} \equiv k_{j^*} + y_{j^*}$. For individuals choosing the schooling option j^*, their schooling investment is not distorted. Moreover, their intertemporal consumption allocation is not distorted either if their parental transfer levels are at least $\bar{b}$. On the other hand, for those with parental transfers lower than $\bar{b}$, the borrowing constraint would bind had they chose the schooling option j^*, which may induce them to choose other schooling levels. Thus, their schooling and/or consumption must be distorted.

The fact that individuals with sufficiently high parental transfers attain undistorted allocations implies that fully equalizing parental transfers across individuals can eliminate all distortions if the average amount of parental transfer is large enough. This provides a case for redistribution, even when there is no explicit social preferences for equity.[22] However, such redistribution must be implemented indirectly through education subsidies because the social planner cannot observe parental transfers. Next, I characterize subsidy policies that minimize aggregate distortions.

3.3 Optimal Subsidy Conditional on Family Resources

I first characterize optimal subsidy for individuals with identical family resources. Consider otherwise identical individuals who are endowed with different amounts of parental transfers. Let $\mathcal{I}$ be the finite set of individual "types." Individuals of type $i \in \mathcal{I}$ are endowed with a parental transfer $b_i \in \mathbb{R}_+$. Let $f_i \in (0, 1]$ be the fraction of type i individuals. Although the type distribution is public information, individual type is private information. Unless stated otherwise, consider the non-trivial case where there are more than two distinct individual types. That is, $|\mathcal{I}| \geq 2$ and $b_i \neq b_{i'}$ for all $(i, i') \in \mathcal{I} \times \mathcal{I}$ such that $i \neq i'$.

The social planner minimizes aggregate distortions by distributing an exogenously given budget $G \in \mathbb{R}_+$ through education subsidy. Let $g_j \in \mathbb{R}$ be the amount of subsidy for schooling j and

[21] Appendix B.5 shows that the marginal value of wealth is positively associated with a *relative* measure of intertemporal consumption distortion—the maximum willingness to pay as a fraction of lifetime consumption.

[22] In general, however, full redistribution is not necessarily optimal because discrete choice induces local non-concavities in the indirect utility functions $V(\cdot)$ and $U(\cdot)$, even though $V_j(\cdot)$ and $U_j(\cdot)$ are concave (e.g., Kwang, 1965; Vereshchagina and Hopenhayn, 2009). Appendix B.6 provides a necessary and sufficient condition for the optimality of full redistribution. In practice, the concavity of the indirect utility function can be achieved by integrating out additional heterogeneity over choices (Gomes, Greenwood, and Rebelo, 2001). Appendix B.16 shows that introducing a sufficient degree of heterogeneity in returns to schooling makes the indirect utility function concave and thus makes fully equalizing parental transfers optimal.

$\boldsymbol{g} = (g_j)_{j\in\mathcal{J}} \in \mathbb{R}^{|\mathcal{J}|}$ the "subsidy schedule." Only in this subsection, I assume that the social planner can also alter the choice set of schooling itself by setting the post-schooling earnings schedule $\boldsymbol{y} = (y_j)_{j\in\mathcal{J}} \in \Lambda$, where $\Lambda \equiv \{\boldsymbol{y} \in \mathbb{R}_+^{|\mathcal{J}|} \,|\, y_j < y_{j'} \text{ for all } (j, j') \in \mathcal{J}\times\mathcal{J} \text{ such that } j < j'\}$.[23] Notice that changing y_j also changes k_j through the relationship $k_j = k(y_j)$. Therefore, choosing the level of earnings associated with a particular schooling level can be thought of as setting productive quality of education for each level of schooling.

Since both $\boldsymbol{g}$ and $\boldsymbol{y}$ affect the monetary return to schooling, they must be appropriately accounted for when defining the set of feasible choices, distortions, and indirect utilities. With updated definitions provided in Appendix B.7, the planning problem can be written as follows:

Problem 1.

$$\min_{\boldsymbol{g},\boldsymbol{y},(\boldsymbol{p}(b_i))_{i\in\mathcal{I}}} \sum_{i\in\mathcal{I}} f_i \tau(b_i; \boldsymbol{g}, \boldsymbol{y}) \tag{6}$$

$$\textit{subject to} \sum_{i\in\mathcal{I}} f_i \sum_{j\in\mathcal{J}} p_j(b_i) g_j \leq G, \tag{7}$$

$$\boldsymbol{p}(b_i) \in \operatorname*{argmax}_{\boldsymbol{p}\in\Pi} \sum_{j\in\mathcal{J}(b_i;\boldsymbol{g},\boldsymbol{y})} p_j V_j(b_i + g_j; y_j), \quad \forall i \in \mathcal{I}, \tag{8}$$

$$(\boldsymbol{g}, \boldsymbol{y}) \in \mathbb{R}^{|\mathcal{J}|} \times \Lambda. \tag{9}$$

In addition to $\boldsymbol{g}$ and $\boldsymbol{y}$, the social planner also chooses a lottery $\boldsymbol{p}(b_i)$ over schooling choices for each i. The constraint (8) states that the lottery must be consistent with individual utility maximization. When individuals are indifferent between multiple schooling options, their utility-maximizing schooling choice is not uniquely determined. However, those schooling options might be associated with different levels of aggregate distortions or aggregate public spending. Therefore, allowing for the social planner to dictate individual schooling choice ensures that the socially optimal outcome is achieved.

From the definition of $\tau(b; \boldsymbol{g}, \boldsymbol{y})$, we have the following identity:

$$\sum_{i\in\mathcal{I}} f_i V(b_i; \boldsymbol{g}, \boldsymbol{y}) = \max_{y\in\mathbb{R}_+} \{y - k(y)\} + \sum_{i\in\mathcal{I}} f_i b_i + \sum_{i\in\mathcal{I}} f_i \sum_{j\in\mathcal{J}} p_j(b_i) g_j - \sum_{i\in\mathcal{I}} f_i \tau(b_i; \boldsymbol{g}, \boldsymbol{y}), \tag{10}$$

where $V(b_i; \boldsymbol{g}, \boldsymbol{y}) \equiv \sum_{j\in\mathcal{J}(b_i;\boldsymbol{g},\boldsymbol{y})} p_j(b_i) V_j(b_i + g_j; y_j)$. Therefore, if the social planner's budget constraint (7) binds, minimizing aggregate distortions is equivalent to maximizing the monetary measure of social welfare, $\sum_{i\in\mathcal{I}} f_i V(b_i; \boldsymbol{g}, \boldsymbol{y})$.[24]

The following set of assumptions facilitates analytical characterization of the planning problem.

Assumption 1. *(i)* $|\mathcal{J}| \geq |\mathcal{I}|$; *and (ii)* $k(\cdot)$ *is continuously differentiable and strictly convex, satisfying* $k(0) = k'(0) = 0$.

[23] Assuming that $\boldsymbol{y}$ is exogenously given would impose restrictions on the social planner's choice set, which is not common in the optimal taxation literature based on Mirrlees (1971).

[24] The social planner's budget constraint does not bind when there are enough resources to fully eliminate aggregate distortions.

Part (i) of the assumption states that the number of schooling options is large enough for a separating equilibrium—where different types choose different options—to exist. Part (ii) is a standard convex cost assumption ensuring that identical types do not make different schooling choices.

Proposition 1. *Suppose that Assumption 1 holds and $(\hat{\boldsymbol{g}}, \hat{\boldsymbol{y}}, (\hat{\boldsymbol{p}}(b_i))_{i \in \mathcal{I}})$ solves Problem 1. Let $\hat{\mathcal{J}} \equiv \{j \in \mathcal{J} | \hat{p}_j(b_i) > 0$ for some $i \in \mathcal{I}\}$. Then (i) $\hat{g}_j > \hat{g}_{j'}$ for all $(j, j') \in \hat{\mathcal{J}} \times \hat{\mathcal{J}}$ such that $j < j'$; (ii) $k(\hat{y}_j) - \hat{g}_j < k(\hat{y}_{j'}) - \hat{g}_{j'}$ for all $(j, j') \in \hat{\mathcal{J}} \times \hat{\mathcal{J}}$ such that $j < j'$; and (iii) $\hat{p}_j(b_i) \in \{0, 1\}$ for all $(i, j) \in \mathcal{I} \times \hat{\mathcal{J}}$.*

We can restrict our attention to the schooling options that are chosen by some individuals (i.e., $j \in \hat{\mathcal{J}}$). Part (i) of the proposition states that the subsidy amount is strictly decreasing in schooling level. Since the net cost of schooling, $k(\hat{y}_j) - \hat{g}_j$, is strictly increasing in schooling levels (part (ii)), by Lemma 1, higher types choose higher schooling levels. Moreover, all identical types make identical schooling choices (part (iii)). Therefore, the optimal policy effectively gives larger subsidies to those with lower parental transfers.

The optimality of a declining subsidy schedule reflects the desire to redistribute toward those with lower parental transfers: as shown by Lemma 3, their allocation is more distorted by borrowing constraints, and an additional dollar given to them reduces the distortions. Notice that this is not driven by social concerns about equity. If the social planner were inequality averse, providing larger subsidies for lower schooling levels would generate an additional social benefit by reducing inequality in lifetime consumption, strengthening the case for a declining subsidy schedule.

The declining subsidy schedule does not depend on the nature of the relationship between schooling and earnings either: although the convexity of $k(\cdot)$ is assumed for technical reasons, how earnings net of schooling cost, $y_j - k_j$, vary across schooling levels is endogenously determined by the social planner.

3.4 Variation of Optimal Subsidy with Family Resources

Next, I characterize how optimal subsidy varies across family resources for each level of schooling when there exists inequality in parental transfers conditional on family resources. Consider individuals who differ not only by parental transfers but also by observable family resources, or "parental wealth," which only affects the distribution of parental transfers. Parental wealth is positively correlated with parental transfers in the sense that those with high parental wealth are more likely to receive high parental transfers. However, the correlation is imperfect, so individuals with identical parental wealth may receive different parental transfers.

Specifically, suppose that $\mathcal{I}$ is partitioned into a finite number of parental wealth groups indexed by $h \in \mathcal{H}$ such that $\mathcal{I} = \cup_{h \in \mathcal{H}} \mathcal{I}_h$. Let $f_{i|h} \equiv f_i \mathbb{I}_{i \in \mathcal{I}_h} / \sum_{i' \in \mathcal{I}_h} f_{i'}$ be the fraction of type i in group h and $\Phi(b|h) \equiv \sum_{i \in \mathcal{I}_h} f_{i|h} \mathbb{I}_{b_i \le b}$ be the cumulative distribution function of parental transfers conditional on group h. The index for parental wealth group is ordered such that the distribution of parental transfers for high h first-order stochastically dominates that for low h. Moreover, there is sufficient within-group inequality in parental transfers in the sense that the support of the parental transfer distribution for one group can overlap with that of all other groups.

Assumption 2. *For all* $(h, h') \in \mathcal{H} \times \mathcal{H}$*: (i)* $\Phi(b|h) \geq \Phi(b|h')$ *for all* $b \in \mathbb{R}$ *and* $h' > h$*; and (ii)* $\min_{i \in \mathcal{I}_h}\{b_i\} < \max_{i \in \mathcal{I}_{h'}}\{b_i\}$.

The social planner can only observe what group each individual belongs to and sets subsidy schedules separately for each group. The budget for each group $(G_h)_{h \in \mathcal{H}}$ is exogenously given. Let $g_{j,h}$ be the subsidy amount for $j \in \mathcal{J}$ and $h \in \mathcal{H}$, and let $\boldsymbol{g}_h \equiv (g_{j,h})_{j \in \mathcal{J}}$ be the subsidy schedule for $h \in \mathcal{H}$. Since I am interested in how optimal $g_{j,h}$ varies with h for a given y_j, the earnings schedule $\boldsymbol{y}$ is assumed to be exogenously given from now on and held constant across groups. The planning problem can be written as follows:

Problem 2.

$$
\begin{aligned}
\min_{(\boldsymbol{g}_h, \boldsymbol{p}(b_i;\boldsymbol{g}_h)_{i \in \mathcal{I}_h})_{h \in \mathcal{H}}} \quad & \sum_{h \in \mathcal{H}} \sum_{i \in \mathcal{I}_h} f_i \tau(b_i; \boldsymbol{g}_h) \\
\textit{subject to} \quad & \sum_{i \in \mathcal{I}_h} f_{i|h} \sum_{j \in \mathcal{J}} p_j(b_i; \boldsymbol{g}_h) g_{j,h} \leq G_h, \quad \forall h \in \mathcal{H}, \\
& \boldsymbol{p}(b_i; \boldsymbol{g}_h) \in \operatorname*{argmax}_{\boldsymbol{p} \in \Pi} \sum_{j \in \mathcal{J}(b_i; \boldsymbol{g}_h)} p_j V_j(b_i + g_{j,h}), \quad \forall i \in \mathcal{I}_h, \forall h \in \mathcal{H}, \\
& \boldsymbol{g}_h \in \mathbb{R}^{|\mathcal{J}|}, \quad \forall h \in \mathcal{H}.
\end{aligned}
$$

Problem 2 generalizes Problem 1 to the case with multiple observable groups (except that $\boldsymbol{y}$ is now taken as given). Since the budget is set for each group and is not allowed to be reallocated across groups, the planning problem can be solved separately by each group. However, it is still difficult to characterize how the optimal subsidy varies across parental wealth groups for the general case. Thus I consider the simple case where $\mathcal{J} = \{1, 2\}$ and $\mathcal{I}_h = \{(L, h), (H, h)\}$ for all $h \in \mathcal{H}$. That is, there are two schooling options—low ($j = 1$) and high ($j = 2$) schooling—and two types of individuals—low ($i = (L, h)$) and high ($i = (H, h)$) parental transfer types—in each group. Then Assumption 2 implies that $b_{L,h}$, $b_{H,h}$, and $f_{H,h|h}$ are increasing in h.[25] Furthermore, I assume a constant budget for each group and normalize its value to zero.

Assumption 3. $G_h = 0$ *for all* $h \in \mathcal{H}$.

I make additional assumptions on the costs of and returns to schooling and their relationship with parental transfer levels.

Assumption 4. *(i)* $k_2 - k_1 < (y_2 - y_1)y_1/y_2$*; (ii)* $\sum_{i \in \mathcal{I}_h} f_{i|h} b_i < (y_2 k_2 - y_1 k_1)/(y_2 - y_1) < b_{H,h}$ *for all* $h \in \mathcal{H}$*; and (iii)* $b_{H,h} - (y_2/y_1) b_{L,h} \geq (y_2 - y_1)(y_2 - k_1)/y_1$ *for all* $h \in \mathcal{H}$.

Part (i) implies $y_1 - k_1 < y_2 - k_2$, meaning the higher level of schooling is more profitable. Therefore, in the absence of subsidy (i.e., $\boldsymbol{g} = \boldsymbol{0}$) and borrowing constraints, everyone chooses the higher level. Part (ii) implies that with zero subsidy, low types are borrowing constrained and choose the low schooling, while high types choose the high schooling. This ensures that the optimal schooling

[25] See Appendix B.13 for proof.

allocation is fully separating, where different types do not make identical schooling choices.[26] Finally, part (iii) imposes that the difference in parental transfers between low and high types is so large that the social planner redistributes as much as she can in each parental wealth group.[27]

Proposition 2. *Suppose that Assumptions 2, 3, and 4 hold and* $(\hat{\boldsymbol{g}}_h, (\hat{\boldsymbol{p}}(b_i))_{i\in\mathcal{I}_h})_{h\in\mathcal{H}}$ *solves Problem 2. Then: (i)* $\hat{g}_{1,h} > \hat{g}_{2,h}$ *and* $\hat{p}_1(b_{L,h}) = \hat{p}_2(b_{H,h}) = 1$ *for all* $h \in \mathcal{H}$*; and (ii)* $\hat{g}_{1,h'} \geq \hat{g}_{1,h}$ *for all* $(h', h) \in \mathcal{H} \times \mathcal{H}$ *such that* $h' > h$.

As already shown in Proposition 1, part (i) of Proposition 2 demonstrates that the optimal policy redistributes from high to low types by assigning larger subsidies to the lower schooling level. Part (ii) states that such redistribution is stronger for rich families with high parental wealth. As implied by Assumption 2, there exists a larger fraction of high parental transfer types among rich families, and the amount high types receive is also larger for rich families than for poor families. Since low types receive higher subsidies by choosing the low schooling, the presence of fewer of them among rich families makes such a policy less costly, thus raising the amount of subsidy each low type individual from rich families can get. Moreover, because high types from rich families receive higher parental transfers than those from poor families, they are less likely to be borrowing constrained, and the social planner is able to extract more from them without distorting their allocations too much.

The result that the amount of optimal subsidy for the low schooling level increases with family resources holds under the assumption that the budget stays constant across groups; it can be offset when the budget for rich families is lower than that for poor families. When the budget can be reallocated across groups, spending more on poor families than on rich families might reduce aggregate distortions because the former are more likely to be borrowing constrained.[28] Section 3.5.1 provides a principle for the optimal allocation of budget across groups, and Section 4.3.2 quantitatively explores the joint determination of subsidy schedules and budgets for each group.

3.5 Incorporating Heterogeneous Returns to Schooling

Thus far, it has been assumed that all differences in educational attainment are driven by unequal parental transfers along with borrowing constraints. I close this section by incorporating unobservable heterogeneity in returns to schooling, another potentially important determinant of educational attainment. The analysis in this subsection highlights that the nature of heterogeneity leading to differences in schooling choice matters for the design of education subsidy, and it also serves as a building block for the quantitative model.

Individuals with an identical parental transfer may choose different schooling levels because of differences in wage returns to schooling (i.e., ability) or because of differences in other non-

[26]For a pooling schooling allocation with $\hat{p}_j(b_i) = 1$ for all $i \in \mathcal{I}_h$, we have $\hat{g}_{j,h} = 0$. That is, no within-group redistribution is feasible since low and high types choose the same level of schooling.

[27]As a result of this assumption, high types are made indifferent between the two schooling options. This assumption would not be needed if $\boldsymbol{y}$ for each $h \in \mathcal{H}$ were optimally chosen by the social planner.

[28]See Colas, Findeisen, and Sachs (2018) for other mechanisms inducing a positive relationship between optimal college subsidy and family resources.

pecuniary, or "psychic," returns to schooling.[29] As discussed below, these two forms of unobservable heterogeneity in returns to schooling have similar policy implications. This subsection incorporates unobservable heterogeneity in psychic returns because it gives an intuitive formula for how the optimal subsidy schedule is determined. Appendix B.14 provides theoretical analysis on unobservable heterogeneity in ability. The quantitative model in Section 4 incorporates heterogeneity in ability as well as psychic returns to schooling.

Let ε_j be the psychic return for schooling j and let $\boldsymbol{\varepsilon} = (\varepsilon_j)_{j \in \mathcal{J}} \in \mathbb{R}^{|\mathcal{J}|}$ be the vector of psychic returns. Then, the schooling choice problem for those with a parental transfer b and psychic returns $\boldsymbol{\varepsilon}$ is

$$\max_{j \in \mathcal{J}(b;\boldsymbol{g})} \left\{ V_j(b + g_j) + \varepsilon_j \right\},$$

where $\mathcal{J}(b;\boldsymbol{g}) \equiv \{j \in \mathcal{J} | k_j - g_j \leq b\}$. Notice that, when (BC) does not bind so that $V_j(b + g_j) = y_j - k_j + b + g_j$ for all $j \in \mathcal{J}$, parental transfers do not affect schooling decisions. Therefore, parental transfers affect schooling choice only though borrowing constraints. The lack of direct wealth effects is a desirable property because the wealth effects would generate a counterfactual negative relationship between educational attainment and parental resources for empirically plausible values of psychic returns (Caucutt, Lochner, and Park, 2017).

The distortions can be defined as before, but it is convenient to take averages over $\boldsymbol{\varepsilon}$. Let $F(\cdot)$ be the continuous cumulative distribution function of $\boldsymbol{\varepsilon}$ and define $p_j(b;\boldsymbol{g})$ as the fraction of individuals choosing $j \in \mathcal{J}$, conditional on $(b, \boldsymbol{g})$:

$$p_j(b;\boldsymbol{g}) \equiv \int \mathbb{I}_{j \in \text{argmax}_{j' \in \mathcal{J}(b;\boldsymbol{g})} \{V_{j'}(b+g_{j'})+\varepsilon_{j'}\}} dF(\boldsymbol{\varepsilon}), \tag{11}$$

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

$$\tau_c(b;\boldsymbol{g}) \equiv \sum_{j \in \mathcal{J}(b;\boldsymbol{g})} p_j(b;\boldsymbol{g}) \left[(y_j - k_j + b + g_j) - V_j(b + g_j) \right],$$

and the average schooling distortion is

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

which also takes into account lost psychic returns due to distorted schooling choices.

Finally, by adding the two distortions, the average total distortion conditional on $(b, \boldsymbol{g})$, $\tau(b;\boldsymbol{g}) \equiv$

[29] Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005) find substantial non-pecuniary 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).

$\tau_c(b;\boldsymbol{g}) + \tau_s(b;\boldsymbol{g})$, is given by

$$\tau(b;\boldsymbol{g}) = \int \max_{j\in\mathcal{J}} \left\{y_j - k_j + \varepsilon_j\right\} dF(\boldsymbol{\varepsilon}) + b + \sum_{j\in\mathcal{J}} p_j(b;\boldsymbol{g})g_j - V(b;\boldsymbol{g}), \tag{12}$$

where $V(b;\boldsymbol{g}) \equiv \int \max_{j\in\mathcal{J}(b;\boldsymbol{g})}\{V_j(b+g_j)+\varepsilon_j\}dF(\boldsymbol{\varepsilon})$.

The social planner, who cannot observe individual values of $\boldsymbol{\varepsilon}$, solves the following problem:

Problem 3.

$$\begin{aligned} \min_{(\boldsymbol{g}_h)_{h\in\mathcal{H}}} & \sum_{h\in\mathcal{H}} \sum_{i\in\mathcal{I}_h} f_i \tau(b_i;\boldsymbol{g}_h) \\ \textit{subject to} & \sum_{h\in\mathcal{H}} \sum_{i\in\mathcal{I}_h} f_i \sum_{j\in\mathcal{J}} p_j(b_i;\boldsymbol{g}_h) g_{j,h} \le G, \\ & \boldsymbol{g}_h \in \mathbb{R}^{|\mathcal{J}|}, \quad \forall h\in\mathcal{H}, \end{aligned} \tag{13}$$

where $p_j(b;\boldsymbol{g})$ and $\tau(b;\boldsymbol{g})$ are given by (11) and (12).

Notice that separate budget contraints for each group imposed in Problem 2 are now replaced by a single budget constraint (13). For analytical tractability, I assume that the psychic returns 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), $p_j(b;\boldsymbol{g}) = \exp\left(V_j(b+g_j)\right) / \sum_{j'\in\mathcal{J}(b;\boldsymbol{g})}\exp\left(V_{j'}(b+g_{j'})\right)$ for all $j\in\mathcal{J}(b;\boldsymbol{g})$ and $V(b;\boldsymbol{g}) = \ln(\sum_{j\in\mathcal{J}(b;\boldsymbol{g})}\exp(V_j(b+g_j)))+\rho$ hold, where $\rho \approx 0.57721$ is the Euler-Mascheroni constant.[30] These formulas are also useful for quantitative analysis because they make numerical integration over $\boldsymbol{\varepsilon}$ unnecessary.

3.5.1 Equal Parental Transfers Within Group

I begin with the simple case where everyone in the same parental wealth group receives the same amount of parental transfers. Then individuals in the same group are heterogeneous only in the psychic returns to schooling.

Assumption 5. *$b_i = b_{i'}$ for all $(i,i')\in\mathcal{I}_h\times\mathcal{I}_h$ and for all $h\in\mathcal{H}$.*

When Assumption 5 holds, let b_h be the amount of parental transfer for individuals in group h, and let $f_h \equiv \sum_{i\in\mathcal{I}_h} f_i$ denote the fraction of group h individuals.

Proposition 3. *Suppose that Assumption 5 holds and $(\hat{\boldsymbol{g}}_h)_{h\in\mathcal{H}}$ solves Problem 3. Then, for all $h\in\mathcal{H}$: (i) $\hat{g}_{j,h} \le \hat{g}_{j',h}$ for all $(j,j')\in\mathcal{J}(b_h;\hat{\boldsymbol{g}}_h)\times\mathcal{J}(b_h;\hat{\boldsymbol{g}}_h)$ such that $j<j'$; and (ii) for all $j\in\mathcal{J}(b_h;\hat{\boldsymbol{g}}_h)$,*

$$\hat{g}_{j,h} = \sum_{j'\in\mathcal{J}} p_{j'}(b_h;\hat{\boldsymbol{g}}_h)\hat{g}_{j',h} + \frac{1}{1+\lambda} - \frac{1}{V_j'(b_h+\hat{g}_{j,h})}, \tag{14}$$

where λ is the Lagrangian multiplier on the constraint (13).

[30] The Euler-Mascheroni constant is also the mean of the standard Gumbel distribution.

Formula (14) shows that the optimal subsidy for schooling j and group h is determined by the level of public spending for group h; the social marginal value of public funds, as captured by the Lagrangian multiplier on the social planner's budget constraint (13); and the marginal value of wealth conditional on (j, h).

Within-Group Allocation Taking a difference of (14) between two schooling levels gives

$$\hat{g}_{j',h} - \hat{g}_{j,h} = \frac{1}{V_j'(b_h + \hat{g}_{j,h})} - \frac{1}{V_{j'}'(b_h + \hat{g}_{j',h})}, \tag{15}$$

which shows that the optimal subsidy is larger for schooling levels with higher marginal value of wealth, where an additional subsidy has greater impact. As shown by Lemma 2, the marginal value of wealth is larger for higher schooling levels, which makes it optimal to provide an increasing subsidy schedule.

For a given level of parental transfer, borrowing constraints are more likely to bind for those choosing higher schooling levels because they pay higher costs during schooling and earn more after schooling. Therefore, contrary to the case with unobservable heterogeneity in parental transfers, the allocations of those attaining higher schooling are more distorted. Thus, giving them larger subsidies reduces aggregate distortions. Notice that the optimal policy might increase inequality in lifetime consumption by giving higher subsidies to more educated individuals with higher lifetime earnings. Therefore, the absence of strong social preferences for equity is crucial for the increasing optimal subsidy schedule.

As shown in Appendix B.14, this result carries over to the case with unobservable heterogeneity in wage returns to schooling, provided that more able individuals attain higher schooling due to complementarity between ability and schooling in generating earnings: an increasing subsidy schedule is optimal because it helps higher ability individuals who are more likely to be borrowing constrained.[31]

Between-Group Allocation Formula (14) also gives the optimality condition for allocation of public spending across parental wealth groups: for all $h \in \mathcal{H}$,

$$\frac{1}{1+\lambda} = \sum_{j \in \mathcal{J}(b_h; \hat{\boldsymbol{g}}_h)} \frac{p_j(b_h; \hat{\boldsymbol{g}}_h)}{V_j'(b_h + \hat{g}_{j,h})}. \tag{16}$$

The term $1/V_j'(b_h + g_{j,h})$ is the cost of a marginal increase in the money-metric utility of individuals in group h choosing schooling j. Therefore, Condition (16) states that the average cost of increasing monetary measure of welfare, $V(b_h; \boldsymbol{g}_h)$, is equalized across groups.[32] If the average cost were not equal across groups, then the social planner could raise $\sum_{h \in \mathcal{H}} f_h V(b_h; \boldsymbol{g}_h)$ by transferring resources

[31] In general, however, borrowing-constrained individuals with higher ability may attain lower schooling as a result of a strong intertemporal consumption smoothing motive (Lochner and Monge-Naranjo, 2011). The potential non-monotonicity between ability and schooling makes theoretical analysis on the optimal policy design challenging (Gottlieb and Moreira, 2012), but it helps identify individual preferences for intertemporal consumption smoothing based on the empirical relationship between measured ability and educational attainment, as discussed in Section 4.2.3.

[32] A similar condition is derived in the literature of optimal income taxation with "tagging" (e.g., Mankiw and Weinzierl, 2010).

to the group for which this cost was relatively low. As noted earlier, this also means that such a transfer between groups reduces aggregate distortions when the planner's budget constraint binds (i.e., $\lambda > 0$).

The equalization of average cost across groups generally involves giving higher subsidies to lower parental wealth groups because $1/V_j'(b)$ is increasing in b (Lemma 2). In particular, the optimal subsidy might fully compensate for the differences in parental transfers across groups, satisfying the following condition: for all $j \in \mathcal{J}(b_h; \boldsymbol{g}_h) \cap \mathcal{J}(b_{h'}; \boldsymbol{g}_{h'})$ and $(h, h') \in \mathcal{H} \times \mathcal{H}$,

$$g_{j,h'} - g_{j,h} = b_h - b_{h'}. \tag{17}$$

It is easy to see that if $\boldsymbol{g}_h$ satisfies (14) for group h, then $\boldsymbol{g}_{h'}$ with $g_{j,h'} = g_{j,h} + (b_h - b_{h'})$ also satisfies (14) for group h'. Therefore, (17) is consistent with necessary conditions for optimality. Appendix B.16 establishes conditions under which (17) is satisfied at optimum.

The two properties of the optimal policy—larger subsidy amounts for those with higher schooling levels and lower family resources—are consistent with the features of current US financial aid policy. However, they are derived under the assumption that parental transfers are perfectly correlated with parental wealth; they are not necessarily optimal when the social planner cannot precisely predict parental transfers based on parental wealth. Although the presence of within-group inequality in parental transfers makes it difficult to sharply characterize the properties of the optimal policy, I next show how it changes the optimality condition (14).

3.5.2 Within-Group Inequality in Parental Transfers

When parental transfers also differ across individuals in the same parental wealth group, those with different parental transfers in the same group may choose the same schooling levels due to differences in psychic returns. In this case, we can derive a formula similar to (14) but aggregated across types within each schooling level and parental wealth group.

Corollary 1. *Suppose that $(\hat{\boldsymbol{g}}_h)_{h\in\mathcal{H}}$ solves Problem 3. If $\mathcal{J}(b_i; \hat{\boldsymbol{g}}_h) = \mathcal{J}$ for all $i \in \mathcal{I}_h$ and for all $h \in \mathcal{H}$, then*

$$\hat{g}_{j,h} = \sum_{i\in\mathcal{I}_h} \hat{q}_{i|j} \sum_{j'\in\mathcal{J}} p_{j'}(b_i; \hat{\boldsymbol{g}}_h)\hat{g}_{j',h} + \frac{1}{1+\lambda} - \frac{1}{\sum_{i\in\mathcal{I}_h} \hat{f}_{i|j} V_j'(b_i + \hat{g}_{j,h})}, \tag{18}$$

where $\hat{f}_{i|j} \equiv f_i p_j(b_i; \hat{\boldsymbol{g}}_h) / \sum_{i'\in\mathcal{I}_h} f_{i'} p_j(b_{i'}; \hat{\boldsymbol{g}}_h)$ and $\hat{q}_{i|j} \equiv f_i p_j(b_i; \hat{\boldsymbol{g}}_h) V_j'(b_i + \hat{g}_{j,h}) / \sum_{i'\in\mathcal{I}} f_{i'} p_j(b_{i'}; \hat{\boldsymbol{g}}_h) V_j'(b_{i'} + \hat{g}_{j,h})$.

Formula (18) shows how the relationship between the *average* marginal value of wealth and schooling affects the shape of the optimal subsidy schedule for a given group. The shape is ambiguous as a result of two opposing forces. For a given level of parental transfer, those choosing higher schooling levels are more likely to be borrowing constrained and have a higher marginal value of wealth. However, those with more schooling are also more likely to have received larger parental transfers, which would yield a lower marginal value of wealth. Therefore, the relative heterogeneity between returns to schooling vs. parental transfers is an important determinant of the optimal subsidy schedule. In the following section, I investigate these forces quantitatively, along with the optimal subsidy policy for the US.

4 Quantitative Analysis

I now explore the quantitative implications of unobservable heterogeneity in parental transfers and returns to schooling for the design of education subsidies using a more general framework. I first extend the model to a life-cycle setting with heterogeneity in ability and endogenous parental transfers. Then, I calibrate it to the relationship between schooling, ability, parental income, and parental transfers, as well as other features of the US economy. Finally, I solve for optimal policies numerically.

4.1 A More General Framework

Time is discrete and a model period represents a calendar year. A family consists of a parent and a child. In period $t = 0$, children have finished high school and are about to start post-secondary education. Children live until $t = T_k$ and parents live until $t = T_p$.

4.1.1 Intertemporal Consumption Allocation

Children, or "youth," accrue a flow of utility from consumption c_t in period t and discount future utility flows at a subjective discount factor $\beta \in (0, 1)$. The present discounted value of lifetime utility (as at period $t = 0$) from a consumption profile $(c_t)_{t=1}^{T_k}$ is

$$U = \sum_{t=1}^{T_k} \beta^t u(c_t), \tag{19}$$

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

Let $\mathcal{J} = \mathcal{J}_+ \cup \{0\}$, where $\mathcal{J}_+$ is a subset of strictly positive integers. Each schooling option $j \in \mathcal{J}$ represents the highest year of college education completed. The choice $j = 0$ corresponds to not attending college, in which case youth start working in $t = 1$. Those choosing $j \in \mathcal{J}_+$ attend college from $t = 1$ until $t = j$ and start working in $t = j + 1$.

Individuals can save and borrow at an annual gross interest rate $R = 1/\beta$. Let $y_j(a)$ be the present discounted value of lifetime earnings for those with schooling j and ability $a \in \mathcal{A}$, where $\mathcal{A}$ is a totally ordered set and $y_j : \mathcal{A} \to \mathbb{R}_+$ is a strictly increasing function. Then the lifetime budget constraint conditional on schooling j is

$$\sum_{t=1}^{T_k} R^{-t} c_t \leq y_j(a) - k_j + b_j + g_j, \tag{20}$$

where $k_j \in \mathbb{R}_+$, $b_j \in \mathbb{R}_+$, and $g_j \in \mathbb{R}_+$ are the present discounted values of college costs, parental transfers, and subsidies, respectively. Notice that parental transfers can be schooling-specific and subsidies are restricted to be positive. I also assume that there are no monetary costs and subsidies for those who do not attend college, that is, $k_0 = g_0 = 0$. Let $\Gamma \equiv \{\boldsymbol{g} \in \mathbb{R}_+^{|\mathcal{J}|} | g_0 = 0\}$ be the set of $\boldsymbol{g}$ satisfying these assumptions.

While those who are not enrolled in college can fully smooth consumption over time, individuals who are enrolled in college face borrowing constraints: for $j \in \mathcal{J}_+$,

$$\sum_{t=1}^{j} R^{-t} c_t + k_j - b_j - g_j \leq \overline{d}_j, \tag{21}$$

where $\overline{d}_j \in \mathbb{R}_+$ is the borrowing limit during college. The schooling-specific borrowing limits reflect the feature of the US government student loan program that offers higher annual as well as total loan limits to students in higher grades.

Assuming individuals cannot borrow more than they will earn, that is, $\overline{d}_j \leq y_j(a)$ for all $(a, j) \in \mathcal{A} \times \mathcal{J}_+$, the set of feasible schooling choices can be written as $\mathcal{J}(\boldsymbol{b}; \boldsymbol{g}) = \{0\} \cup \{j \in \mathcal{J}_+ | b_j \geq k_j - g_j - \overline{d}_j\}$, where $\boldsymbol{b} \equiv (b_j)_{j \in \mathcal{J}} \in \mathbb{R}_+^{|\mathcal{J}|}$ is the parental transfer schedule. For a feasible choice $j \in \mathcal{J}(\boldsymbol{b}; \boldsymbol{g})$, let $U_j(b_j + g_j; a)$ be its indirect utility, which is the maximized lifetime utility (19) subject to the lifetime budget constraint (20) and the borrowing constraints (21) for $j \in \mathcal{J}_+$. As before, let $V_j(b_j + g_j; a) \equiv e(U_j(b_j + g_j; a))$ be the money-metric indirect utility, where $e(\cdot)$ is the expenditure function defined similarly as (4):

$$e(U) \equiv \inf_{(\tilde{c}_t)_{t=1}^{T_k} \in \mathbb{R}_+^{T_k}} \left\{ \sum_{t=1}^{T_k} R^{-t} \tilde{c}_t \middle| \sum_{t=1}^{T_k} \beta^t u(\tilde{c}_t) \geq U \right\}.$$

Let $V_j'(z; a)$ be the derivative of $V_j(z; a)$ with respect to z. Then a version of Lemma 2 holds.

Lemma 4. *For all $j \in \mathcal{J}(\boldsymbol{b}; \boldsymbol{g})$, $V_j'(z; a) \geq 1$ holds. Moreover, for $j \in \mathcal{J}_+$, $V_j'(z; a) > 1$ and $V_j'(z; a)$ is increasing in a if and only if (21) binds.*

As discussed earlier, an additional dollar of initial wealth z raises welfare by more than a dollar for borrowing-constrained individuals because it improves intertemporal consumption smoothing. It has stronger effects for higher ability individuals who earn more because their desired level of consumption during schooling is higher. These are important for understanding parental transfer decisions in Section 4.1.3.

4.1.2 Education Choice

For a given subsidy schedule $\boldsymbol{g}$, the schooling choice problem for youth with schooling returns $(a, \boldsymbol{\varepsilon})$ and a parental transfer schedule $\boldsymbol{b}$ is

$$\max_{j \in \mathcal{J}(\boldsymbol{b}; \boldsymbol{g})} \left\{ V_j(b_j + g_j; a) + \mu_j + \sigma \varepsilon_j \right\},$$

where $\mu_j \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$ are the location and scale parameters of the distribution of psychic returns conditional on schooling j. I continue to assume that $\boldsymbol{\varepsilon}$ is identically and independently distributed with standard Gumbel distribution. I normalize $\mu_0 = 0$ so that μ_j for $j \in \mathcal{J}_+$ measures the common psychic returns of j relative to not attending college. The scale parameter σ determines the degree of heterogeneity in psychic returns. When $\sigma = 0$, psychic returns are homogeneous and thus differences in schooling choices are entirely driven by differences in other dimensions such as monetary returns to schooling. On the other hand, when $\sigma \to \infty$, the heterogeneity in psychic returns completely dominates, and all feasible options are equally likely to be chosen. As shown in the next subsection, the location and scale parameters are important in fitting the marginal distribution of schooling.

Conditional on $(\boldsymbol{b}, \boldsymbol{g}, a)$, the fraction of youth choosing an option $j \in \mathcal{J}$ is

$$p_j(\boldsymbol{b}; \boldsymbol{g}, a) \equiv \int \mathbb{I}_{j \in \operatorname{argmax}_{j' \in \mathcal{J}(\boldsymbol{b};\boldsymbol{g})}\{V_{j'}(b_j + g_j; a) + \mu_{j'} + \sigma\varepsilon_{j'}\}} dF(\boldsymbol{\varepsilon}), \tag{22}$$

and the average money-metric indirect utility is

$$V(\boldsymbol{b}; \boldsymbol{g}, a) \equiv \int \max_{j \in \mathcal{J}(\boldsymbol{b};\boldsymbol{g})} \left\{V_j(b_j + g_j; a) + \mu_j + \sigma\varepsilon_j\right\} dF(\boldsymbol{\varepsilon}). \tag{23}$$

4.1.3 Parental Transfer Decision

In $t = 0$, parents are endowed with wealth $w \in \mathbb{R}_+$—which represents the sum of initial net worth and present discounted value of future earnings—and make transfers to their child. Parents derive utility from their own consumption and also care about their child's welfare. That is, parents are altruistic toward their child. The degree of parental altruism is captured by a parameter $\delta \in [0, 1]$, which is heterogeneous across families. In order to simplify parents' problem, I assume that parents do not know about their child's psychic returns, $\boldsymbol{\varepsilon}$, when making a transfer decision, although they have the same information as their child otherwise. Not knowing their child's preferences for education, parents offer and commit to a schedule of transfers, $\boldsymbol{b}$.

For a given parental transfer schedule, a schooling option $j \in \mathcal{J}$ is chosen with a probability $p_j(\boldsymbol{b}; \boldsymbol{g}, a)$, which leads to an expected parental transfer, $\sum_{j \in \mathcal{J}} p_j(\boldsymbol{b}; \boldsymbol{g}, a) b_j$. Therefore, the expected annual consumption of the parent, c_p, satisfies

$$\sum_{t=1}^{T_p} R^{-t} c_p = w - \sum_{j \in \mathcal{J}} p_j(\boldsymbol{b}; \boldsymbol{g}, a) b_j. \tag{24}$$

Similarly, the expected money-metric indirect utility of the child, $V(\boldsymbol{b}; \boldsymbol{g}, a)$, can be expressed in terms of annual consumption, c_k, that satisfies

$$\sum_{t=1}^{T_k} R^{-t} c_k = V(\boldsymbol{b}; \boldsymbol{g}, a). \tag{25}$$

Although parents face uncertainty about their child's preferences for schooling, I assume that they are risk neutral so that they only care about expected values, c_p and c_k. The parents' problem follows.

Problem 4. *Consider a family with $(a, \delta, w, \boldsymbol{g})$. Taking $(p_j(\boldsymbol{b}; \boldsymbol{g}, a))_{j \in \mathcal{J}}$ and $V(\boldsymbol{b}; \boldsymbol{g}, a)$ as given, the* *parent solves*

$$\max_{\boldsymbol{b}} \left\{ (1-\delta) \sum_{t=1}^{T_p} \beta^t v(c_p) + \delta \sum_{t=1}^{T_k} \beta^t v(c_k) \right\}$$

$$\text{subject to (24), (25), and } b_j \in [0, w] \quad \forall j \in \mathcal{J},$$

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

The parameter $\eta \in \mathbb{R}_+$ is the elasticity of intergenerational substitution (EGS) that reflects the desire to smooth consumption across generations (Córdoba and Ripoll, forthcoming), which could differ from the EIS (γ). The distinction between the attitudes toward intertemporal and intergenerational consumption smoothing makes the parents' behavior formulated in Problem 4 more general than the behavior typically considered (e.g., Becker and Tomes, 1986). In particular, when $\gamma = \eta$, $\mu_j = 0$ for all $j \in \mathcal{J}$, and $\sigma = 0$, the parents' objective function is the weighted average of each family member's utility from own consumption:

$$(1-\delta) \sum_{t=1}^{T_p} \beta^t u(c_{t,p}) + \delta \sum_{t=1}^{T_k} \beta^t u(c_{t,k}), \tag{26}$$

where $c_{t,p}$ and $c_{t,k}$ are actual consumption of the parent and the child in each year.

Allowing for the EGS to differ from the EIS enables the model to better match the empirical relationship between educational attainment, ability, parental transfers, and parental income. The roles of the two parameters in determining parental transfers are emphasized in the following proposition.

Proposition 4. *For a family with* $(w, \delta, a, \boldsymbol{g})$, *suppose that* $\hat{\boldsymbol{b}}$ *solves Problem 4. Then, for* $j \in \mathcal{J}(\hat{\boldsymbol{b}}; \boldsymbol{g})$ *with* $\hat{b}_j \in (0, w)$,

$$\hat{b}_j = \sum_{j' \in \mathcal{J}} p_{j'}(\hat{\boldsymbol{b}}; \boldsymbol{g}, a) \hat{b}_{j'} + \sigma \left\{ \left(\frac{\delta}{1-\delta} \right) \frac{v'(\hat{c}_k)}{v'(\hat{c}_p)} - \frac{1}{V_j'(\hat{b}_j + g_j; a)} \right\}, \tag{27}$$

where $\hat{c}_p$ *and* $\hat{c}_k$ *are given by (24) and (25), evaluated at* $\boldsymbol{b} = \hat{\boldsymbol{b}}$.

Condition (27) is similar to the condition for the optimal subsidy (14), but the term for intergenerational smoothing, $\delta v'(c_k)/[(1-\delta)v'(c_p)]$, replaces $1/(1+\lambda)$. When $\hat{b}_j \in (0, w)$ for all $j \in \mathcal{J}(\hat{\boldsymbol{b}}; \boldsymbol{g})$, rearranging (27) gives

$$(1-\delta) v'(\hat{c}_p) = \delta v'(\hat{c}_k) \left(\sum_{j \in \mathcal{J}(\hat{\boldsymbol{b}}; \boldsymbol{g})} \frac{p_j(\hat{\boldsymbol{b}}; \boldsymbol{g}, a)}{V_j'(\hat{b}_j + g_j; a)} \right)^{-1}. \tag{28}$$

When all borrowing constraints (21) are slack, $V_j'(b_j + g_j; a) = 1$ for all $j \in \mathcal{J}$ and (28) becomes $(1-\delta)v'(\hat{c}_p) = \delta v'(\hat{c}_k)$. As discussed by Becker and Tomes (1986) and Brown, Scholz, and Seshadri (2012), this condition implies that wealthier or more altruistic parents give larger transfers. Youth with higher ability get lower parental transfers because they earn more and are better off. The gradients of parental transfers with respect to these factors depend on the EGS. However, the EIS does not play a role because it does not affect $\hat{c}_k$ when consumption is fully smoothed over time.

On the other hand, when the borrowing constraint for some $j \in \mathcal{J}_+$ binds, (28) implies $(1-\delta)v'(\hat{c}_p) > \delta v'(\hat{c}_k)$. Now the EIS affects parental transfers because it influences the magnitude of intertemporal consumption distortion: a given consumption fluctuation is more costly with strong preferences for intertemporal consumption smoothing or a small EIS. Unlike before, higher ability youth may receive higher parental transfers: although they are better off (i.e., high $\hat{c}_k$), their consumption is more likely to be intertemporally distorted by borrowing constraints (i.e., low $\sum_{j\in\mathcal{J}(\hat{\boldsymbol{b}};\boldsymbol{g})} p_j(\hat{\boldsymbol{b}};\boldsymbol{g},a)/V_j'(\hat{b}_j+g_j;a)$) because they borrow more for a given level of schooling and also are more likely to choose higher schooling levels. Therefore, when the EIS is small relative to the EGS, the latter effect dominates and parental transfers increase with ability.

Taking a difference of (27) between two schooling levels gives a formula similar to (15):

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

It shows that both borrowing constraints and incomplete information about children's preferences for schooling are necessary for parents to give different amounts of transfers across schooling levels. If (21) does not bind for either of the two choices (i.e., $V_j'(\hat{b}_j+g_j;a) = V_{j'}'(\hat{b}_{j'}+g_{j'};a) = 1$) or there is no incomplete information about psychic returns (i.e., $\sigma = 0$), then $\hat{b}_{j'} = \hat{b}_j$ holds, reflecting the fact that there is no need for non-paternalistic parents to manipulate children's schooling choices by conditioning transfer amounts on schooling levels.[33] With incomplete information (i.e., $\sigma > 0$), $\hat{b}_{j'} \geq \hat{b}_j$ holds if and only if $V_j'(\hat{b}_{j'}+g_{j'};a) \geq V_j'(\hat{b}_j+g_j;a)$. That is, parents, not knowing precisely which schooling option will be chosen, offer larger transfers for schooling levels for which children's consumption is more intertemporally distorted by borrowing constraints. Since youth are not subject to borrowing constraints when they do not attend college, identical parents will offer larger amounts for attending college than for not attending college.

This model provides an explanation for why parents would condition transfer amounts on schooling choices made by their children, instead of giving lump-sum transfers. Such "conditional parental transfer rule" is often taken as exogenously given (Keane and Wolpin, 2001; Johnson, 2013) or motivated by paternalistic preferences of parents who directly value their children's education (Abbott et al., forthcoming; Colas, Findeisen, and Sachs, 2018). However, as shown above, paternalism is not necessary to justify the conditional parental transfer rule.

4.1.4 Planning Problem

I continue to assume that the social objective is to minimize aggregate distortions. Distortions are defined similarly as in Section 3.5. Conditional on $(\boldsymbol{b},\boldsymbol{g},a)$, the average total distortion is (see

[33] For this reason, parents typically make lump-sum, or unconditional, transfers in dynastic models with complete information (e.g., Becker and Tomes, 1986; Krueger and Ludwig, 2016; Caucutt and Lochner, forthcoming).

Appendix B.20 for the definitions of consumption and schooling distortions)

$$\tau(\boldsymbol{b};\boldsymbol{g},a) = \int \max_{j\in\mathcal{J}} \left\{ y_j(a) - k_j + \mu_j + \sigma\varepsilon_j \right\} dF(\boldsymbol{\varepsilon}) + \sum_{j\in\mathcal{J}} p_j(\boldsymbol{b};\boldsymbol{g},a)(b_j + g_j) - V(\boldsymbol{b};\boldsymbol{g},a). \quad (29)$$

Let $\mathcal{I}$ be the set of family types. Families of type $i \in \mathcal{I}$ have youth's ability $a_i \in \mathcal{A}$, parental altruism $\delta_i \in [0,1]$, and parental wealth $w_i \in \mathbb{R}_+$. Let $f_i \in [0,1]$ be the fraction of type $i \in \mathcal{I}$ families. Suppose that $\mathcal{I}$ is partitioned into a finite number of groups indexed by $h \in \mathcal{H}$ such that $\mathcal{I} = \cup_{h\in\mathcal{H}}\mathcal{I}_h$. The social planner can only observe what group each family belongs to and sets subsidy schedules separately for each group. Let $g_{j,h}$ be the subsidy amount for $j \in \mathcal{J}$ and $h \in \mathcal{H}$, and let $\boldsymbol{g}_h \equiv (g_{j,h})_{j\in\mathcal{J}}$ be the subsidy schedule for $h \in \mathcal{H}$. In Section 4.3.2, the partition represents quartiles of parental income.

The social planner solves the following problem.

Problem 5.

$$\begin{aligned} \min_{(\boldsymbol{g}_h)_{h\in\mathcal{H}}} & \sum_{h\in\mathcal{H}} \sum_{i\in\mathcal{I}_h} f_i \tau(\boldsymbol{b}_i;\boldsymbol{g}_h,a_i) \\ \textit{subject to} & \sum_{h\in\mathcal{H}} \sum_{i\in\mathcal{I}_h} f_i \sum_{j\in\mathcal{J}} p_j(\boldsymbol{b}_i;\boldsymbol{g}_h,a_i) g_{j,h} \le G, \\ & \boldsymbol{g}_h \in \Gamma, \quad \forall h \in \mathcal{H}, \end{aligned}$$

where $\boldsymbol{b}_i \equiv (b_{j,i})_{j\in\mathcal{J}}$ *solves Problem 4 for* $(a_i, \delta_i, w_i, \boldsymbol{g}_h)$ *for all* $i \in \mathcal{I}_h$ *and for all* $h \in \mathcal{H}$.

With a binding social budget constraint, minimizing aggregate distortions is equivalent to maximizing families' welfare, as measured by $\sum_{t=1}^{T_p} R^{-t} c_p + \sum_{t=1}^{T_k} R^{-t} c_k$ aggregated across families.[34] (See Appendix B.20 for details.)

4.2 Calibration

I now discuss the calibration procedure. I first externally 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 section are denominated in 2004 US dollars using the consumer price index (CPI-U-RS).

The choice set of schooling is $\mathcal{J} = \{0, 1, 2, 4\}$. Therefore, in the model, individuals may not attend college, or attend college for one, two, or four years. Since relatively few 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 highest completed schooling.[35]

[34]In Section 3.3, it is shown that minimizing aggregate distortions is equivalent to maximizing youth's welfare if the social planner's budget constraint binds. This result no longer holds since parental transfers are endogenously determined. It reflects the fact that the social planner who cares only about youth's welfare, but not about parents', would encourage parents to give more to their children. See Phelan (2006) and Farhi and Werning (2007, 2010) for the implications of the intergenerational consumption distortion arising from the disagreement between the social planner and parents.

[35]Although the choice of college quality is not explicitly modeled, it is implicitly bundled with years of college in the calibration procedure because those attaining higher years are more likely to attend high quality institutions.

4.2.1 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 earnings for those choosing $j \in \mathcal{J}$ is

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

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 $T(\tilde{y})$ is the amount of income taxes paid.[36] I assume an annual interest rate of 3%, which implies $R = 1/\beta = 1.03$, and use the tax function $T(\tilde{y}) = 0.264[1-(0.012\tilde{y}^{0.964}+1)^{-1/0.964}]\tilde{y}$, estimated by Guner, Kaygusuz, and Ventura (2014).

Since the NLSY97 respondents, aged 12–17 in 1997, are still too young for the estimation of life-cycle earnings profiles, the parameters of the earnings function are estimated using data from the National Longitudinal Surveys of Youth 1979 (NLSY79) for the years 1979–2012. As a measure of ability, I use quartiles of the Armed Forces Qualifying Test (AFQT) scores.[37] I select all individuals from the random sample with at least 12 years and at most 17 years of completed schooling. Earnings of 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 annual earnings $\tilde{y}_j(a,x)$ as a function of years of schooling, ability, and experience. Appendix C.1 shows the OLS estimates of the regression (Table C.1) as well as the present discounted value of lifetime earnings, $y_j(a)$ (Table C.2).

4.2.2 Monetary Costs of Schooling

The monetary costs (k_j) and subsidies (g_j) for each schooling level are computed based on the average annual tuition and grant aid, estimated using data from the NPSAS:04 for full-time, full-year dependent students who applied for federal financial aid. The NPSAS:04 is chosen because the academic year 2003–2004 overlaps with the years most NLSY97 respondents attended college.[38] To capture the key features of the need-based financial aid system in a simple way, I partition the population by quartiles of parental income and estimate the average amount of grant aid separately.[39] Given the relatively small size of the market for private student loans, I assume that students can only borrow from the government. I take the cumulative limit, implied by annual limits of the Stafford Loan Program in

[36]Notice that the definition of schooling distortion does not account for the discouraging effects of taxation on education investment (Bovenberg and Jacobs, 2005). Taking it into account can have substantial effects on the level of optimal subsidy (Krueger and Ludwig, 2016), as well as the differences in optimal subsidy amounts by parental income (Colas, Findeisen, and Sachs, 2018). However, it would also require incorporating other education subsidies that are not distributed through financial aid, such as direct state appropriations to public post-secondary institutions, which is beyond the scope of this paper.

[37]The 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.

[38]The identity of post-secondary institutions the NLSY97 respondents attended is provided in the geocode file that is only available to US researchers.

[39]Since the NPSAS:04 contains only those enrolled in college, its parental 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.

2003–2004, as the borrowing limit ($\overline{d}_j$). Tables C.3 and C.4 in Appendix C.1 show the estimated annual and total amounts.

4.2.3 Preference Parameters

I assume that unobserved family characteristics—youth's psychic returns to schooling and parental altruism—are independently distributed with each other and with observed characteristics, such as parental income and youth's ability. 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.[40] They are chosen so that the model replicates empirical relationships between educational attainment, ability, parental income, and parental transfers from the NLSY97 data. Below, I lay out how each parameter affects certain features of the data.

The consumption-smoothing parameters can be inferred from how educational attainment and parental transfers vary with parental income and youth's ability. As shown in Section 4.1.3, the EIS has no effects on either of these decisions when the borrowing constraints do not bind. Therefore, we can learn more about the EGS from very wealthy families for which the borrowing constraints are less likely to bind. With binding borrowing constraints, both intertemporal and intergenerational elasticities of substitution affect parental transfer functions. For example, a strong and positive ability-transfer gradient would suggest a large value of the EGS relative to the EIS. Moreover, for a given value of the EGS, the educational attainment–transfer gradient is informative about the EIS: with a small EIS, the intertemporal consumption distortion associated with further schooling is large, and therefore educational attainment depends strongly on parental income. A small EIS also implies that the negative wealth effect associated with high ability is large, which weakens the strong, postive effects of ability on educational attainment that would prevail in the absence of borrowing constraints.

For given consumption-smoothing parameters and observed family characteristics, the distribution of psychic returns is crucial for the distribution of educational attainment, while the distribution of parental transfers is shaped by the distribution of parental altruism. Furthermore, the dispersion of parental altruism affects the parental income gradient of educational attainment: greater heterogeneity in parental altruism makes parental income a noisier signal of parental transfers, so the parental income gradient is attenuated toward zero.

Based on this argument, I choose three sets of target statistics (total of 18 targets) that would provide enough information to pin down the remaining eight parameters: (i) marginal distribution of educational attainment; (ii) OLS estimates of the regression of educational attainment on parental income quartiles and youth's ability quartiles; and (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.

The target statistics are computed using the NLSY97 data. The NLSY97 is a longitudinal survey of 8,984 Americans 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

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

parental transfers, which have been recently used to study the role of parental transfers in education choices (e.g., Johnson, 2013; Abbott et al., forthcoming; Colas, Findeisen, and Sachs, 2018). From these questions, I compute the amount of total parental transfers each youth received between ages 18 and 26, discounted back to age 17.[41] To be consistent with the measurement of parental transfers, education outcomes are also measured at age 26.

The fact that parental transfers are measured only during early adulthood in the data makes it difficult to compare them with those in the model, which should be thought of as the total amount received over the remaining lifetime. In particular, those who receive very high parental transfers during college are likely to receive more later in life because parents may not want to make all transfers when their children are still young. For example, parents might use bequests as compensation for certain services provided by their children (Bernheim, Shleifer, and Summers, 1985). Parents might also be afraid that young children would spend too much and ask for more later (Bruce and Waldman, 1990). On the other hand, parents who intend to give very little might not want to postpone giving because their children are likely to be borrowing constrained during schooling (Brown, Scholz, and Seshadri, 2012). These possibilities suggest that, although measured parental transfers are probably biased downward, the bias is likely to be small at the low end of the distribution, justifying using a binary variable with a low threshold value (instead of a continuous variable) as a dependent variable in the parental transfer regression.[42]

I choose the parameter vector that minimizes the weighted sum of squared differences between the statistics based on actual and simulated data. The calibration procedure is further explained in Appendix C.4. Table 4 reports the calibrated parameter values, and Table 5 shows the target statistics, along with the statistics based on the calibrated model. Table 6 compares the fraction of youth in each parental transfer quartile, separately by parental income quartile. Since they are not directly targeted, comparing them is another way to assess the calibrated model.

The EIS is smaller than 1, consistent with most estimates reported in Browning, Hansen, and Heckman (1999).[43] On the other hand, the EGS is greater than 1, which is in line with Córdoba and Ripoll (forthcoming).[44] Since the EGS is larger than the EIS, the desire to smooth consumption across generations is weaker than that across time (i.e., within the generation), and the model is able to replicate the positive relationship between parental transfers and youth's ability reported in Table 5c, although the relationship is weaker in the model. As Table 5b shows, the implied consumption-smoothing motives also replicate the parental income and youth's ability gradients of educational attainment observed in the data reasonably well, with the exception of the coefficient on the second

[41]The questions about parental transfers are available only until 2010, when the youngest cohort was 26 years old. See Appendix C.3 for details.

[42]Alternatively, the model could be modified to allow parents to give transfers more than once in order to be consistent with the measurement of parental transfers in the data. However, as shown by Brown, Scholz, and Seshadri (2012), the repeated interaction introduces another source of inefficiency that would exist even in the absence of borrowing constraints, which complicates the policy objective.

[43]The EIS is higher than 0.5, a commonly used value in the literature. The large EIS could reflect the lack of additional margins that individuals may adjust in response to borrowing constraints, such as college quality choice, delayed college entry, and labor supply during schooling.

[44]Córdoba and Ripoll (forthcoming) identify the EGS differently, based on its effect on fertility in a dynastic model of Becker and Barro (1988).

Table 4: Calibrated Preference Parameters

Parameter	Description	Value
Consumption smoothing:		
γ	EIS	0.840
η	EGS	1.167
Distribution of parental altruism:		
$E(\delta)$	Mean	0.342
$SD(\delta)$	Standard deviation	0.217
Distribution of psychic returns to schooling:		
μ_1		-64,301
μ_2	Location parameters for each schooling	-65,619
μ_4		-171,295
σ	Scale parameter	15,899

Table 5: Target Statistics and Model Fit

	Schooling Choice			
	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 the Highest Year of College Completed

	Parental Income Quartile			Ability 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 the Highest Year of College Completed

	Parental Income Quartile			Ability 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

Table 6: Distribution of Parental Transfers by Parental Income

Parental Income		% with Parental Transfer			
		Quartile 1	Quartile 2	Quartile 3	Quartile 4
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

parental income quartile.[45]

The parameters for the distribution of parental altruism imply that the average weight parents put on their children is around 34%, although the weights differ across families. As Table 6 shows, the calibrated distribution of parental altruism replicates the dispersion of parental transfers conditional on parental income, the key feature of the model. Finally, the large negative location parameters for the distribution of psychic returns suggest substantial psychic costs of college education, consistent with other studies in the literature (e.g., Carneiro, Hansen, and Heckman, 2003; Cunha, Heckman, and Navarro, 2005). Finishing the first year of college involves paying a psychic cost that is worth more than $60,000. Further, completing four years requires an additional $100,000, although the scale parameter implies that there is moderate heterogeneity in psychic returns.

4.3 Optimal Policies

Now I use the calibrated model to solve for optimal policies. In principle, subsidy amounts can depend on all family characteristics that are observed by the social planner. Therefore, separate subsidy schedules can be set for each level of parental wealth and youth's ability. However, solving Problem 5 with a very fine partition is computationally challenging, as the number of control variables ($g_{1,h}$, $g_{2,h}$, and $g_{4,h}$ for each h) increases with the number of groups. Therefore, I first solve for a single subsidy schedule for a group of families with identical parental wealth and youth's ability that only differ in unobservable characteristics—parental altruism and youth's psychic returns to schooling. This simple setup helps illustrate the mechanisms that were previously derived analytically. Next, I consider all families used for calibration, assuming that the social planner can only distinguish families by their parental income quartile. This assumption is an approximation to current financial aid policy,

[45] The fact that some of the targets in Tables 5b and 5c are not well fitted might reflect a correlation between observed and unobserved characteristics. For example, the stronger ability-transfer gradient in the data can be explained if more altruistic parents are more likely to have high-ability children because they invested more in early childhood education. However, introducing a positive correlation between parental altruism and youth's ability in the model could make the educational attainment–ability gradient stronger than that in the data.

parsimoniously capturing its need-based feature.

4.3.1 Policy Conditional on Parental Wealth and Youth's Ability

Consider families with identical parental wealth and youth's ability. In particular, I consider families in the highest ability and lowest parental income quartile, a group mostly likely to be affected by borrowing constraints, and set their parental wealth equal to the group average.

For this group of families, Table 7 shows the distribution of undistorted schooling that would be chosen if there were no borrowing constraints and a subsidy policy. Most youth from these families would attain a four-year college education even without subsidy if they were able to borrow as much as they wanted, although there is still variation in schooling choice due to heterogeneity in psychic returns. With borrowing constraints, however, only those with high parental transfers might be able to complete four years.

Table 7: Undistorted Schooling Distribution

	Schooling Choice			
	0	1	2	4
Fraction (%)	1.1	1.9	5.8	91.2

Differences in parental transfers among this group are driven by differences in parental altruism and psychic returns to schooling. While the heterogeneity in psychic returns induces parents to give different transfer amounts across schooling levels (i.e., $b_{j,i}$ varies with j), the parental transfer schedule itself differs across parents (i.e., $b_{j,i}$ varies with i) because of varying degrees of altruism, generating inequality in parental transfers conditional on schooling. Although youth with identical parental altruism might receive different parental transfer amounts depending on their schooling choice, they can be considered to have equal "parental support" in the sense that they face identical parental transfer schedules.

Panel A of Table 8 reports various model outcomes under the current subsidy schedule (first row) for the lowest parental income quartile. Those who attain higher schooling levels receive higher total subsidies, with annual amounts around \$6,000 for the first two years and \$7,000 for the last two years, and their parents also give higher transfers on average.[46] Despite a substantial amount of subsidy given to those completing four years of college, only 40% of youth finish four years. As Table 7 shows, this is less than half of what it would be if there were no borrowing constraints. The fourth row suggests that the borrowing constraints (21) bind for most youth (with high ability and low parental income) attending college.[47] Conditional on attending college, those who stay in school longer are less likely to

[46]Under the current financial aid policy, the annual amount of financial aid (except loans) does not explicitly depend on the number of years in college. The higher annual subsidy for later years reflects that (i) students who stay in school longer attend more expensive institutions (e.g., four-year rather than two-year institutions) and (ii) the amount of need-based financial aid is increasing in college costs.

[47]While those who do not attend college are able to fully smooth consumption, they could be still considered borrowing constrained in the sense that most of them would attend college if all borrowing constraints were removed.

Table 8: Effects of Current and Optimal Policies

	Schooling Choice				All
	0	1	2	4	
A. Current Policy					
Total Subsidy ($)		6,042	12,466	26,531	13,570
Parental transfer ($)	298	4,578	12,149	62,100	27,737
Fraction (%)	29.1	14.6	15.9	40.3	
borrowing constrained (%)	0	97.6	95.4	85.6	63.9
Marginal value of wealth	1	9.11	5.89	2.32	3.49
B. Optimal Policy					
Total Subsidy ($)		12,665	16,018	19,157	13,570
Parental transfer ($)	821	1,678	9,533	87,239	24,743
Fraction (%)	12.8	35.0	27.7	24.5	
borrowing constrained (%)	0	98.5	93.9	77.1	79.4
Marginal value of wealth	1	2.54	4.51	1.62	2.66

be borrowing constrained and have lower average marginal value of wealth.[48] As discussed in Section 3.5, this indicates an important role of differences in parental support (rather than psychic returns to schooling) for schooling choice.

Panel B of Table 8 reports the same outcomes under the optimal policy given the same budget as the current policy. The optimal policy reallocates the budget toward youth who attend college without completing four years. Subsidy for one year doubles while subsidy for four years decreases by a third. As a result, more youth attend college, but fewer stay for four years. While the optimal policy reduces the marginal value of wealth and lowers the fraction of students who are borrowing constrained among those attending one or two years, the borrowing constraints are more likely to bind overall because of higher college attendance. Increases in subsidy for one and two years are accompanied by reduced parental transfers, while a lower subsidy for four years leads to higher parental transfers.

Table 9: Changes in Distortions Between Current and Optimal Policies

Distortions	Change ($)	% Change
Consumption	-6,548	-35.6
Schooling	4,798	18.2
Total	-1,750	-3.9

Table 9 shows how the optimal policy affects distortions. The policy change improves consumption smoothing of those who attain low schooling because of low parental support. Since some youth who currently complete four years are now incentivized to leave college early, their consumption smoothing

[48]For each j, the marginal value of wealth is calculated by taking the average of $V_j'(b_{j,i} + g_j; a_i)$ across i.

is also enhanced. However, the policy change further distorts the intertemporal consumption allocation of those who stay until four years but remain borrowing constrained. Moreover, some non-college attendees with full consumption smoothing are induced to attend college and become borrowing constrained. Aggregate distortions in intertemporal consumption allocation are reduced, suggesting that the positive effects dominate. By contrast, distortions in schooling allocation rise. Although providing a generous subsidy during early years of college helps youth with low parental support attend college, it also provides incentives to attend college without finishing four years, even for those whose schooling decisions are mainly based on psychic returns rather than parental support. The efficiency loss in schooling investment is smaller than the gain in consumption smoothing, and the optimal policy reduces total distortions by $1,750 per youth, a 3.9% decrease.

These results demonstrate that policy improvements need not enhance schooling outcomes, which are often the emphasis of education policy analyses, because of the trade-off between schooling investment and consumption smoothing. For example, a policy that promotes college attendance is not necessarily optimal if it means that many college students have very distorted life-cycle consumption profiles. It is important to assess policies based on a criterion that captures this trade-off.

Table 10: Effects of Optimal Policy without Heterogeneity in Parental Altruism

	Schooling Choice			
	0	1	2	4
Total Subsidy ($)		0	5	19,545
Parental transfer ($)	17,152	20,128	26,690	28,499
Fraction (%)	4.0	8.0	18.6	69.4
borrowing constrained (%)	0	100	100	100
Marginal value of wealth	1	1.23	2.50	3.49

Role of Unobservable Heterogeneity in Parental Support The optimal subsidy schedule is shaped by the two dimensions of unobservable heterogeneity. To isolate the contribution of heterogeneity in parental altruism, Table 10 presents the optimal subsidy with homogeneous parental altruism (i.e., $\mathrm{SD}(\delta) = 0$) that also holds the budget fixed. In this case, all youth have equal parental support, so differences in schooling choices are entirely driven by differences in psychic returns. Consistent with Proposition 3, the optimal policy gives larger subsidies to those with more schooling. Those who stay in school longer also have a higher marginal value of wealth and thus, as shown by Proposition 4, receive higher parental transfers. With equal parental support, the borrowing constraints bind more severely for those investing more in schooling, in contrast to the case of heterogeneous parental altruism. This exercise reveals that inequality in parental support is responsible for the generous optimal subsidies during early years of college reported in Panel B of Table 8, confirming the insights provided by Proposition 1. Still, the increasing optimal subsidy schedule (shown in Panel B of Table 8) suggests that the heterogeneity in psychic returns to schooling also plays an important role in shaping the optimal subsidy schedule.

Table 11: Effects of Optimal Policy with Doubled Parental Wealth

	Schooling Choice			
	0	1	2	4
Total Subsidy ($)		14,316	17,738	12,210
Parental transfer ($)	5,338	6,915	23,878	167,960
Fraction (%)	7.1	25.0	30.7	37.2
borrowing constrained (%)	0	95.9	87.3	55.0
Marginal value of wealth	1	1.99	3.40	1.27

Effects of Higher Family Resources As described in Section 3.4, heterogeneity in parental support conditional on family resources also affects the optimal subsidy schedule by family resources. To demonstrate this, Table 11 reports the optimal policy that spends $13,570 per youth when parental wealth is doubled. Parental transfers increase substantially as parents become richer. Compared with the baseline case presented in Panel B of Table 8, the optimal policy gives higher subsidies to youth completing fewer than four years of college while those who finish four years receive less. This illustrates the mechanism described in Proposition 2—the force to redistribute toward youth with low parental support choosing low schooling levels becomes stronger for richer families. Of course, the negative relationship between optimal subsidy amounts and parental wealth during early years of college hinges on the assumption that the budget for education subsidy stays constant. Next, I explore the joint determination of subsidy schedules and budgets for families with different resources.

4.3.2 Differential Policy by Parental Income Quartiles

Now, I consider all families in the NLSY97 sample that are used to compute the calibration targets reported in Table 5 and partition them into four groups based on quartiles of parental income. Therefore, families in the same group facing identical subsidy schedules could differ by parental wealth and youth's ability as well as by parental altruism and youth's psychic returns to schooling.

Panel A of Table 12 shows current subsidy amounts by parental income quartile and schooling level, estimated from the NPSAS:04 data. It also reports the average amounts of subsidy, or budgets, that are computed based on the model under current policy. The amount of subsidy increases with schooling level and decreases with parental income quartile. For each level of schooling, the differences in subsidy amounts between students from the highest and lowest parental income quartiles are around 50%, with most of the gap occurring in the bottom three quartiles. However, the average amount of subsidy does not vary much across parental income quartiles, and it is the highest for the top quartile. The weak and non-monotonic relationship between budget and parental income is due to the positive correlation between schooling and parental income. Although students from high-income families receive low subsidies for each schooling level, they attain high levels for which subsidies are large. Their greater schooling levels are due to their high ability and large parental transfers, on average.

Panel B of Table 12 shows subsidy amounts under the optimal policy that spends the same total amount ($7,885) per youth. The optimal policy reallocates the budget toward lower quartiles of

Table 12: Current and Optimal Policies

Parental Income	Total Subsidy ($)			
	Schooling Choice			Budget
	1	2	4	
A. Current Policy				
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 Policy				
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

parental income, with the largest amount transferred to the bottom quartile. Despite the substantial decrease in the budget for the top quartile, the optimal subsidy for the first year is higher than the current amount as a result of within-group reallocation toward lower schooling levels. For youth in the top two parental income quartiles, optimal subsidies do not necessarily increase with years of college, meaning that they need to pay more than tuition in each year during later years of college. Compared with the current policy, the optimal subsidies for completing one or two years are generally higher, while they are lower for finishing four years (except for those in the top income quartile who leave after the second year). Youth in the highest parental income quartile receive significantly lower subsidies than the rest. Subsidy amounts for completing only one or two years of college education vary less among those in the bottom three quartiles: the gaps between the first and the third quartiles are 28.5% and 35.1%, respectively, compared with 42.9% and 42.4% under the current policy.

Table 13: Changes in Educational Attainment Between Current and Optimal Policies

Parental Income	Fraction with Schooling (pp)				Average Years
	0	1	2	4	
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 13 demonstrates how the optimal policy affects educational attainment. Percentage point changes in the fraction of youth choosing each schooling option are presented in columns two to five, and the final column reports changes in average years of college education completed. For all parental income quartiles, more youth leave college after the first or second year while fewer complete four years, reflecting more generous subsidies only during early years. College attendance rates are higher for all quartiles except for the top quartile, which receives considerably lower subsidy amounts. Because fewer youth complete four years, there is a decrease in the average years of college for all but the bottom quartile. Changes in the overall schooling distribution presented in the bottom row show that the optimal policy reduces inequality in educational attainment.

Table 14: Changes in Distortions Between Current and Optimal Policies

Parental Income	Distortions	Change ($)	% Change
Quartile 1	Consumption	-2,356	-32.8
	Schooling	397	5.5
	Total	-1,958	-13.6
Quartile 2	Consumption	0	0.0
	Schooling	-842	-9.3
	Total	-842	-6.5
Quartile 3	Consumption	560	19.0
	Schooling	-1,317	-13.3
	Total	-758	-5.9
Quartile 4	Consumption	-330	-14.1
	Schooling	686	11.9
	Total	356	4.4
All	Consumption	-532	-12.9
	Schooling	-269	-3.4
	Total	-801	-6.6

Table 15: Comparing Education Attainment: Optimal Policy Relative to Undistorted Allocation

Parental Income	Fraction with Schooling (pp)				Average Years
	0	1	2	4	
Quartile 1	-7.2	14.9	12.3	-20.1	-0.41
Quartile 2	1.0	12.2	5.5	-18.7	-0.52
Quartile 3	0.8	11.4	6.9	-19.1	-0.51
Quartile 4	5.1	6.6	4.1	-15.8	-0.48
All	-0.1	11.3	7.2	-18.4	-0.48

Table 14 shows that the reduction in distortions associated with the optimal policy are greatest for lower parental income groups. Indeed, distortions increase for the top income quartile. Within each parental income quartile, consumption and schooling distortions change in opposite directions, which reflects the intrinsic trade-off between them. As discussed earlier, in the presence of within-group inequality in parental support, providing large subsidies for early years can improve efficiency by helping those with low parental support attend college and smooth consumption, but it also has adverse effects on the schooling decisions of others.

For the bottom quartile, the efficiency gain of the policy change is entirely driven by better consumption smoothing, while schooling investment is more distorted partly as a result of an inefficiently high college attendance rate. To illustrate this, Table 15 compares the schooling distribution under the optimal policy with that of the undistorted allocation. It shows that the optimal policy raises the college attendance rate for the bottom quartile beyond the undistorted level, suggesting that some youth with low schooling returns are induced to attend college just because of the generous early subsidy.

In contrast, the increases in college attendance rates for the second and third parental income quartiles are not excessive in the sense that their college attendance rates are still below the undistorted levels. They indeed improve the efficiency of their schooling investment despite large reductions in the fraction of those completing four years. For these groups, the efficiency gain of the policy change is driven by more efficient schooling investment (rather than better consumption smoothing) for those with low parental support.

Distortions for the top parental income quartile increase because the large reduction in their budget aggravates the efficiency of schooling investment for those who are borrowing constrained. The budget cut is largely met by lower subsidy for four years and subsidy for one year is instead increased, limiting potential adverse effects of the policy change on those who attain low schooling due to low parental support. The increase in subsidy for one year, along with a general decline in educational attainment, improves overall consumption smoothing for this group, but its magnitude is small.

Table 16: Effects of Removing Current Subsidy

Distortions	Change ($)	% Change
Consumption	-2,743	-66.6
Schooling	4,290	53.8
Total	1,547	12.8

Overall, the optimal policy reduces aggregate distortions in schooling by $269 per youth, despite the lower average years of college education. Combined with a larger reduction in the distortion in intertemporal consumption allocation ($532), the optimal policy reduces aggregate distortions by $801 per youth, or 6.6%. To evaluate the magnitude of this change, Table 16 presents the effects of removing the current policy (i.e., setting $\boldsymbol{g} = \mathbf{0}$) for all quartiles. The removal raises aggregate distortions by $1,547, which is about twice as large as the amount of aggregate distortions reduced by moving from current to the optimal policy. Put differently, starting from zero subsidy, the current policy spends $7,885 per youth and reduces aggregate distortions by $1,547. The optimal policy further reduces distortions by $801 by simply reallocating the same amount of budget. This represents a dramatic improvement in allocative efficiency.

5 Conclusions

As the costs of higher education are rising, there is a growing concern that many youth without significant help from their parents are unable to access the resources they need to attend college. While the current financial aid system and related policy discussions are primarily focused on addressing disparity in financial resources between rich and poor families, this paper provides evidence that substantial inequality in parental support exists even among college students with similar family resources. This presents a challenge for the financial aid system because, unlike family resources, parental transfers cannot be easily observed by financial aid authorities.

I show that the presence of unobservable heterogeneity in parental transfers conditional on family resources has important implications for the design of need-based financial aid, even ignoring any concerns about equity. Since those with low parental transfers are more likely to be borrowing constrained and to under-invest, they can be targeted by providing high subsidies at low schooling levels. Because parental transfers are imperfectly correlated with family resources, the generous subsidy for low schooling levels need not decline steeply with family resources. The quantitative analysis suggests a reform that reallocates public spending toward those attaining one or two years of college education in order to better target borrowing-constrained students who leave college early (or do not attend) because of low parental transfers. The reform also reduces the gap in subsidy amounts by parental income during early years of college, providing enhanced public support for some college education to all youth—even those from high-income families. The potential efficiency gains of such a budget-neutral reform would be substantial.

Universally expanding public support for the first two years of college has been a widely debated topic recently, especially since President Obama introduced America's College Promise, a legislative initiative to make community college tuition-free, in 2015. Although free community college proposals drew considerable media attention and led to legislation at the local and state level, critics argue that they are badly targeted because "covering the full tuition of all community college students would mean middle-income, and even upper-income, students would get hefty subsidies, even though many do not need the help" (Butler, 2015, January, 20, para. 5). The results presented in this paper suggest that eliminating community college tuitions may improve efficiency because there exist a large number of students from middle- and high-income families who do not receive substantial financial support from parents and thus would benefit from free tuition. However, this paper also implies that the generous subsidy for the first two years of college must be accompanied by lower subsidies for later years because those with sufficient parental support (i.e., those who do not need the subsidy) are more likely to attend four-year institutions. Therefore, free community college proposals should be discussed in a broader context of restructuring the entire tuition and financial aid policies.

The idea of concentrating financial aid on early years of college, known as "front-loading," has already been proposed as a strategy to promote college access (e.g., Kane, 1999), although primary attention has been given to reforming the Federal Pell Grant program for students from low-income families. While proposals to front-load Pell Grants come up often during reauthorization of the Higher Education Act, they have been too controversial to be adopted because of concerns that they might create incentives for students to drop out in later years when grants are reduced (Mercer, 2007). However, the arguments for and against front-loading are concerned with its effects on particular outcomes (e.g., college attendance, persistence of enrollment, and program cost) without a framework that weighs all potential impacts. Moreover, most policy discussions are focused on improving education outcomes and ignore providing better consumption smoothing over the life cycle. This paper contributes to this debate by showing that front-loading can improve efficiency in both schooling and consumption smoothing, because it enables better targeting of borrowing-constrained students.

This paper abstracts from several important issues in order to focus on the problem of reallocating the current budget to reduce inefficiencies caused by borrowing constraints. The social planner's ability to allocate resources across time or generations is limited because the social planner cannot increase the borrowing limits or raise additional tax revenues for higher education purposes. While it might be possible to provide sufficient student loans or subsidies that would allow everyone to get a college education, benefits of such a policy must be weighed against costs associated with loan defaults and financing the expenditure. This paper also abstracts from risks in the returns from schooling and the role of family in providing insurance against them (e.g., Kaplan, 2012). If parents who give higher transfers during school also offer greater insurance against post-schooling outcomes, then the case for providing public support for those without such parental support will be strengthened. Since it is difficult for the social planner to identify who has access to family insurance, designing a public insurance program (e.g., income-contingent student loan repayment) that effectively targets those who would benefit the most faces a simliar challenge as the one described in this paper. While recent studies explore the design of student loan programs that provide public insurance against labor market risk (Gary-Bobo and Trannoy, 2015; Findeisen and Sachs, 2016; Lochner and Monge-Naranjo, 2016), they do not account for the existence of and potential heterogeneity in access to family insurance, which is found to play an important role in repaying student loans (Lochner, Stinebrickner, and Suleymanoglu, 2018). Future work should devote attention to these issues.

References

Abbott, Brant, Giovanni Gallipoli, Costas Meghir, and Giovanni L. Violante. forthcoming. "Education Policy and Intergenerational Transfers in Equilibrium." *Journal of Political Economy*.

Becker, Gary S. 1975. *Human Capital*. New York, NY: Columbia University Press.

Becker, Gary S. and Robert J. Barro. 1988. "A Reformulation of the Economic Theory of Fertility." *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):S1–S39.

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.

Bernheim, B. Douglas, Andrei Shleifer, and Lawrence H. Summers. 1985. "The Strategic Bequest Motive." *Journal of Political Economy* 93 (6):1045–1076.

Bils, Mark and Peter J. Klenow. 2000. "Does Schooling Cause Growth?" *American Economic Review* 90 (5):1160–1183.

Bohacek, Radim and Marek Kapicka. 2008. "Optimal Human Capital Policies." *Journal of Monetary Economics* 55 (1):1–16. Carnegie-Rochester Conference Series on Public Policy: Advances in Dynamic Public Finance April 20–21, 2007.

Bovenberg, A. Lans and Bas Jacobs. 2005. "Redistribution and Education Subsidies are Siamese Twins." *Journal of Public Economics* 89 (11):2005–2035.

Brown, Meta, John Karl Scholz, and Ananth Seshadri. 2012. "A New Test of Borrowing Constraints for Education." *Review of Economic Studies* 79 (2):511–538.

Browning, Martin, Lars Peter Hansen, and James J. Heckman. 1999. "Micro Data and General Equilibrium Models." In *Handbook of Macroeconomics*, vol. 1, Part A, edited by John B. Taylor and Michael Woodford, chap. 8. Elsevier, 543–633.

Bruce, Neil and Michael Waldman. 1990. "The Rotten-Kid Theorem Meets the Samaritan's Dilemma." *Quarterly Journal of Economics* 105 (1):155–165.

Butler, Stuart M. 2015, January, 20. "Obama's SOTU Free College Plan is Bad for Poor Americans." https://www.brookings.edu/opinions/why-obamas-free-community-college-idea-is-a-bad-one/. (accessed: January 15, 2019).

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. Vytlacil. 2011. "Estimating Marginal Returns to Education." *American Economic Review* 101 (6):2754–2781.

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. forthcoming. "Early and Late Human Capital Investments, Borrowing Constraints, and the Family." *Journal of Political Economy*.

Caucutt, Elizabeth M., Lance Lochner, and Youngmin Park. 2017. "Correlation, Consumption, Confusion, or Constraints: Why Do Poor Children Perform So Poorly?" *Scandinavian Journal of Economics* 119 (1):102–147.

Colas, Mark, Sebastian Findeisen, and Dominik Sachs. 2018. "Optimal Need-Based Financial Aid." Unpublished.

College Board. 2016. *Trends in Student Aid 2016*. New York, NY: College Board. (ERIC Document Reproduction Service No. ED 572 538).

Consumer Financial Protection Bureau. 2012. *Private Student Loans*. Washington, DC: Consumer Financial Protection Bureau. (ERIC Document Reproduction Service No. ED 534 984).

Córdoba, Juan Carlos and Marla Ripoll. forthcoming. “The Elasticity of Intergenerational Substitution, Parental Altruism, and Fertility Choice.” *Review of Economic Studies* .

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.” *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.

Do, Chau. 2004. “The Effects of Local Colleges on the Quality of College Attended.” *Economics of Education Review* 23 (3):249–257.

Dorans, Neil J. 1999. “Correspondences Between ACT and SAT I Scores.” *ETS Research Report Series* 1999 (1):i–18.

Edlin, Aaron S. and Chris Shannon. 1998. “Strict Single Crossing and the Strict Spence-Mirrlees Condition: A Comment on Monotone Comparative Statics.” *Econometrica* 66 (6):1417–1425.

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, 283–324.

Farhi, Emmanuel and Iván Werning. 2007. “Inequality and Social Discounting.” *Journal of Political Economy* 115 (3):365–402.

———. 2010. “Progressive Estate Taxation.” *Quarterly Journal of Economics* 125 (2):635–673.

Federal Student Aid. 2011. *2011–2012 Federal Student Aid Handbook*. Jessup, MD: Federal Student Aid, US Department of Education. (ERIC Document Reproduction Service No. ED 537 360).

Findeisen, Sebastian and Dominik Sachs. 2016. “Education and Optimal Dynamic Taxation: The Role of Income-Contingent Student Loans.” *Journal of Public Economics* 138:1–21.

Flodén, Martin. 2001. “The Effectiveness of Government Debt and Transfers as Insurance.” *Journal of Monetary Economics* 48 (1):81–108.

Gary-Bobo, Robert J. and Alain Trannoy. 2015. "Optimal Student Loans and Graduate Tax Under Moral Hazard and Adverse Selection." *The RAND Journal of Economics* 46 (3):546–576.

Gomes, Joao, Jeremy Greenwood, and Sergio Rebelo. 2001. "Equilibrium Unemployment." *Journal of Monetary Economics* 48 (1):109–152.

Gottlieb, Daniel and Humberto Moreira. 2012. "Should Educational Policies Be Regressive?" *Journal of Public Economic Theory* 14 (4):601–623.

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. 2003. "Redistribution through Education and Other Transfer Mechanisms." *Journal of Monetary Economics* 50 (8):1719–1750.

———. 2014. "Borrowing Constraints, College Aid, and Intergenerational Mobility." *Journal of Human Capital* 8 (1):1–41.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante. 2008. "Insurance and Opportunities: A Welfare Analysis of Labor Market Risk." *Journal of Monetary Economics* 55 (3):501–525.

———. 2017. "Optimal Tax Progressivity: An Analytical Framework." *Quarterly Journal of Economics* 132 (4):1693–1754.

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.

Hellwig, Martin F. 2010. "Incentive Problems with Unidimensional Hidden Characteristics: A Unified Approach." *Econometrica* 78 (4):1201–1237.

Hicks, J. R. 1940. "The Valuation of the Social Income." *Economica* 7 (26):105–124.

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." *Economic Journal* 49 (195):549–552.

Kane, Thomas J. 1999. *The Price of Admission: Rethinking How Americans Pay for College*. Brookings Institution Press.

Kaplan, Greg. 2012. "Moving Back Home: Insurance against Labor Market Risk." *Journal of Political Economy* 120 (3):446–512.

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.

Kehoe, Timothy J. and David K. Levine. 1993. "Debt-Constrained Asset Markets." *Review of Economic Studies* 60 (4):865–888.

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.

Kwang, Ng Yew. 1965. "Why do People Buy Lottery Tickets? Choices Involving Risk and the Indivisibility of Expenditure." *Journal of Political Economy* 73 (5):530–535.

Lochner, Lance and Alexander Monge-Naranjo. 2012. "Credit Constraints in Education." *Annual Review of Economics* 4 (1):225–256.

———. 2016. "Student Loans and Repayment: Theory, Evidence, and Policy." In *Handbook of the Economics of Education*, vol. 5, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, chap. 8. Elsevier, 397–478.

Lochner, Lance, Todd Stinebrickner, and Utku Suleymanoglu. 2018. "Parental Support, Savings and Student Loan Repayment." Centre for Human Capital and Productivity (CHCP) Working Paper Series 2018-2, University of Western Ontario.

Lochner, Lance J. and Alexander Monge-Naranjo. 2011. "The Nature of Credit Constraints and Human Capital." *American Economic Review* 101 (6):2487–2529.

Mankiw, N. Gregory and Matthew Weinzierl. 2010. "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution." *American Economic Journal: Economic Policy* 2 (1):155–176.

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.

McKenzie, Lionel. 1957. "Demand Theory Without a Utility Index." *Review of Economic Studies* 24 (3):185–189.

McPherson, Michael S. and Morton Owen Schapiro. 1998. *The Student Aid Game: Meeting Need and Rewarding Talent in American Higher Education.* Princeton University Press.

Mercer, Charmaine. 2007. "Federal Pell Grant Program of the Higher Education Act: Background and Reauthorization." Congressional Research Service Report for Congress Order Code RL31668, Washington, DC: Congressional Research Service.

Milgrom, Paul and Ilya Segal. 2002. "Envelope Theorems for Arbitrary Choice Sets." *Econometrica* 70 (2):583–601.

Milgrom, Paul and Chris Shannon. 1994. "Monotone Comparative Statics." *Econometrica* 62 (1):157–180.

Mirrlees, James A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38 (2):175–208.

Phelan, Christopher. 2006. "Opportunity and Social Mobility." *Review of Economic Studies* 73 (2):487–504.

Restuccia, Diego and Carlos Urrutia. 2004. "Intergenerational Persistence of Earnings: The Role of Early and College Education." *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. 2012. *How America Pays for College 2012*. Newark, DE: Sallie Mae Bank. (ERIC Document Reproduction Service No. ED 540 414).

———. 2017. *How America Pays for College 2017*. Newark, DE: Sallie Mae Bank. (ERIC Document Reproduction Service No. ED 577 456).

Samuelson, Paul A. 1974. "Complementarity: An Essay on the 40th Anniversary of the Hicks-Allen Revolution in Demand Theory." *Journal of Economic Literature* 12 (4):1255–1289.

Stantcheva, Stefanie. 2017. "Optimal Taxation and Human Capital Policies over the Life Cycle." *Journal of Political Economy* 125 (6):1931–1990.

Stinebrickner, Ralph and Todd Stinebrickner. 2008. "The Effect of Credit Constraints on the College Drop-Out Decision: A Direct Approach Using a New Panel Study." *American Economic Review* 98 (5):2163–2184.

Vereshchagina, Galina and Hugo A. Hopenhayn. 2009. "Risk Taking by Entrepreneurs." *American Economic Review* 99 (5):1808–30.

Appendices

A Additional Empirical Results from the NPSAS:12 Data

A.1 Conditional Distribution of Parental Contribution

Table A.1: EFC and Actual Parental Contribution by Subgroups

EFC	% with Amount Parents Paid toward Expenses					
	$0	$1 to $2,000	$2,001 to $5,000	$5,001 to $10,000	$10,001 to $20,000	$20,001 or More
			A. Cost of Attendance $20,000 or Less			
$0	42.7	35.3	8.3	6.4	4.2	3.1
$1 to $2,000	27.6	34.8	12.9	14.4	8.8	1.4
$2,001 to $5,000	33.2	35.6	13.1	11.1	5.1	1.9
$5,001 to $10,000	20.0	33.5	21.0	14.0	8.4	3.1
$10,001 to $20,000	15.6	27.7	20.9	15.7	13.6	6.4
$20,001 or More	15.8	23.0	16.9	18.9	14.1	11.2
			B. No Own Income			
$0	39.8	37.5	8.3	7.8	4.4	2.2
$1 to $2,000	19.3	37.7	14.8	11.3	13.2	3.6
$2,001 to $5,000	22.5	41.1	11.5	11.3	9.3	4.2
$5,001 to $10,000	11.0	35.1	16.7	20.5	12.1	4.5
$10,001 to $20,000	9.9	25.7	17.8	16.1	20.2	10.4
$20,001 or More	9.0	17.3	8.6	17.8	25.9	21.3
			C. No PLUS Loans			
$0	41.0	35.2	8.0	7.8	4.9	3.0
$1 to $2,000	29.2	37.2	11.4	10.4	9.4	2.4
$2,001 to $5,000	32.5	34.8	12.8	10.5	6.4	3.0
$5,001 to $10,000	21.2	31.5	16.8	14.8	11.3	4.4
$10,001 to $20,000	16.0	24.3	17.1	16.1	17.2	9.2
$20,001 or More	11.5	15.7	11.9	19.0	22.0	19.9
			D. No Private Loans			
$0	40.8	34.8	8.8	7.8	4.9	2.8
$1 to $2,000	28.0	38.6	11.2	10.0	9.5	2.7
$2,001 to $5,000	30.6	34.8	12.6	12.0	6.8	3.2
$5,001 to $10,000	19.8	31.0	16.8	15.3	12.5	4.6
$10,001 to $20,000	12.1	26.2	17.8	15.2	19.4	9.4
$20,001 or More	11.4	15.9	12.0	17.5	24.0	19.3

There are several financial factors that might drive the differences in reported parental contribution conditional on EFC. First, those differences could reflect the variation in college costs that exists even among four-year public institutions as a result of differences in tuition fees across states, as well as differences in cost of living across regions. Second, the EFC may not correctly measure parental resources because it includes students' contribution from their own earnings and assets as well as their parents' contribution. Third, the parental contribution reported by students may not reflect all forms of parental help. For example, parents can borrow from the federal government through Parent Loan for Undergraduate Students (PLUS), which may not be acknowledged by their children because PLUS loans are directly disbursed to schools to lower students' out-of-pocket payments. Moreover, parents can help their children take out private loans by co-signing with them. Since co-signing makes parents liable for the debt in the case of default, it could be considered an alternative form of help: a future transfer at the time of an adverse labor market outcome.

To address these concerns, Table A.1 presents results separately by subgroups. Panel A shows results for students with costs less than $20,000, which is lower than both the average cost ($22,336) and the median cost ($21,440). Panel B selects students who did not have income in the previous year (38.6%), for whom the EFC mostly reflects the expected contribution of their parents. Panel C are those whose parents did not take out PLUS loans (82%), and Panel D shows results for those who did not take out private loans (90.6%). These results show that the results in Table 2 are robust to accounting for these factors.

A.2 Determinants of Parental Contribution

This subsection estimates the effects of various individual and family characteristics on parental contribution. Table A.2 reports OLS estimates for a linear probability model using the probability of receiving more than certain amounts from parents as dependent variables.

Consistent with Table 2, students with higher EFC are significantly more likely to receive greater amounts from their parents. The probability of receiving higher parental transfer is also significantly increasing in the net cost of college. This is not surprising because one might expect that the main considerations for parents making transfer decisions are how much money they have and how much their children need to pay for college. Class level is another variable that determines students' out-of-pocket payment through borrowing limits. As mentioned earlier, students in their second (third or higher) year could borrow $1,000 ($2,000) more from the federal government in 2011–2012 compared with first-year students, so their parents are more likely to make lower annual contributions. However, this effect might be offset by selection if those with lower parental contributions are more likely to drop out and do not pursue higher levels. Indeed, the empirical relationship between class levels and parental contribution is non-monotonic: the marginal effects of higher class level is negative only up to the third year, when the annual borrowing limit reaches its maximum, suggesting a potential role for selection in inducing the positive marginal effect beyond the third year.

Despite the importance of financial factors for parental contribution, some characteristics that are not directly related to financial resources currently available for the family, and thus not used for the

Table A.2: Estimated Effects on Parental Transfers

	Parents Contributed More Than				
	$0	$2,000	$5,000	$10,000	$20,000
	(1)	(2)	(3)	(4)	(5)
EFC:					
$1 to $2,000	0.105*	0.068*	0.042	0.029	-0.012
	(0.030)	(0.030)	(0.028)	(0.022)	(0.012)
$2,001 to $5,000	0.083*	0.058*	0.011	-0.010	-0.009
	(0.027)	(0.027)	(0.023)	(0.018)	(0.011)
$5,001 to $10,000	0.169*	0.150*	0.058*	0.020	-0.012
	(0.031)	(0.029)	(0.027)	(0.022)	(0.013)
$10,001 to $20,000	0.220*	0.248*	0.149*	0.110*	0.022
	(0.026)	(0.029)	(0.024)	(0.022)	(0.013)
$20,001 or more	0.236*	0.350*	0.296*	0.230*	0.116*
	(0.027)	(0.030)	(0.027)	(0.021)	(0.016)
Net cost of attendance:					
$10,001 to $15,000	0.045*	0.081*	0.059*	0.008	-0.007
	(0.023)	(0.024)	(0.019)	(0.015)	(0.008)
$15,001 to $20,000	0.044	0.089*	0.091*	0.058*	0.023*
	(0.024)	(0.024)	(0.020)	(0.017)	(0.010)
$20,001 or more	0.059*	0.143*	0.182*	0.137*	0.060*
	(0.024)	(0.028)	(0.023)	(0.020)	(0.013)
Class level for loans:					
2nd year	-0.034	-0.039*	-0.041*	-0.044*	-0.019
	(0.019)	(0.019)	(0.018)	(0.016)	(0.011)
3rd year	-0.101*	-0.085*	-0.083*	-0.067*	-0.029*
	(0.020)	(0.022)	(0.020)	(0.018)	(0.012)
4th year or higher	-0.083*	-0.033	-0.045*	-0.057*	-0.023
	(0.018)	(0.017)	(0.018)	(0.017)	(0.013)
Parents' highest education:					
High school diploma or equivalent	0.086*	0.061	0.056	0.011	0.018
	(0.042)	(0.033)	(0.032)	(0.030)	(0.018)
Some post-secondary education	0.085	0.069*	0.049	-0.011	0.002
	(0.046)	(0.034)	(0.032)	(0.028)	(0.017)
Bachelor's degree	0.133*	0.112*	0.092*	0.019	0.018
	(0.045)	(0.035)	(0.033)	(0.026)	(0.017)
Master's degree or higher	0.148*	0.169*	0.157*	0.085*	0.023

Continued on next page

Table A.2 *(Continued)*

	Parents Contributed More Than				
	$0	$2,000	$5,000	$10,000	$20,000
	(1)	(2)	(3)	(4)	(5)
	(0.045)	(0.038)	(0.037)	(0.030)	(0.019)
Foreign-born parent	0.039	0.018	-0.006	0.011	-0.008
	(0.023)	(0.022)	(0.021)	(0.017)	(0.013)
Female	-0.009	0.023	0.041*	0.027*	0.021*
	(0.014)	(0.016)	(0.015)	(0.012)	(0.010)
Race:					
White	-0.001	0.054	0.009	-0.010	-0.024
	(0.030)	(0.029)	(0.029)	(0.026)	(0.017)
Black	0.023	0.040	-0.009	-0.020	-0.032
	(0.032)	(0.033)	(0.032)	(0.026)	(0.018)
Asian	0.068	0.158*	0.107*	0.072*	0.055
	(0.043)	(0.042)	(0.036)	(0.034)	(0.030)
Selectivity of institution:[50]					
Minimally selective	0.004	-0.012	0.027	-0.016	-0.007
	(0.044)	(0.038)	(0.036)	(0.025)	(0.015)
Moderately selective	-0.006	0.005	0.048	0.031	-0.005
	(0.034)	(0.033)	(0.029)	(0.020)	(0.010)
Very selective	0.052	0.051	0.117*	0.072*	0.016
	(0.040)	(0.036)	(0.035)	(0.026)	(0.013)
Grade point average in high school:					
2.0 to 2.4	0.066	0.001	-0.026	-0.021	-0.024
	(0.114)	(0.087)	(0.088)	(0.070)	(0.045)
2.5 to 2.9	0.081	0.064	0.027	0.031	0.006
	(0.108)	(0.083)	(0.082)	(0.073)	(0.044)
3.0 to 3.4	0.093	0.017	-0.013	0.015	-0.002
	(0.108)	(0.077)	(0.080)	(0.071)	(0.043)
3.5 to 4.0	0.073	0.001	-0.024	0.000	-0.007
	(0.107)	(0.076)	(0.080)	(0.069)	(0.042)
SAT math score:[51]					
351 to 450	-0.008	-0.038	-0.007	0.013	0.005

Continued on next page

[50] The selectivity measure was developed for the Integrated Postsecondary Education Data System (IPEDS), based on the following criteria: whether the institution was open admission (no minimal requirements); the number of applicants; the number of students admitted; the 25th and 75th percentiles of ACT and/or SAT scores; and whether or not test scores were required.

[51] For those who took ACT only, their scores are converted to SAT scores according to a concordance table from Dorans (1999).

Table A.2 *(Continued)*

	Parents Contributed More Than				
	\$0	\$2,000	\$5,000	\$10,000	\$20,000
	(1)	(2)	(3)	(4)	(5)
	(0.045)	(0.048)	(0.037)	(0.034)	(0.022)
451 to 550	-0.035	-0.033	-0.007	0.007	0.005
	(0.047)	(0.049)	(0.040)	(0.033)	(0.025)
551 to 650	-0.046	-0.054	-0.023	0.006	0.010
	(0.048)	(0.053)	(0.041)	(0.033)	(0.024)
651 to 800	-0.075	-0.034	-0.002	0.012	0.053
	(0.057)	(0.063)	(0.052)	(0.043)	(0.032)
SAT verbal score:					
351 to 450	0.054	0.052	-0.010	-0.020	0.003
	(0.044)	(0.045)	(0.042)	(0.034)	(0.021)
451 to 550	0.076	0.061	-0.016	-0.016	0.002
	(0.047)	(0.046)	(0.044)	(0.037)	(0.021)
551 to 650	0.062	0.076	0.010	0.003	0.004
	(0.049)	(0.051)	(0.048)	(0.038)	(0.021)
651 to 800	0.073	0.067	-0.044	-0.033	-0.007
	(0.056)	(0.056)	(0.056)	(0.045)	(0.027)
Field of study:					
Computer and information sciences	-0.004	-0.077	-0.063	-0.026	-0.058
	(0.064)	(0.076)	(0.072)	(0.065)	(0.033)
Engineering and engineering technology	0.062	0.024	0.011	0.016	-0.012
	(0.061)	(0.061)	(0.050)	(0.043)	(0.032)
Bio & phys science, sci tech, math,	0.065	0.018	0.036	0.028	0.004
agriculture	(0.052)	(0.058)	(0.053)	(0.043)	(0.031)
General studies and other	0.157*	0.096	0.105	0.042	0.016
	(0.060)	(0.068)	(0.063)	(0.051)	(0.034)
Social sciences	0.004	-0.038	-0.035	-0.032	-0.032
	(0.058)	(0.060)	(0.049)	(0.042)	(0.029)
Humanities	0.054	-0.031	0.013	0.016	-0.002
	(0.055)	(0.051)	(0.048)	(0.044)	(0.031)
Health care fields	0.046	-0.019	-0.043	-0.049	-0.036
	(0.055)	(0.061)	(0.050)	(0.042)	(0.030)
Business	0.041	0.005	0.020	-0.015	-0.014
	(0.055)	(0.058)	(0.050)	(0.042)	(0.030)
Education	0.033	-0.075	-0.060	-0.038	-0.024
	(0.058)	(0.058)	(0.050)	(0.042)	(0.029)

Continued on next page

Table A.2 *(Continued)*

	Parents Contributed More Than				
	\$0	\$2,000	\$5,000	\$10,000	\$20,000
	(1)	(2)	(3)	(4)	(5)
Other applied	0.037	-0.013	-0.024	-0.033	-0.019
	(0.055)	(0.056)	(0.045)	(0.041)	(0.028)
Constant	0.351*	0.048	0.038	0.050	0.048
	(0.139)	(0.109)	(0.097)	(0.088)	(0.056)
R^2	0.093	0.158	0.166	0.147	0.083

Notes: Linear probability models estimated using OLS. Sample weights are used. Standard errors in parentheses.

* Significant at the 5% level.

EFC calculation, matter as well. Higher educated parents, especially those with at least a bachelor's degree, are significantly more likely to contribute more. Female students are more likely to receive larger amounts (more than \$5,000) than males, although the effects are small. Asian students tend to receive larger parental transfers, but other variables of race and immigrant status of parents do not have significant effects on parental contribution. However, most variables related to students' academic ability or returns to education—selectivity of institution, high school GPA, SAT scores, and college major—are not significantly correlated with parental transfers, although attending very selective institutions has some significant effects.[49] Moreover, the R-squared statistics at the bottom of the table reveal that the observed variables used as regressors do not explain most of the variation in parental contribution.

B Proofs and Analytical Details

B.1 Proof of Lemma 1

The empty set is always both less than and greater than any other set in the strong set order, so it suffices to only consider the values of b satisfying $\operatorname{argmax}_{j\in\mathcal{J}(b)} U_j(b) \neq \emptyset$. Since $\operatorname{argmax}_{j\in\mathcal{J}(b)} U_j(b) = \operatorname{argmax}_{j\in\mathcal{J}(b)} V_j(b)$ (where $V_j(b)$ is defined in Section 3.2) in this case, I switch to the money-metric utility function $e(\ln c_1 + \ln c_2) = 2(c_1 c_2)^{1/2}$ from now on.

[49]This does not necessarily mean that the amount of parental transfer over the lifetime is not correlated with ability since those with higher ability are more likely to stay in college longer.

Define $\Theta \equiv \{(y, n, b) \in \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+ \,|n \leq b\}$ and $\mathcal{V} : \Theta \to \mathbb{R}_+$ as follows:

$$\mathcal{V}(y, n, b) \equiv \max_{(c_1, c_2) \in \mathbb{R}^2_+} 2(c_1 c_2)^{\frac{1}{2}} \tag{30}$$

$$\text{subject to } c_1 + c_2 \leq b - n + y, \tag{31}$$

$$c_1 \leq b - n. \tag{32}$$

It is easy to see that $\mathcal{V}(y, n, b)$ is strictly increasing in y (unless $b = n$), strictly decreasing in n, strictly increasing in b, and quasiconcave and continuously differentiable in (y, n). Importantly, the following Mirrlees-Spence condition is satisfied: for all $b' > b$,

$$\frac{\mathcal{V}_y(y, n, b)}{|\mathcal{V}_n(y, n, b)|} \leq \frac{\mathcal{V}_y(y, n, b')}{|\mathcal{V}_n(y, n, b')|}, \tag{MS}$$

where $\mathcal{V}_y(y, n, b) \equiv \partial\mathcal{V}(y, n, b)/\partial y = \min\{[(b - n)/y]^{1/2}, 1\}$ and $|\mathcal{V}_n(y, n, b)| \equiv |\partial\mathcal{V}(y, n, b)/\partial n| = \max\{[y/(b - n)]^{1/2}, 1\} > 0$ are calculated by the envelope theorem. As shown by Milgrom and Shannon (1994) and Edlin and Shannon (1998), (MS) is equivalent to the following single-crossing differences condition: for all $y' > y$ and $b' > b$,

$$\mathcal{V}(y', n', b) \geq (>) \mathcal{V}(y, n, b) \Rightarrow \mathcal{V}(y', n', b') \geq (>) \mathcal{V}(y, n, b'). \tag{SCD}$$

It follows that $V_j(b) = \mathcal{V}(y_j, k_j, b)$ also satisfies (SCD): for all $b' > b$ and $(j, j') \in \mathcal{J}(b) \cap \mathcal{J}(b') \times \mathcal{J}(b) \cap \mathcal{J}(b')$ such that $j' > j$, $V_{j'}(b) \geq (>) V_j(b)$ implies $V_{j'}(b') \geq (>) V_j(b')$. Moreover, $\mathcal{J}(b)$ is increasing in b in the sense that $\mathcal{J}(b')$ dominates $\mathcal{J}(b)$ in the strong set order whenever $b' > b$. Therefore, by Milgrom and Shannon (1994), $\text{argmax}_{j \in \mathcal{J}(b)} V_j(b)$ is increasing in b.

B.2 Comparison with Bénabou (2002)'s Criterion of Efficiency

Consider a continuum of individuals, indexed by $i \in [0, 1]$. Let $c_t(i)$ for $t \in \{1, 2\}$ be the consumption of individual i in period t. Then, aggregating (5) across individuals gives a monetary measure of aggregate efficiency:

$$\int_0^1 [c_1(i) + c_2(i)]\, di - \mathcal{T}_c, \tag{33}$$

where the first term is aggregate lifetime consumption and the second term is aggregate intertemporal consumption distortion, defined as follows:

$$\mathcal{T}_c \equiv \int_0^1 [c_1(i) + c_2(i)]\, di - \int_0^1 2[c_1(i)c_2(i)]^{\frac{1}{2}} di.$$

Next, I construct Bénabou (2002)'s criterion of aggregate efficiency. The first step is to replace stochastic consumption in each period with certainty-equivalents, which can be omitted since there is no uncertainty in this environment. The second step is to linearly aggregate individuals' consumptions (i.e., $\int_0^1 c_t(i) di$). Finally, the sequence of aggregate consumptions is evaluated through individuals' common utility function:

$$\ln\left(\int_0^1 c_1(i)di\right) + \ln\left(\int_0^1 c_2(i)di\right). \tag{34}$$

To compare (34) with my monetary measure of efficiency (33), I transform (34) into monetary units through the expenditure function:

$$\int_0^1 [c_1(i) + c_2(i)]di - \tilde{\mathcal{T}}_c, \tag{35}$$

where

$$\tilde{\mathcal{T}}_c \equiv \int_0^1 [c_1(i) + c_2(i)]di - 2\left[\left(\int_0^1 c_1(i)di\right)\left(\int_0^1 c_2(i)di\right)\right]^{\frac{1}{2}}.$$

While (35) is similar to (33), its definition of aggregate intertemporal distortion is different: rather than aggregating individuals' willingness to pay to eliminate the consumption fluctuations, (35) evaluates the distortion from the perspective of a representative agent facing a sequence of aggregate consumptions in each period. In this way, the magnitude of the intertemporal distortion is understated in (35):

$$\mathcal{T}_c - \tilde{\mathcal{T}}_c = 2\left[\left(\int_0^1 c_1(i)di\right)\left(\int_0^1 c_2(i)di\right)\right]^{\frac{1}{2}} - \int_0^1 2[c_1(i)c_2(i)]^{\frac{1}{2}}di \geq 0,$$

which holds because of Jensen's inequality.

B.3 Proof of Lemma 2

For all $j \in \mathcal{J}(b)$, we have

$$V_j(b) = \begin{cases} 2[(b-k_j)y_j]^{\frac{1}{2}}, & \text{for } b < k_j + y_j, \\ b - k_j + y_j, & \text{for } b \geq k_j + y_j. \end{cases}$$

Therefore,

$$V_j'(b) = \max\left\{\left(\frac{y_j}{b-k_j}\right)^{\frac{1}{2}}, 1\right\} \geq 1,$$

where the inequality is strict if and only if $b < k_j + y_j$. It is easy to see that $V_j'(b)$ is decreasing in b and increasing in j. Moreover, $V_j'(b)$ is strictly decreasing in b and strictly increasing in j if and only if $b < k_j + y_j$.

B.4 Proof of Lemma 3

Consider a function $\hat{j} : \mathbb{R}_+ \to \mathcal{J}$ such that $\hat{j}(b) \in \text{argmax}_{j \in \mathcal{J}(b)} V_j(b)$ for all $b \in \mathbb{R}_+$. By the envelope theorem of Milgrom and Segal (2002), for $b'' > b'$,

$$V(b'') - V(b') = \int_{b'}^{b''} V'_{\hat{j}(b)}(b) db.$$

Since $V'_j(b) \geq 1$ for all $j \in \mathcal{J}(b)$ by Lemma 2, we have $V(b'') - V(b') \geq b'' - b'$. Therefore, $\tau(b'') \leq \tau(b')$ follows.

Next, define $\bar{b}$ as follows:

$$\bar{b} \equiv \min_{j \in \text{argmax}_{j' \in \mathcal{J}} \{y_{j'} - k_{j'}\}} \{k_j + y_j\}.$$

Then, for those with $b \geq \bar{b}$, there exists a schooling option that (i) maximizes lifetime consumption and (ii) permits full consumption smoothing.

B.5 Marginal Value of Wealth and Consumption Distortion

Consider an individual with b and a feasible choice $j \in \mathcal{J}(b)$. Let $(\hat{c}_1, \hat{c}_2)$ be the solution to the following problem:

$$V_j(b) \equiv \max_{(c_1, c_2) \in \mathbb{R}_+^2} \left\{ 2(c_1 c_2)^{1/2} | c_1 + c_2 \leq b - k_j + y_j, c_1 \leq b - k_j \right\}.$$

Then the marginal value of wealth is $V'_j(b) = (\hat{c}_2/\hat{c}_1)^{1/2}$.

In order to eliminate the intertemporal consumption distortion, the individuals is willing to pay as much as $\hat{c}_1 + \hat{c}_2 - 2(\hat{c}_1\hat{c}_2)^{1/2}$, which can be expressed as a fraction of lifetime consumption:

$$\frac{\hat{c}_1 + \hat{c}_2 - 2(\hat{c}_1\hat{c}_2)^{1/2}}{\hat{c}_1 + \hat{c}_2} = 1 - \frac{2(\hat{c}_2/\hat{c}_1)^{1/2}}{1 + \hat{c}_2/\hat{c}_1} = \frac{[V'_j(b) - 1]^2}{1 + [V'_j(b)]^2},$$

which is increasing in $V'_j(b)$ since $V'_j(b) \geq 1$.

B.6 Optimality of Full Redistribution

Let $\hat{V}(\cdot)$ be the "upper concave envelope" of $V(\cdot)$, the infimum of all concave functions that lie over $V(\cdot)$. Formally, for each $b \in \mathbb{R}_+$,

$$\hat{V}(b) \equiv \inf_{\nu, \alpha} \{\nu | \nu + \alpha(b - b') \geq V(b'), \forall b' \in \mathbb{R}_+\}.$$

Let $\mathrm{E}(\cdot)$ be the expectation operator over the distribution of parental transfers. Since $\hat{V}(\cdot)$ is concave, Jensen's inequality implies $\hat{V}(\mathrm{E}(b)) \geq \mathrm{E}(\hat{V}(b))$. Therefore, the expected value of $\hat{V}(\cdot)$ is maximized by fully redistributing parental transfers. This also means that the full redistribution maximizes the expected value of $V(\cdot)$ (i.e., it minimizes the expected value of $\tau(\cdot)$) if and only if $V(\mathrm{E}(b)) = \hat{V}(\mathrm{E}(b))$. First, suppose that $V(\mathrm{E}(b)) = \hat{V}(\mathrm{E}(b))$. Then $V(\mathrm{E}(b)) = \hat{V}(\mathrm{E}(b)) \geq \mathrm{E}(\hat{V}(b)) \geq \mathrm{E}(V(b))$ holds because $\hat{V}(b) \geq V(b)$ for all $b \in \mathbb{R}_+$. Next, suppose that $V(\mathrm{E}(b)) < \hat{V}(\mathrm{E}(b))$. Then there exists a distribution of parental transfers that gives an expected value of $V(\cdot)$ that is strictly higher than $V(\mathrm{E}(b))$. Thus, the full redistribution is not optimal.

B.7 Updated Definitions

$$
\begin{aligned}
\mathcal{J}(b; \boldsymbol{g}, \boldsymbol{y}) &\equiv \left\{ j \in \mathcal{J} | k(y_j) - g_j \leq b \right\}, \\
\tau_c(b; \boldsymbol{g}, \boldsymbol{y}) &\equiv \sum_{j \in \mathcal{J}(b; \boldsymbol{g}, \boldsymbol{y})} p_j(b) \left[(y_j - k(y_j) + b + g_j) - V_j(b + g_j; y_j) \right], \\
\tau_s(b; \boldsymbol{g}, \boldsymbol{y}) &\equiv \max_{y \in \mathbb{R}_+} \{ y - k(y) \} - \sum_{j \in \mathcal{J}} p_j(b)(y_j - k(y_j)), \\
\tau(b; \boldsymbol{g}, \boldsymbol{y}) &\equiv \tau_c(b; \boldsymbol{g}, \boldsymbol{y}) + \tau_s(b; \boldsymbol{g}, \boldsymbol{y}), \\
V_j(b + g_j; y_j) &\equiv \max_{(c_1, c_2) \in \mathbb{R}_+^2} \left\{ 2(c_1 c_2)^{1/2} | c_1 + c_2 \leq b - k(y_j) + y_j, c_1 \leq b - k(y_j) \right\}.
\end{aligned}
$$

B.8 Proof of Proposition 1

First, notice that Problem 1 becomes trivial when there are enough resources to fully eliminate all distortions. For example, when $\min_{i \in \mathcal{I}} \{b_i\} + G \geq k(y^*) + y^*$, setting $\hat{y}_{\hat{j}} = y^*$ for some $\hat{j} \in \mathcal{J}$ and setting $\hat{g}_j = G$ for all $j \in \mathcal{J}$ would induce $\hat{p}_{\hat{j}}(b_i) = 1$ for all $i \in \mathcal{I}$ and achieve zero aggregate distortions. In this case, the social marginal value of budget, represented by the Lagrangian multiplier on the social planner's budget constraint (7), is zero because an additional budget does not affect the value of the social objective function. Therefore, for the rest of this section, I consider the case where the Lagrangian multiplier on (7) is strictly positive.

I make an assumption, which will be verified later, that each individual chooses only one schooling option. Therefore, I solve Problem 1 with the following additional constraints:

$$
p_j(b_i) \in \{0, 1\}, \quad \forall (i, j) \in \mathcal{I} \times \mathcal{J}.
$$

By the revelation principle, Problem 1 can be alternatively formulated as designing a direct revelation mechanism where both schooling and subsidy are determined based on individuals' truthful reports about their parental transfers (or their type), an approach taken by Mirrlees (1971). The social planner lets an individual attain a certain level of schooling—or, equivalently, earnings $y \in \mathbb{R}_+$ associated with that schooling—in return for a net payment, $n \in \mathbb{R}$. Therefore, the difference between the schooling cost and net payment, $k(y) - n$, is the amount of subsidy. Then, as defined in Appendix B.1, $\mathcal{V}(y, n, b)$ for $(y, n, b) \in \Theta$ is the money-metric indirect utility of (y, n) for individuals with a parental transfer $b \in \mathbb{R}_+$, which satisfies (MS).

The condition (MS) implies that those with different amounts of parental transfers have different preferences over (y, n), as the slope of the indifference curve is increasing in b. However, as can be seen above, the slope of the indifference curve is strictly increasing in b when the borrowing constraint binds, while it stays constant otherwise. This suggests that all unconstrained individuals have identical preferences over (y, n), behaving as if they are identical types. As discussed by Hellwig (2010), the fact that (MS) may hold only weakly makes this problem different from standard mechanism design problems.

Let $\mathcal{T}(y, n, b)$ be the distortion for $(y, n, b) \in \Theta$. The intertemporal consumption distortion is $\mathcal{T}_c(y, n, b) \equiv y - n + b - \mathcal{V}(y, n, b)$ and the schooling distortion is $\mathcal{T}_s(y, n, b) \equiv [y^* - k(y^*)] - [y - k(y)]$, where $y^* \in \operatorname{argmax}_{y \in \mathbb{R}_+}\{y - k(y)\}$. Therefore, $\mathcal{T}(y, n, b) = y^* - k(y^*) + k(y) - n + b - \mathcal{V}(y, n, b)$.

A direct mechanism consists of a message space $\mathcal{I}$ and an allocation function $(y, n) : \mathcal{I} \to \mathbb{R}_+ \times \mathbb{R}$ that assigns an allocation (y_i, n_i) to those reporting to have parental transfer b_i. Without loss of generality, assume that $\mathcal{I} = \{1, 2, \ldots, I\}$ such that $I \geq 2$ and $b_{i-1} < b_i$ for all $i \in \mathcal{I} \setminus \{1\}$.

Problem 6.

$$\min_{(y_i, n_i)_{i \in \mathcal{I}}} \sum_{i \in \mathcal{I}} f_i \mathcal{T}(y_i, n_i, b_i)$$

$$\textit{subject to} \sum_{i \in \mathcal{I}} f_i [k(y_i) - n_i] \leq G, \tag{RC}$$

$$\mathcal{V}(y_i, n_i, b_i) \geq \mathcal{V}(y_{i'}, n_{i'}, b_i), \quad \forall (i, i') \in \mathcal{I} \times \mathcal{I} \textit{ such that } n_{i'} \leq b_i, \tag{IC}$$

$$y_i \geq 0, \quad \forall i \in \mathcal{I}, \tag{36}$$

$$n_i \leq b_i, \quad \forall i \in \mathcal{I}. \tag{37}$$

Since parental transfers cannot be publicly observed, the allocation function must be incentive compatible in the sense that individuals have no incentives to lie about their type, which is imposed by (IC). For $i \neq i'$, (IC) can be broken down into downward and upward incentive compatibility constraints:

$$\mathcal{V}(y_i, n_i, b_i) \geq \mathcal{V}(y_{i'}, n_{i'}, b_i), \quad \forall (i, i') \in \mathcal{I} \times \mathcal{I} \text{ such that } i > i', \tag{DIC}$$

$$\mathcal{V}(y_i, n_i, b_i) \geq \mathcal{V}(y_{i'}, n_{i'}, b_i), \quad \forall (i, i') \in \mathcal{I} \times \mathcal{I} \text{ such that } i < i' \text{ and } n_{i'} \leq b_i. \tag{UIC}$$

Instead of Problem 6, I consider a problem where (UIC) is replaced by (MC):

Problem 7.

$$\min_{(y_i, n_i)_{i \in \mathcal{I}}} \sum_{i \in \mathcal{I}} f_i \mathcal{T}(y_i, n_i, b_i)$$

subject to (RC), (DIC), (36), (37), *and*

$$y_i \geq y_{i'}, \quad \forall (i, i') \in \mathcal{I} \times \mathcal{I} \textit{ such that } i \geq i'. \tag{MC}$$

The next two lemmas show that Problem 7 is identical to Problem 6.

Lemma 5. *A solution to Problem 6 satisfies* (MC).

Proof. See Appendix B.9. □

Lemma 6. *A solution to Problem 7 satisfies* (UIC).

Proof. See Appendix B.10. □

In the remainder of this subsection, let $(\hat{y}_i, \hat{n}_i)_{i \in \mathcal{I}}$ be a solution to Problem 7. The following lemma shows its properties.

Lemma 7. *For all $i \in \mathcal{I} \setminus \{1\}$, $\hat{n}_i \geq \hat{n}_{i-1}$ and $k(\hat{y}_i) - \hat{n}_i \leq k(\hat{y}_{i-1}) - \hat{n}_{i-1}$, where the inequalities are strict if and only if $\hat{y}_i > \hat{y}_{i-1}$.*

Proof. See Appendix B.11. □

It remains to verify the guess that all individuals of identical type receive identical (y, n).

Lemma 8. *There does not exist $(y', n', i') \in \mathbb{R}_+ \times \mathbb{R} \times \mathcal{I}$ such that (i) $y' \neq \hat{y}_{i'}$ and $n' \neq \hat{n}_{i'}$, (ii) $\mathcal{V}(y', n', b_{i'}) \geq \mathcal{V}(\hat{y}_{i'}, \hat{n}_{i'}, b_{i'})$, (iii) $k(y') - n' \leq k(\hat{y}_{i'}) - \hat{n}_{i'}$, and (iv) $\mathcal{V}(\hat{y}_i, \hat{n}_i, b_i) \geq \mathcal{V}(y', n', b_i)$ for all $i \in \mathcal{I}$ such that $n' \leq b_i$.*

Proof. See Appendix B.12. □

Therefore, $(\hat{y}_i, \hat{n}_i)_{i \in \mathcal{I}}$ strictly dominates any other allocations that assign different pairs of (y, n) to some individuals of identical type. This proves that Problem 1 is identical to Problem 6.

Finally, I construct $(\hat{\boldsymbol{y}}, \hat{\boldsymbol{g}}, (\hat{\boldsymbol{p}}(b_i))_{i \in \mathcal{I}})$ from $(\hat{y}_i, \hat{n}_i)_{i \in \mathcal{I}}$ as follows. Define the set of types with distinct schooling choices $\hat{\mathcal{J}} \equiv \{1\} \cup \{i \in \mathcal{I} \setminus \{1\} | \hat{y}_{i-1} < \hat{y}_i\}$. Then $\hat{y}_j$ is well defined for all $j \in \hat{\mathcal{J}}$. Similarly, define $\hat{g}_j \equiv k(\hat{y}_j) - \hat{n}_j$ for all $j \in \hat{\mathcal{J}}$. Finally, for all $(i, j) \in \mathcal{I} \times \hat{\mathcal{J}}$, set $\hat{p}_j(b_i) = \mathbb{I}_{\hat{y}_j = \hat{y}_i}$. Then it is easy to see that Proposition 1 follows.

B.9 Proof of Lemma 5

I begin with two implications of (DIC).

Claim 1. (DIC) *implies $\mathcal{V}(y_i, n_i, b_i) > \mathcal{V}(y_{i-1}, n_{i-1}, b_{i-1})$ for all $i \in \mathcal{I} \setminus \{1\}$.*

Proof. Notice that

$$\mathcal{V}(y_i, n_i, b_i) \geq \mathcal{V}(y_{i-1}, n_{i-1}, b_i) > \mathcal{V}(y_{i-1}, n_{i-1}, b_{i-1}),$$

where the first inequality holds due to (DIC) for $i \in \mathcal{I} \setminus \{1\}$ and $i' = i - 1$ and the second inequality follows because $\mathcal{V}(y, n, b)$ is strictly increasing in b. □

Claim 2. (DIC) *implies $n_i < b_i$ for all $i \in \mathcal{I} \setminus \{1\}$.*

Proof. First, notice that (37) implies $\mathcal{V}(y_i, n_i, b_i) \geq 0$ for all $i \in \mathcal{I}$. Next, suppose that $n_i = b_i$ for some $i \in \mathcal{I} \setminus \{1\}$. Then $\mathcal{V}(y_i, n_i, b_i) = 0$. Therefore, by Claim 1, $\mathcal{V}(y_i, n_i, b_i) = 0 > \mathcal{V}(y_{i'}, n_{i'}, b_{i'})$ for all $i' < i$, leading to a contradiction. □

In the remainder of this subsection, let $(y_i, n_i)_{i \in \mathcal{I}}$ be the solution to Problem 6.

Claim 3. *If $y_i = y_{i'}$ for some $(i, i') \in \mathcal{I} \times \mathcal{I}$, then $n_i = n_{i'}$.*

Proof. Suppose that $y_i = y_{i'}$. Without loss of generality, consider $i \leq i'$. (IC) for (i', i) implies $n_i \geq n_{i'}$. Since it also implies $b_i \geq n_{i'}$, (IC) for (i, i') gives $n_i \leq n_{i'}$. Therefore, $n_i = n_{i'}$ holds. □

Claim 4. *If $y_i > y_{i'}$ for some $(i, i') \in \mathcal{I} \times \mathcal{I}$ such that $i < i'$, then $n_i > n_{i'}$.*

Proof. Suppose that $y_i > y_{i'}$ for some $i' > i$. Since $i' > i \geq 1$, $b_{i'} > n_{i'}$ holds by Claim 2. Therefore, $y_i > y_{i'}$ and $n_i \leq n_{i'}$ imply $\mathcal{V}(y_{i'}, n_{i'}, b_{i'}) < \mathcal{V}(y_i, n_i, b_{i'})$, violating (IC) for (i', i). □

Claim 5. *$y_i - n_i = y_{i'} - n_{i'}$ for all $(i, i') \in \mathcal{I} \times \mathcal{I}$ such that* $\min\{b_i, b_{i'}\} \geq \max\{n_i + y_i, n_{i'} + y_{i'}\}$.

Proof. If $\min\{b_i, b_{i'}\} \geq \max\{n_i + y_i, n_{i'} + y_{i'}\}$, then $\mathcal{V}(y_i, n_i, b) = y_i - n_i + b$ and $\mathcal{V}(y_{i'}, n_{i'}, b) = y_{i'} - n_{i'} + b$ for all $b \in \{b_i, b_{i'}\}$. Therefore, (IC) for (i, i') and (i', i) imply $y_i - n_i \geq y_{i'} - n_{i'}$ and $y_i - n_i \leq y_{i'} - n_{i'}$, respectively. □

Claim 6. *$y_i \geq y_{i'}$ for all $(i, i') \in \mathcal{I} \times \mathcal{I}$ such that $i \geq i'$.*

Proof. Suppose that there exists $(i_0, i_1) \in \mathcal{I} \times \mathcal{I}$ such that $y_{i_0} > y_{i_1}$ and $i_0 < i_1$. Then, by Claim 4, $n_{i_1} < n_{i_0}$. Due to $y_{i_0} > y_{i_1}$ and (SCD), (IC) for (i_1, i_0) implies $\mathcal{V}(y_{i_1}, n_{i_1}, b_i) \geq \mathcal{V}(y_{i_0}, n_{i_0}, b_i)$ for all $i \leq i_1$ and (IC) for (i_1, i_0) implies $\mathcal{V}(y_{i_1}, n_{i_1}, b_i) \geq \mathcal{V}(y_{i_0}, n_{i_0}, b_i)$ for all $i \leq i_1$. Thus, $\mathcal{V}(y_{i_1}, n_{i_1}, b_i) = \mathcal{V}(y_{i_0}, n_{i_0}, b_i)$ holds for all i such that $i_0 \leq i \leq i_1$.

When $b < n + y$, the borrowing constraint binds and individuals have different preferences over (y, n) in the sense that $\mathcal{V}_y(y, n, b)/|\mathcal{V}_n(y, n, b)|$ is strictly increasing in b. Therefore, $\mathcal{V}(y_{i_1}, n_{i_1}, b_{i_0}) = \mathcal{V}(y_{i_0}, n_{i_0}, b_{i_0})$ and $\mathcal{V}(y_{i_1}, n_{i_1}, b_{i_1}) = \mathcal{V}(y_{i_0}, n_{i_0}, b_{i_1})$ suggest that $\min\{b_{i_0}, b_{i_1}\} \geq \max\{n_{i_0} + y_{i_0}, n_{i_1} + y_{i_1}\}$ must hold. Then, by Claim 5, we have $y_{i_0} - n_{i_0} = y_{i_1} - n_{i_1}$.

Next, let $(\underline{y}, \underline{n})$ be a solution to the following problem:

$$\min_{(y,n) \in \mathbb{R}_+ \times \mathbb{R}} \left\{ k(y) - n \middle| y - n = y_{i_0} - n_{i_0}, y \in [y_{i_1}, y_{i_0}] \right\}.$$

Since $k(\cdot)$ is strictly increasing and strictly convex, $(\underline{y}, \underline{n})$ is unique. Moreover, $\underline{y} - \underline{n} = y_{i_0} - n_{i_0}$ and $\underline{y} \leq y_{i_0}$ suggest $\underline{n} \leq n_{i_0}$. Therefore, $\underline{n} + \underline{y} \leq n_{i_0} + y_{i_0} \leq \min\{b_{i_0}, b_{i_1}\}$ and $\mathcal{V}(\underline{y}, \underline{n}, b_i) = \mathcal{V}(y_i, n_i, b_i)$ for all $i \in \{i_0, i_1\}$.

Then we can construct an alternative allocation $(\tilde{y}_i, \tilde{n}_i)_{i \in \mathcal{I}}$ such that (i) $(\tilde{y}_i, \tilde{n}_i) = (y_i, n_i)$ for all $i \in \mathcal{I} \setminus \{i_0, i_1\}$ and (ii) $(\tilde{y}_i, \tilde{n}_i) = (\underline{y}, \underline{n})$ for $i \in \{i_0, i_1\}$. Because $(\underline{y}, \underline{n})$ is unique and $y_{i_0} neq y_{i_1}$, we have

$$f_{i_0}[k(\tilde{y}_{i_0}) - \tilde{n}_{i_0}] + f_{i_1}[k(\tilde{y}_{i_1}) - \tilde{n}_{i_1}] < f_{i_0}[k(y_{i_0}) - n_{i_0}] + f_{i_1}[k(y_{i_1}) - n_{i_1}],$$

implying that the alternative allocation $(\tilde{y}_i, \tilde{n}_i)_{i \in \mathcal{I}}$ reduces aggregate spending. Moreover, since $\mathcal{V}(\tilde{y}_i, \tilde{n}_i, b_i) = \mathcal{V}(y_i, n_i, b_i)$ for all $i \in \mathcal{I}$, (10) suggests that the alternative allocation also reduces aggregate distortions.

Next, I show that $(\tilde{y}_i, \tilde{n}_i)_{i \in \mathcal{I}}$ satisfies (IC). Since the case $\underline{y} \in \{y_{i_0}, y_{i_1}\}$ is trivial, consider the case $\underline{y} \in (y_{i_1}, y_{i_0})$. First, consider the following set of constraints:

$$\mathcal{V}(\tilde{y}_i, \tilde{n}_i, b_i) \geq \mathcal{V}(\tilde{y}_{i'}, \tilde{n}_{i'}, b_i), \quad \forall (i, i') \in \{i_0, i_1\} \times \mathcal{I} \setminus \{i_0, i_1\} \text{ such that } \tilde{n}_{i'} \leq b_i.$$

These are satisfied because $(\tilde{y}_{i'}, \tilde{n}_{i'}) = (y_{i'}, n_{i'})$ for all $i' \in \mathcal{I} \setminus \{i_0, i_1\}$ implies that $\mathcal{V}(\tilde{y}_{i'}, \tilde{n}_{i'}, b_i) = \mathcal{V}(y_{i'}, n_{i'}, b_i)$ for all $(i, i') \in \{i_0, i_1\} \times \mathcal{I} \setminus \{i_0, i_1\}$, and $(y_i, n_i)_{i\in\mathcal{I}}$ satisfies (IC).

Next, consider the following set of constraints:

$$\mathcal{V}(\tilde{y}_i, \tilde{n}_i, b_i) \geq \mathcal{V}(\tilde{y}_{i'}, \tilde{n}_{i'}, b_i), \quad \forall (i, i') \in \mathcal{I} \setminus \{i_0, i_1\} \times \{i_0, i_1\} \text{ such that } \tilde{n}_{i'} \leq b_i.$$

Since $b_{i_0} \geq \max\{n_{i_0} + y_{i_0}, n_{i_1} + y_{i_1}\} \geq \underline{n} + \underline{y} = \tilde{n}_{i_0} + \tilde{y}_{i_0} = \tilde{n}_{i_1} + \tilde{y}_{i_1}$, $\mathcal{V}(\tilde{y}_{i'}, \tilde{n}_{i'}, b_i) = \mathcal{V}(y_{i'}, n_{i'}, b_i) = \underline{y} - \underline{n} + b_i$ for all $i \geq i_0$ and $i' \in \{i_0, i_1\}$. Next, by $\tilde{y}_{i_1} > y_{i_1}$ and (SCD), $\mathcal{V}(\tilde{y}_{i_1}, \tilde{n}_{i_1}, b_{i_0}) = \mathcal{V}(y_{i_1}, n_{i_1}, b_{i_0})$ implies $\mathcal{V}(\tilde{y}_{i_1}, \tilde{n}_{i_1}, b_i) \leq \mathcal{V}(y_{i_1}, n_{i_1}, b_i)$ for all $i \leq i_0$. Therefore, $\mathcal{V}(\tilde{y}_{i_0}, \tilde{n}_{i_0}, b_i) = \mathcal{V}(\tilde{y}_{i_1}, \tilde{n}_{i_1}, b_i) \leq \mathcal{V}(y_{i_1}, n_{i_1}, b_i) \leq \mathcal{V}(y_i, n_i, b_i) = \mathcal{V}(\tilde{y}_i, \tilde{n}_i, b_i)$ holds for all $i \in \mathcal{I} \setminus \{i_0, i_1\}$ such that $\tilde{n}_{i_1} \leq b_i$.

Therefore, $(\tilde{y}_i, \tilde{n}_i)_{i\in\mathcal{I}}$ satisfies (RC), (IC), and (36), and it achieves the aggregate distortion that is strictly lower than that under $(y_i, n_i)_{i\in\mathcal{I}}$. This contradicts the assumption that $(y_i, n_i)_{i\in\mathcal{I}}$ solves Problem 6. □

B.10 Proof of Lemma 6

(MC) implies that the following adjacent downward incentive compatibility constraints are sufficient for (DIC):

$$\mathcal{V}(y_i, n_i, b_i) \geq \mathcal{V}(y_{i-1}, n_{i-1}, b_i), \quad \forall i \in \mathcal{I} \setminus \{1\}. \tag{ADIC}$$

Claim 7. (MC) *and* (ADIC) *imply* (DIC).

Proof. Suppose that $(y_i, n_i)_{i\in\mathcal{I}}$ satisfies (MC) and (ADIC). Consider $i \in \mathcal{I} \setminus \{1\}$ and note that (ADIC) implies that (DIC) for $(i, i-1)$ holds. For induction, suppose that (DIC) for (i, i') holds for some $i' \leq i - 1$ (i.e., $\mathcal{V}(y_i, n_i, b_i) \geq \mathcal{V}(y_{i'}, n_{i'}, b_i)$) and consider the validity of (DIC) for $(i, i'-1)$. By (ADIC), we have $\mathcal{V}(y_{i'-1}, n_{i'-1}, b_{i'}) \leq \mathcal{V}(y_{i'}, n_{i'}, b_{i'})$, where $y_{i'-1} \leq y_{i'}$ holds due to (MC). If $y_{i'-1} = y_{i'}$, then $\mathcal{V}(y_{i'-1}, n_{i'-1}, b_{i'}) \leq \mathcal{V}(y_{i'}, n_{i'}, b_{i'})$ implies $n_{i'-1} \geq n_{i'}$, so $\mathcal{V}(y_{i'-1}, n_{i'-1}, b_i) \leq \mathcal{V}(y_{i'}, n_{i'}, b_i)$ holds. If $y_{i'-1} < y_{i'}$, then $\mathcal{V}(y_{i'-1}, n_{i'-1}, b_i) \leq \mathcal{V}(y_{i'}, n_{i'}, b_i)$ is implied by (SCD). From $\mathcal{V}(y_{i'-1}, n_{i'-1}, b_i) \leq \mathcal{V}(y_{i'}, n_{i'}, b_i)$ and $\mathcal{V}(y_{i'}, n_{i'}, b_i) \leq \mathcal{V}(y_i, n_i, b_i)$, we have $\mathcal{V}(y_{i'-1}, n_{i'-1}, b_i) \leq \mathcal{V}(y_i, n_i, b_i)$. Therefore, (DIC) for $(i, i'-1)$ holds. The induction is thus complete. □

Therefore, (DIC) in Problem 7 can be replaced by (ADIC). (MC) also implies that it suffices to consider only $y_1 \geq 0$ instead of (36). Moreover, due to Claim 2, (37) can be reduced to a single constraint $n_1 \leq b_1$. With reduced constraints, the Lagrangian for Problem 7 is

$$\sum_{i\in\mathcal{I}} f_i \Big\{ \mathcal{V}(y_i, n_i, b_i) - (1+\lambda)[k(y_i) - n_i] \Big\} + \lambda G - \sum_{i\in\mathcal{I}} f_i [b_i + y^* - k(y^*)]$$
$$+ \sum_{i\in\mathcal{I}\setminus\{1\}} \Big\{ \psi_i [\mathcal{V}(y_i, n_i, b_i) - \mathcal{V}(y_{i-1}, n_{i-1}, b_i)] + \varphi_i (y_i - y_{i-1}) \Big\} + \zeta(b_1 - n_1) + \xi y_1,$$

where λ, φ_i, ψ_i, ζ, and ξ are Lagrangian multipliers on (RC), (MC), (ADIC), $n_1 \leq b_1$, and $y_1 \geq 0$, respectively.

In the remainder of this subsection, let $(\hat{y}_i, \hat{n}_i)_{i\in\mathcal{I}}$ be a solution to Problem 7. The first order conditions are, for all $i \in \mathcal{I}$,

$$(f_i + \psi_i)\mathcal{V}_y(\hat{y}_i, \hat{n}_i, b_i) - (1+\lambda)f_i k'(\hat{y}_i) - \psi_{i+1}\mathcal{V}_y(\hat{y}_i, \hat{n}_i, b_{i+1}) \leq \varphi_{i+1} - \varphi_i, \tag{38}$$

$$(f_i + \psi_i)|\mathcal{V}_n(\hat{y}_i, \hat{n}_i, b_i)| - (1+\lambda)f_i - \psi_{i+1}|\mathcal{V}_n(\hat{y}_i, \hat{n}_i, b_{i+1})| \leq 0, \tag{39}$$

where $\psi_i = \varphi_i = 0$ for $i = 1$ and $\psi_{i+1} = \varphi_{i+1} = 0$ for $i = I$. (38) holds as equality if $i > 1$ or $\hat{y}_1 > 0$, while (39) holds as equality if $i > 1$ or $\hat{n}_1 < b_1$. When (39) holds as equality, we can combine (38) and (39) to get

$$\begin{aligned}\varphi_{i+1} - \varphi_i \geq &(1+\lambda)f_i\left(\frac{\mathcal{V}_y(\hat{y}_i, \hat{n}_i, b_i)}{|\mathcal{V}_n(\hat{y}_i, \hat{n}_i, b_i)|} - k'(\hat{y}_i)\right)\\ &+ \psi_{i+1}\left(\frac{\mathcal{V}_y(\hat{y}_i, \hat{n}_i, b_i)}{|\mathcal{V}_n(\hat{y}_i, \hat{n}_i, b_i)|}|\mathcal{V}_n(\hat{y}_i, \hat{n}_i, b_{i+1})| - \mathcal{V}_y(\hat{y}_i, \hat{n}_i, b_{i+1})\right).\end{aligned} \tag{40}$$

I first show that it is not possible that only one of $\hat{y}_1 \geq 0$ and $\hat{n}_1 \leq b_1$ holds as an equality.

Claim 8. *If $\hat{n}_1 < b_1$, then $\hat{y}_1 > 0$.*

Proof. Suppose that $\hat{n}_1 < b_1$ and $\hat{y}_1 = 0$. Then $\min\{b_1, b_2\} > \hat{n}_1 + \hat{y}_1 = \hat{n}_1$. Therefore, $\mathcal{V}_y(\hat{y}_1, \hat{n}_1, b_i) = |\mathcal{V}_n(\hat{y}_1, \hat{n}_1, b_i)| = 1$ for $i \in \{1, 2\}$. Then, also using $k'(\hat{y}_1) = 0$, the first order conditions for $i = 1$ can be written as follows:

$$\begin{aligned}f_1 - \psi_2 \leq &\varphi_2,\\ -\lambda f_1 = &\psi_2.\end{aligned}$$

Since $f_1 > 0$, $\lambda \geq 0$, and $\psi_2 \geq 0$, we have $\lambda = \psi_2 = 0$ from the second equation. Then the first equation gives $f_1 \leq \varphi_2$, which implies $\varphi_2 > 0$, and thus $\hat{y}_2 = \hat{y}_1 = 0$.

Since $\hat{y}_2 = \hat{y}_1$ and (ADIC) for $i = 2$ imply $\hat{n}_2 \leq \hat{n}_1$, $\min\{b_2, b_3\} \geq \hat{n}_2 + \hat{y}_2$ holds. Therefore, the first order conditions for $i = 2$ are

$$\begin{aligned}f_2 - \psi_3 = &\varphi_3 - \varphi_2,\\ -\lambda f_2 = &\psi_3.\end{aligned}$$

Since $\psi_3 = 0$ from the second equation, the first equation is $f_2 = \varphi_3 - \varphi_2$, which implies $\varphi_3 > \varphi_2 > 0$. Therefore, $\hat{y}_3 = \hat{y}_2 = 0$.

In this way, we can show that $\hat{y}_i = 0$ and $\hat{n}_i \leq \hat{n}_1$ for all $i \in \mathcal{I}$, and $\psi_i = 0$ and $\varphi_i > 0$ for all $i \in \mathcal{I} \setminus \{1\}$. However, the first order condition for y_I is $f_I = -\varphi_I$, which contradicts $f_I > 0$. □

Claim 9. *If $\hat{n}_1 = b_1$, then $\hat{y}_1 = 0$.*

Proof. Suppose that $\hat{n}_1 = b_1$ and $\hat{y}_1 > 0$. Then $\mathcal{V}(\hat{y}_1, \hat{n}_1, b_1) = 0$. Consider an alternative allocation $(\tilde{y}_i, \tilde{n}_i)_{i\in\mathcal{I}}$ such that $(\tilde{y}_i, \tilde{n}_i) = (\hat{y}_i, \hat{n}_i)$ for all $i \in \mathcal{I} \setminus \{1\}$, $\tilde{y}_1 = 0$, and $\tilde{n}_1 = \hat{n}_1$. Then $\mathcal{V}(\tilde{y}_1, \tilde{n}_1, b_1) = \mathcal{V}(\hat{y}_1, \hat{n}_1, b_1) = 0$ and $\mathcal{V}(\tilde{y}_2, \tilde{n}_2, b_2) = \mathcal{V}(\hat{y}_2, \hat{n}_2, b_2) \geq \mathcal{V}(\hat{y}_1, \hat{n}_1, b_2) > \mathcal{V}(\tilde{y}_1, \tilde{n}_1, b_2)$, where the first inequality holds due to (ADIC) for $i = 2$ and the second inequality is because of $b_2 > b_1 = \hat{n}_1 = \tilde{n}_1$ and $\hat{y}_1 > \tilde{y}_1$. However, $k(\tilde{y}_1) - \tilde{n}_1 < k(\hat{y}_1) - \hat{n}_1$. Therefore, $(\tilde{y}_i, \tilde{n}_i)_{i\in\mathcal{I}}$ satisfies all constraints and achieves lower aggregate distortions, contradicting the assumption that $(\hat{y}_i, \hat{n}_i)_{i\in\mathcal{I}}$ is optimal. □

The next two lemmas characterize the nature of distortions induced by private information. The key to this characterization is to determine whether $(\hat{y}_i, \hat{n}_i)$ delivers the utility of type i, $\mathcal{V}(\hat{y}_i, \hat{n}_i, b_i)$, in an efficient (i.e., cost-minimizing) manner. The most efficient way to deliver a given level of utility v to type i is defined as follows: for $v \in \mathbb{R}_+$,

$$(y_i^*(v), n_i^*(v)) \equiv \underset{(y,n)\in\mathbb{R}_+\times\mathbb{R}}{\text{argmin}} \left\{k(y) - n | \mathcal{V}(y, n, b_i) \geq v, n \leq b_i\right\}.$$

Claim 10. *For $v = 0$, $n_i^*(v) = b_i$ and $y_i^*(v) = 0$. For $v \in (0, 2y^*)$, $y_i^*(v) \in (0, y^*)$ and $n_i^*(v) \in (b_i - y_i^*(v), b_i)$. For $v \geq 2y^*$, $y_i^*(v) = y^*$ and $n_i^*(v) = y^* + b_i - v \leq b_i - y_i^*(v)$.*

Proof. Since the case with $v = 0$ is trivial, consider $v > 0$. Since $n_i^*(v) = b_i$ implies $\mathcal{V}(y_i^*(v), n_i^*(v), b_i) = 0 < v$, $n_i^*(v) < b_i$ must hold. In this case, the first order conditions imply

$$\frac{\mathcal{V}_y(y_i^*(v), n_i^*(v), b_i)}{|\mathcal{V}_n(y_i^*(v), n_i^*(v), b_i)|} = k'(y_i^*(v)). \tag{41}$$

Suppose that $n_i^*(v) + y_i^*(v) \leq b_i$. Then $\mathcal{V}(y_i^*(v), n_i^*(v), b_i) = y_i^*(v) - n_i^*(v) + b_i = v$ and (41) implies $y_i^*(v) = y^*$. Moreover, $n_i^*(v) + y_i^*(v) \leq b_i$ also implies $v = y_i^*(v) - n_i^*(v) + b_i \geq y_i^*(v) - [b_i - y_i^*(v)] + b_i = 2y_i^*(v)$. Therefore, for $v < 2y^*$, $y_i^*(v) > b_i - n_i^*(v) > 0$ must hold and (41) gives

$$k'(y_i^*(v)) = \frac{\mathcal{V}_y(y_i^*(v), n_i^*(v), b_i)}{|\mathcal{V}_n(y_i^*(v), n_i^*(v), b_i)|} = \frac{b_i - n_i^*(v)}{y_i^*(v)} < 1 = k'(y^*), \tag{42}$$

which implies $y_i^*(v) < y^*$. □

Claim 11. *$y_i^*(v)$ and $n_i^*(v)$ are continuous in v.*

Proof. First, notice that $y_i^*(v)$ and $n_i^*(v)$ are continuous in v for $v \geq 2y^*$. For $v \in (0, 2y^*)$, $(y_i^*(v), n_i^*(v))$ solves $b_i - n_i^*(v) = k'(y_i^*(v))y_i^*(v)$ and $v = 2\{[b_i - n_i^*(v)]y_i^*(v)\}^{\frac{1}{2}}$. Therefore, $y_i^*(v)$ solves $v = 2y_i^*(v)[k'(y_i^*(v))]^{\frac{1}{2}}$. Since $k'(y)$ is continuous in y, $y_i^*(v)$ and $n_i^*(v) = b_i - k'(y_i^*(v))y_i^*(v)$ are continuous in v over $(0, 2y^*)$. Moreover, $\lim_{v\downarrow 0}(y_i^*(v), n_i^*(v)) = (0, b_i)$ and $\lim_{v\uparrow 2y^*}(y_i^*(v), n_i^*(v)) = (y^*, b_i - y^*)$. □

Let $\hat{v}_i \equiv \mathcal{V}(\hat{y}_i, \hat{n}_i, b_i)$ and consider $i \in \mathcal{I}$ with $\hat{v}_i > 0$. When $\hat{y}_i > y_i^*(\hat{v}_i)$ or $\hat{n}_i > n_i^*(\hat{v}_i)$, both $\hat{y}_i > y_i^*(\hat{v}_i)$ and $\hat{n}_i > n_i^*(\hat{v}_i)$ must hold, because $\mathcal{V}(\hat{y}_i, \hat{n}_i, b_i) = \mathcal{V}(y_i^*(\hat{v}_i), n_i^*(\hat{v}_i), b_i)$ and $\mathcal{V}(y, n, b)$ is strictly increasing in y and strictly decreasing in n. Since $\mathcal{V}_y(y, n, b)/|\mathcal{V}_n(y, n, b)|$ is decreasing in (y, n) and $k'(y)$ is strictly increasing in y, $\hat{y}_i > y_i^*(\hat{v}_i)$ and $\hat{n}_i > n_i^*(\hat{v}_i)$ imply

$$\frac{\mathcal{V}_y(\hat{y}_i, \hat{n}_i, b_i)}{|\mathcal{V}_n(\hat{y}_i, \hat{n}_i, b_i)|} < k'(\hat{y}_i). \tag{43}$$

Similarly, when $\hat{y}_i < y_i^*(\hat{v}_i)$ or $\hat{n}_i < n_i^*(\hat{v}_i)$, both $\hat{y}_i < y_i^*(\hat{v}_i)$ and $\hat{n}_i < n_i^*(\hat{v}_i)$ hold and

$$\frac{\mathcal{V}_y(\hat{y}_i, \hat{n}_i, b_i)}{|\mathcal{V}_n(\hat{y}_i, \hat{n}_i, b_i)|} > k'(\hat{y}_i). \tag{44}$$

Claim 12. *If $\hat{n}_i = b_i$ for some $i \in \mathcal{I}$, then $i = 1$, $\hat{y}_1 = y_1^*(\hat{v}_1) = \hat{v}_1 = 0$, and $\hat{n}_1 = n_1^*(\hat{v}_1) = b_1$.*

Proof. Suppose that $\hat{n}_i = b_i$ for some $i \in \mathcal{I}$. Then $i = 1$ by Claim 2. Moreover, by Claim 9, $\hat{n}_1 = b_1$ implies $\hat{y}_1 = 0$. Since $\hat{n}_1 = b_1$ implies $\hat{v}_1 = 0$, we have $y_1^*(\hat{v}_1) = 0$ and $n_1^*(\hat{v}_1) = b_1$ by Claim 10. Therefore, $\hat{y}_1 = y_1^*(\hat{v}_1) = 0$ and $\hat{n}_1 = n_1^*(\hat{v}_1) = b_1$. □

Claim 13. *$\hat{y}_i \leq y_i^*(\hat{v}_i)$ and $\hat{n}_i \leq n_i^*(\hat{v}_i)$ for all $i \in \mathcal{I}$.*

Proof. By Claim 12, it suffices to only consider the case $\hat{n}_i < b_i$. Suppose that there exists $i \in \mathcal{I}$ such that $\hat{y}_i > y_i^*(\hat{v}_i)$ or $\hat{n}_i > n_i^*(\hat{v}_i)$. Therefore, (43) holds for such $i \in \mathcal{I}$. Define $i' \leq i$ to be the smallest type such that $\hat{y}_{i'} = \hat{y}_i$. Then $\hat{n}_{i'} \geq \hat{n}_i$ holds due to (ADIC) and (43) for i' also holds:

$$\frac{\mathcal{V}_y(\hat{y}_{i'}, \hat{n}_{i'}, b_{i'})}{|\mathcal{V}_n(\hat{y}_{i'}, \hat{n}_{i'}, b_{i'})|} \leq \frac{\mathcal{V}_y(\hat{y}_{i'}, \hat{n}_{i'}, b_i)}{|\mathcal{V}_n(\hat{y}_{i'}, \hat{n}_{i'}, b_i)|} \leq \frac{\mathcal{V}_y(\hat{y}_i, \hat{n}_i, b_i)}{|\mathcal{V}_n(\hat{y}_i, \hat{n}_i, b_i)|} < k'(\hat{y}_i) = k'(\hat{y}_{i'}), \tag{45}$$

where the first inequality holds due to (MS) and the second inequality follows from $\hat{y}_{i'} = \hat{y}_i$ and $\hat{n}_{i'} \geq \hat{n}_i$.

For $i' \in \mathcal{I} \setminus \{I\}$, the first order condition (40) is

$$\begin{aligned}
\varphi_{i'+1} - \varphi_{i'} =& (1+\lambda) f_{i'} \left(\frac{\mathcal{V}_y(\hat{y}_{i'}, \hat{n}_{i'}, b_{i'})}{|\mathcal{V}_n(\hat{y}_{i'}, \hat{n}_{i'}, b_{i'})|} - k'(\hat{y}_{i'}) \right) \\
&+ \psi_{i'+1} \left(\frac{\mathcal{V}_y(\hat{y}_{i'}, \hat{n}_{i'}, b_{i'})}{|\mathcal{V}_n(\hat{y}_{i'}, \hat{n}_{i'}, b_{i'})|} |\mathcal{V}_n(\hat{y}_{i'}, \hat{n}_{i'}, b_{i'+1})| - \mathcal{V}_y(\hat{y}_{i'}, \hat{n}_{i'}, b_{i'+1}) \right) \\
<& \psi_{i'+1} \left(\frac{\mathcal{V}_y(\hat{y}_{i'}, \hat{n}_{i'}, b_{i'})}{|\mathcal{V}_n(\hat{y}_{i'}, \hat{n}_{i'}, b_{i'})|} |\mathcal{V}_n(\hat{y}_{i'}, \hat{n}_{i'}, b_{i'+1})| - \mathcal{V}_y(\hat{y}_{i'}, \hat{n}_{i'}, b_{i'+1}) \right) \\
\leq& 0,
\end{aligned}$$

where the first inequality follows from (45) and the second inequality holds due to (MS). Therefore, $\varphi_{i'+1} < \varphi_{i'}$ holds. However, by the definition of i', $\hat{y}_{i'-1} < \hat{y}_{i'}$, so (MC) does not bind and $\varphi_{i'} = 0$. Since $\varphi_{i'+1} \geq 0$, this leads to a contradiction.

For $i' = I$, $\varphi_{i'+1} = \psi_{i'+1} = 0$. Thus, (40) and (45) imply $\varphi_{i'} > 0$, which contradicts $\hat{y}_{i'-1} < \hat{y}_{i'}$. Therefore, for all $i \in \mathcal{I}$, $\hat{y}_i \leq y_i^*(\hat{v}_i)$ and $\hat{n}_i \leq n_i^*(\hat{v}_i)$ must hold. □

Claim 14. *$\hat{y}_i = y_i^*(\hat{v}_i)$ and $\hat{n}_i = n_i^*(\hat{v}_i)$ for $i = I$ or $i \in \mathcal{I} \setminus \{I\}$ with $\psi_{i+1} = \varphi_{i+1} = 0$.*

Proof. By Claim 12, it suffices to only consider the case $\hat{n}_i < b_i$. For $i = I$ or $i \in \mathcal{I} \setminus \{I\}$ with $\psi_{i+1} = \varphi_{i+1} = 0$, (40) implies

$$-\varphi_i = (1+\lambda) f_i \left(\frac{\mathcal{V}_y(\hat{y}_i, \hat{n}_i, b_i)}{|\mathcal{V}_n(\hat{y}_i, \hat{n}_i, b_i)|} - k'(\hat{y}_i) \right) \geq 0,$$

where the inequality follows from Claim 13. Since $\varphi_i \geq 0$, $\varphi_i = 0$ must hold and the above equation implies $\hat{y}_i = y_i^*(\hat{v}_i)$ and $\hat{n}_i = n_i^*(\hat{v}_i)$. □

Therefore, private information distorts allocation downward from efficiency, except for the highest type. This reflects that individuals have incentives to under-report their type. That is, all constraints of (ADIC) hold as equalities.

Claim 15. *For all $i \in \mathcal{I} \setminus \{1\}$, $\mathcal{V}(\hat{y}_i, \hat{n}_i, b_i) = \mathcal{V}(\hat{y}_{i-1}, \hat{n}_{i-1}, b_i)$.*

Proof. Suppose that there exists $i \in \mathcal{I} \setminus \{1\}$ such that $\mathcal{V}(\hat{y}_i, \hat{n}_i, b_i) > \mathcal{V}(\hat{y}_{i-1}, \hat{n}_{i-1}, b_i)$. Because (ADIC) for i is slack, $\psi_i = 0$. Then (39) for $i-1$ and i can be written as

$$\begin{aligned} |\mathcal{V}_n(\hat{y}_{i-1}, \hat{n}_{i-1}, b_{i-1})| \leq & \frac{(1+\lambda) f_{i-1}}{f_{i-1} + \psi_{i-1}}, \\ |\mathcal{V}_n(\hat{y}_i, \hat{n}_i, b_i)| = & (1+\lambda) + \frac{\psi_{i+1}}{f_i} |\mathcal{V}_n(\hat{y}_i, \hat{n}_i, b_{i+1})|, \end{aligned}$$

which imply

$$|\mathcal{V}_n(\hat{y}_{i-1}, \hat{n}_{i-1}, b_{i-1})| \leq 1 + \lambda \leq |\mathcal{V}_n(\hat{y}_i, \hat{n}_i, b_i)|. \tag{46}$$

Since $\lambda > 0$ and $|\mathcal{V}_n(y, n, b)| = \max\{[y/(b-n)]^{1/2}, 1\}$, we have $|\mathcal{V}_n(\hat{y}_i, \hat{n}_i, b_i)| = [\hat{y}_i/(b_i - \hat{n}_i)]^{1/2} > 1$.

If $\hat{y}_i = \hat{y}_{i-1}$, then $\mathcal{V}(\hat{y}_i, \hat{n}_i, b_i) > \mathcal{V}(\hat{y}_{i-1}, \hat{n}_{i-1}, b_i)$ implies $\hat{n}_i < \hat{n}_{i-1}$. However, $\hat{y}_i = \hat{y}_{i-1}$ and $b_i - \hat{n}_i > b_{i-1} - \hat{n}_{i-1}$ imply $|\mathcal{V}_n(\hat{y}_{i-1}, \hat{n}_{i-1}, b_{i-1})| > |\mathcal{V}_n(\hat{y}_i, \hat{n}_i, b_i)|$, which contradicts (46). Therefore, $\hat{y}_i > \hat{y}_{i-1}$ must hold, which implies $\varphi_i = 0$. Moreover, by Claim 14, $\psi_i = \varphi_i = 0$ implies $\hat{y}_{i-1} = y^*_{i-1}(\hat{v}_{i-1})$ and $\hat{n}_{i-1} = n^*_{i-1}(\hat{v}_{i-1})$.

Suppose that $\hat{n}_{i-1} = b_{i-1}$. By Claim 12, $i = 2$ and $(\hat{y}_1, \hat{n}_1) = (y^*_1(\hat{v}_1), n^*_1(\hat{v}_1)) = (y^*_1(0), n^*_1(0))$. Moreover, $|\mathcal{V}_n(y, n, b)|$ is continuous in (y, n, b), and $y^*_i(v)$ and $n^*_i(v)$ are continuous in v by Claim 11. Therefore,

$$|\mathcal{V}_n(\hat{y}_1, \hat{n}_1, b_1)| = \lim_{v \downarrow 0} |\mathcal{V}_n(y^*_1(v), n^*_1(v), b_1)| = \lim_{v \downarrow 0} \left(\frac{y^*_1(v)}{b_1 - n^*_1(v)} \right)^{\frac{1}{2}} = \lim_{v \downarrow 0} \left(\frac{1}{k'(y^*_1(v))} \right)^{\frac{1}{2}} = \infty,$$

where the third equality is due to (41) and the last equality follows from $\lim_{v \downarrow 0} y^*_1(v) = 0$. Therefore, (46) cannot be satisfied, as $\hat{n}_2 < b_2$ implies $|\mathcal{V}_n(\hat{y}_2, \hat{n}_2, b_2)| = \max\{[\hat{y}_2/(b_2 - \hat{n}_2)]^{1/2}, 1\} < \infty$.

Next, suppose that $\hat{n}_{i-1} < b_{i-1}$. If $b_{i-1} \geq \hat{n}_{i-1} + \hat{y}_{i-1}$, then $\mathcal{V}_y(\hat{y}_{i-1}, \hat{n}_{i-1}, b_{i-1})/|\mathcal{V}_n(\hat{y}_{i-1}, \hat{n}_{i-1}, b_{i-1})| = 1$. Therefore, $\hat{y}_{i-1} = y^*_{i-1}(\hat{v}_{i-1}) = y^*$. However, this contradicts $\hat{y}_{i-1} < \hat{y}_i$ and $\hat{y}_i \leq y^*_i(\hat{v}_i) \leq y^*$. Thus, $b_{i-1} < \hat{n}_{i-1} + \hat{y}_{i-1}$ must hold. In this case, $|\mathcal{V}_n(\hat{y}_{i-1}, \hat{n}_{i-1}, b_{i-1})| > 1$, which also implies $|\mathcal{V}_n(\hat{y}_i, \hat{n}_i, b_i)| > 1$ due to (46). Therefore, $b_i < \hat{n}_i + \hat{y}_i$. Because $b_{i-1} < \hat{n}_{i-1} + \hat{y}_{i-1}$ and $b_i < \hat{n}_i + \hat{y}_i$, (46) can be written as

$$\left(\frac{\hat{y}_{i-1}}{b_{i-1} - \hat{n}_{i-1}} \right)^{\frac{1}{2}} \leq 1 + \lambda \leq \left(\frac{\hat{y}_i}{b_i - \hat{n}_i} \right)^{\frac{1}{2}}. \tag{47}$$

However, $\hat{y}_{i-1} = y^*_{i-1}(\hat{v}_{i-1})$, $\hat{n}_{i-1} = n^*_{i-1}(\hat{v}_{i-1})$, $\hat{y}_i \leq y^*_i(\hat{v}_i)$, $\hat{n}_i \leq n^*_i(\hat{v}_i)$, and $\hat{y}_{i-1} < \hat{y}_i$ imply

$$\frac{\mathcal{V}_y(\hat{y}_{i-1}, \hat{n}_{i-1}, b_{i-1})}{|\mathcal{V}_n(\hat{y}_{i-1}, \hat{n}_{i-1}, b_{i-1})|} = \left(\frac{b_{i-1} - \hat{n}_{i-1}}{\hat{y}_{i-1}} \right)^{\frac{1}{2}} = k'(\hat{y}_{i-1}) < k'(\hat{y}_i) \leq \frac{\mathcal{V}_y(\hat{y}_i, \hat{n}_i, b_i)}{|\mathcal{V}_n(\hat{y}_i, \hat{n}_i, b_i)|} = \left(\frac{b_i - \hat{n}_i}{\hat{y}_i} \right)^{\frac{1}{2}},$$

which contradicts (47). ◻

The fact that the constraints (ADIC) hold as equalities implies that (UIC) is satisfied.

Claim 16. *$\hat{n}_i = \hat{n}_{i-1}$ for all $i \in \mathcal{I} \setminus \{1\}$ such that $\hat{y}_i = \hat{y}_{i-1}$.*

Proof. Suppose that $\hat{y}_i = \hat{y}_{i-1}$ for some $i \in \mathcal{I} \setminus \{1\}$. Since $\mathcal{V}(\hat{y}_i, \hat{n}_i, b_i) = \mathcal{V}(\hat{y}_{i-1}, \hat{n}_{i-1}, b_i)$ holds due to Claim 15 and $\mathcal{V}(y, n, b)$ is strictly decreasing in n, $\hat{n}_i = \hat{n}_{i-1}$ follows. ◻

Claim 17. *$(\hat{y}_i, \hat{n}_i)_{i \in \mathcal{I}}$ satisfies (UIC) .*

Proof. I first show that the following adjacent upward incentive compatibility constraints are satisfied:

$$\mathcal{V}(y_i, n_i, b_i) \geq \mathcal{V}(y_{i+1}, n_{i+1}, b_i), \quad \forall i \in \mathcal{I} \setminus \{I\} \text{ such that } n_{i+1} \leq b_i. \tag{AUIC}$$

Consider $i \in \mathcal{I} \setminus \{I\}$ such that $\hat{n}_{i+1} \leq b_i$. Notice that $\mathcal{V}(\hat{y}_{i+1}, \hat{n}_{i+1}, b_{i+1}) = \mathcal{V}(\hat{y}_i, \hat{n}_i, b_{i+1})$ implies $\mathcal{V}(\hat{y}_i, \hat{n}_i, b_i) \geq \mathcal{V}(\hat{y}_{i+1}, \hat{n}_{i+1}, b_i)$. If $\hat{y}_{i+1} = \hat{y}_i$, then $\hat{n}_{i+1} = \hat{n}_i$ by Claim 16 and $\mathcal{V}(\hat{y}_i, \hat{n}_i, b_i) = \mathcal{V}(\hat{y}_{i+1}, \hat{n}_{i+1}, b_i)$ follows. If $\hat{y}_{i+1} > \hat{y}_i$, then $\mathcal{V}(\hat{y}_i, \hat{n}_i, b_i) \geq \mathcal{V}(\hat{y}_{i+1}, \hat{n}_{i+1}, b_i)$ is implied by (SCD). Therefore, Claim 15 implies (AUIC) is satisfied.

Next, I show that (MC) and (AUIC) imply (UIC). Suppose that $(y_i, n_i)_{i \in \mathcal{I}}$ satisfies (MC) and (AUIC). Consider $i \in \mathcal{I} \setminus \{I\}$ such that $n_{i+1} \leq b_i$ and note that (AUIC) implies that (UIC) for $(i, i+1)$ holds. For induction, suppose that (UIC) for (i, i') holds for some $i' \geq i+1$ such that $n_{i'} \leq b_i$ (i.e., $\mathcal{V}(y_i, n_i, b_i) \geq \mathcal{V}(y_{i'}, n_{i'}, b_i)$) and consider the validity of (UIC) for $(i, i'+1)$ such that $n_{i'+1} \leq b_i$. Note that $n_{i'+1} \leq b_i$ implies $n_{i'+1} \leq b_{i'}$; thus from (AUIC), we have $\mathcal{V}(y_{i'+1}, n_{i'+1}, b_{i'}) \leq \mathcal{V}(y_{i'}, n_{i'}, b_{i'})$, where $y_{i'+1} \geq y_{i'}$ holds due to (MC). If $y_{i'+1} = y_{i'}$, then $\mathcal{V}(y_{i'+1}, n_{i'+1}, b_{i'}) \leq \mathcal{V}(y_{i'}, n_{i'}, b_{i'})$ implies $n_{i'+1} \geq n_{i'}$, so $\mathcal{V}(y_{i'+1}, n_{i'+1}, b_i) \leq \mathcal{V}(y_{i'}, n_{i'}, b_i)$ holds. If $y_{i'+1} > y_{i'}$, then $\mathcal{V}(y_{i'+1}, n_{i'+1}, b_i) \leq \mathcal{V}(y_{i'}, n_{i'}, b_i)$ is implied by (SCD). From $\mathcal{V}(y_{i'+1}, n_{i'+1}, b_i) \leq \mathcal{V}(y_{i'}, n_{i'}, b_i)$ and $\mathcal{V}(y_{i'}, n_{i'}, b_i) \leq \mathcal{V}(y_i, n_i, b_i)$, we have $\mathcal{V}(y_{i'+1}, n_{i'+1}, b_i) \leq \mathcal{V}(y_i, n_i, b_i)$. Therefore, (UIC) for $(i, i'+1)$ holds. The induction is thus complete. ◻

B.11 Proof of Lemma 7

The fact that (ADIC) hold as equalities implies that $\hat{y}_i$ and $\hat{n}_i$ are co-monotonic.

Claim 18. *$\hat{n}_i > \hat{n}_{i-1}$ for all $i \in \mathcal{I} \setminus \{1\}$ such that $\hat{y}_i > \hat{y}_{i-1}$.*

Proof. Suppose that there exists $i \in \mathcal{I} \setminus \{1\}$ such that $\hat{y}_i > \hat{y}_{i-1}$ and $\hat{n}_i \leq \hat{n}_{i-1}$. Since $\mathcal{V}(y, n, b_i)$ is strictly increasing in y (due to $b_i > \hat{n}_i$) and strictly decreasing in n, this implies $\mathcal{V}(\hat{y}_i, \hat{n}_i, b_i) > \mathcal{V}(\hat{y}_{i-1}, \hat{n}_{i-1}, b_i)$, contradicting Claim 15. ◻

Claim 19. *$k(\hat{y}_i) - \hat{n}_i < k(\hat{y}_{i-1}) - \hat{n}_{i-1}$ for all $i \in \mathcal{I} \setminus \{1\}$ such that $\hat{y}_i > \hat{y}_{i-1}$.*

Proof. Consider $i \in \mathcal{I} \setminus \{1\}$ such that $\hat{y}_i > \hat{y}_{i-1}$. By Claims 1, 15, and 18, we have $\hat{n}_i > \hat{n}_{i-1}$ and $\hat{v}_i = \mathcal{V}(\hat{y}_i, \hat{n}_i, b_i) = \mathcal{V}(\hat{y}_{i-1}, \hat{n}_{i-1}, b_i) > 0$. Define $\mathcal{N}(y, v, b)$ to be a function satisfying $v = \mathcal{V}(y, \mathcal{N}(y, v, b), b)$ and notice that $\mathcal{N}_y(y, v, b) \equiv \partial \mathcal{N}(y, v, b)/\partial y = \mathcal{V}_y(y, \mathcal{N}(y, v, b), b)/|\mathcal{V}_n(y, \mathcal{N}(y, v, b), b)|$.

Then $\hat{n}_i = \mathcal{N}(\hat{y}_i, \hat{v}_i, b_i)$ and $\hat{n}_{i-1} = \mathcal{N}(\hat{y}_{i-1}, \hat{v}_i, b_i)$ hold and, by the fundamental theorem of calculus, we have

$$\begin{aligned}
\hat{n}_i - \hat{n}_{i-1} &= \int_{\hat{y}_{i-1}}^{\hat{y}_i} \mathcal{N}_y(y, \hat{v}_i, b_i) dy \\
&= \int_{\hat{y}_{i-1}}^{\hat{y}_i} \frac{\mathcal{V}_y(y, \mathcal{N}(y, \hat{v}_i, b_i), b_i)}{|\mathcal{V}_n(y, \mathcal{N}(y, \hat{v}_i, b_i), b_i)|} dy \\
&\geq \int_{\hat{y}_{i-1}}^{\hat{y}_i} \frac{\mathcal{V}_y(\hat{y}_i, \hat{n}_i, b_i)}{|\mathcal{V}_n(\hat{y}_i, \hat{n}_i, b_i)|} dy \\
&\geq \int_{\hat{y}_{i-1}}^{\hat{y}_i} k'(\hat{y}_i) dy \\
&> k(\hat{y}_i) - k(\hat{y}_{i-1}),
\end{aligned}$$

where the first inequality holds because $\mathcal{V}_y(y, n, b)/|\mathcal{V}_n(y, n, b)|$ is decreasing in (y, n) and $\mathcal{N}(y, v, b)$ is increasing in y, the second inequality holds due to Claim 13, and the last inequality holds because $k(\cdot)$ is strictly convex. □

B.12 Proof of Lemma 8

Suppose that the lemma is false, and consider (y', n', i') that satisfies (i)–(iv). Since $\mathcal{V}(y', n', b_{i'}) > \mathcal{V}(\hat{y}_{i'}, \hat{n}_{i'}, b_{i'})$ or $k(y') - n' < k(\hat{y}_{i'}) - \hat{n}_{i'}$ would contradict the assumption that $(\hat{y}_i, \hat{n}_i)_{i \in \mathcal{I}}$ solves Problem 7, consider the case $\mathcal{V}(y', n', b_{i'}) = \mathcal{V}(\hat{y}_{i'}, \hat{n}_{i'}, b_{i'})$ and $k(y') - n' = k(\hat{y}_{i'}) - \hat{n}_{i'}$. Since $\hat{y}_{i'} \leq y^*_{i'}(\hat{v}_{i'})$ and $\hat{n}_{i'} \leq n^*_{i'}(\hat{v}_{i'})$ hold by Claim 13, $\mathcal{V}(y', n', b_{i'}) = \mathcal{V}(\hat{y}_{i'}, \hat{n}_{i'}, b_{i'})$ and $k(y') - n' = k(\hat{y}_{i'}) - \hat{n}_{i'}$ imply $y' > y^*_{i'}(\hat{v}_{i'})$ and $n' > n^*_{i'}(\hat{v}_{i'})$, because $\mathcal{V}(y, n, b)$ is quasiconcave in (y, n) and $k(y)$ is strictly increasing and strictly convex in y.

Next, define $(\tilde{y}_i, \tilde{n}_i)_{i \in \mathcal{I}}$ such that $(\tilde{y}_{i'}, \tilde{n}_{i'}) = (y', n')$ and $(\tilde{y}_i, \tilde{n}_i) = (\hat{y}_i, \hat{n}_i)$ for all $i \in \mathcal{I} \setminus \{i'\}$. Then $(\tilde{y}_i, \tilde{n}_i)_{i \in \mathcal{I}}$ solves Problem 7, since so does $(\hat{y}_i, \hat{n}_i)_{i \in \mathcal{I}}$. However, $\tilde{y}_{i'} = y' > y^*_{i'}(\hat{v}_{i'})$ and $\tilde{n}_{i'} = n' > n^*_{i'}(\hat{v}_{i'})$ contradict Claim 13.

B.13 Proof of Proposition 2

I first characterize how the parameters for parental transfer distributions differ across groups.

Lemma 9. *$b_{L,h} \leq b_{L,h'}$, $b_{H,h} \leq b_{H,h'}$, and $f_{L,h|h} \geq f_{L,h'|h'}$ for all $h' \geq h$.*

Proof. First, notice that Part (ii) of Assumption 2 implies $b_{L,h'} < b_{H,h}$ for all (h, h'). Next, $\Phi(\cdot|\cdot)$ can be written as follows:

$$\Phi(b|h) = \begin{cases} 0, & \text{for } b < b_{L,h}, \\ f_{L,h|h}, & \text{for } b \in [b_{L,h}, b_{H,h}), \\ 1, & \text{for } b \geq b_{H,h}. \end{cases}$$

Now consider (h, h') such that $h' \geq h$. Then $\Phi(b|h) \geq \Phi(b|h')$ for all $b \in \mathbb{R}$. If $b_{L,h} > b_{L,h'}$, then $\Phi(b_{L,h'}|h') = f_{L,h'|h'} > \Phi(b_{L,h'}|h) = 0$. Since this is a contradiction, $b_{L,h} \leq b_{L,h'}$ must hold. Next, if $b_{H,h} > b_{H,h'}$, then $\Phi(b_{H,h'}|h') = 1 > f_{L,h|h} \geq \Phi(b_{H,h'}|h)$, which is also a contradiction. Therefore, $b_{H,h} \leq b_{H,h'}$ must hold. Finally, suppose that $f_{L,h|h} < f_{L,h'|h'}$. Then $\Phi(b_{L,h'}|h') = f_{L,h'|h'} > f_{L,h|h} \geq \Phi(b_{L,h'}|h)$. Since this is a contradiction, $f_{L,h|h} \geq f_{L,h'|h'}$ must hold. □

Since all groups are identical except for the distribution of parental transfers, I first consider the planning problem for a single group (i.e., assuming $|\mathcal{H}| = 1$) and then conduct comparative statics with respect to the parameters for the distribution of parental transfers. Therefore, I omit the group subscript h from now on.

The following lemma is useful to construct a separating equilibrium.

Lemma 10. $\underline{b} \equiv (y_2 k_2 - y_1 k_1)/(y_2 - y_1) \in (k_2, k_1 + y_1)$ *uniquely solves* $V_1(\underline{b}) = V_2(\underline{b})$.

Proof. First, note that the following inequality is implied by part (i) of Assumption 4:

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

where $k_2 < (y_2 k_2 - y_1 k_1)/(y_2 - y_1)$ 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 $k_2 - k_1 < y_1/y_2(y_2 - y_1)$.

For $b < k_2$, $j = 2$ is not feasible. Therefore, consider $b \geq k_2$. For $b \in [k_2, k_1 + y_1)$, $V_1(b) = (b - k_1)y_1$ and $V_2(b) = (b - k_2)y_2$. Therefore, $b = \underline{b}$ uniquely solves $(b - k_1)y_1 = (b - k_2)y_2$. Moreover, since $V_1(b) < V_2(b)$ for $b \in (\underline{b}, k_1 + y_1)$, Lemma 1 suggests that $V_1(b) < V_2(b)$ for all $b > \underline{b}$. □

Similar to the strategy used to prove Proposition 1, I first solve for a direct revelation mechanism, assuming that the choice probabilities are zero or one, and then verify later that the assumption is satisfied. Let $(\hat{y}_i, \hat{n}_i)_{i \in \mathcal{I}}$ be the solution to the following mechanism design problem:

Problem 8.

$$\min_{(y_i, n_i)_{i \in \mathcal{I}}} \sum_{i \in \mathcal{I}} f_i \mathcal{T}(y_i, n_i, b_i)$$

$$\textit{subject to} \sum_{i \in \mathcal{I}} f_i [k(y_i) - n_i] \leq 0, \tag{RC-2}$$

$$\mathcal{V}(y_H, n_H, b_H) \geq \mathcal{V}(y_L, n_L, b_H), \tag{DIC-2}$$

$$\mathcal{V}(y_L, n_L, b_L) \geq \mathcal{V}(y_H, n_H, b_L) \textit{ if } n_H \leq b_L, \tag{UIC-2}$$

$$y_i \in \{y_1, y_2\}, \quad \forall i \in \mathcal{I},$$

$$n_i \leq b_i, \quad \forall i \in \mathcal{I}.$$

Lemma 11. $\hat{y}_L = y_1 < \hat{y}_H = y_2$.

Proof. First, $\hat{y}_L > \hat{y}_H$ cannot hold due to (SCD). Next, suppose that $\hat{y}_L = \hat{y}_H = y_j$ for some $j \in \mathcal{J}$. Then $\hat{n}_L = \hat{n}_H = k_j$ must hold due to (RC-2), (DIC-2), and (UIC-2). Moreover, $b_L < \underline{b} < b_H$

holds due to Assumption 4, which implies $V_1(b_L) > V_2(b_L)$ and $V_1(b_H) < V_2(b_H)$. Therefore, $(y_L, n_L, y_H, n_H) = (y_1, k_1, y_2, k_2)$ satisfies (RC-2), (DIC-2), and (UIC-2), while strictly dominating both $(y_L, n_L, y_H, n_H) = (y_1, k_1, y_1, k_1)$ and $(y_L, n_L, y_H, n_H) = (y_2, k_2, y_2, k_2)$. Thus, $\hat{y}_L = \hat{y}_H$ cannot hold. □

Lemma 12. *$b_L < \hat{n}_L + \hat{y}_L$ or $b_H < \hat{n}_H + \hat{y}_H$.*

Proof. Suppose that $b_L \geq \hat{n}_L + \hat{y}_L$ and $b_H \geq \hat{n}_H + \hat{y}_H$. Then

$$\begin{aligned} f_L[k(\hat{y}_L) - \hat{n}_L] + f_H[k(\hat{y}_H) - \hat{n}_H] \geq & f_L(k_1 + y_1 - b_L) + f_H(k_2 + y_2 - b_H) \\ > & f_L(k_1 + y_1) + f_H(k_2 + y_2) - \underline{b} \\ > & k_1 + y_1 - \underline{b} \\ > & 0, \end{aligned}$$

where the strict inequalities hold due to $f_L b_L + f_H b_H < \underline{b}$, $k_2 + y_2 > k_1 + y_1$, and $\underline{b} < k_1 + y_1$. Therefore, (RC-2) is violated. □

As before, guess that (UIC-2) does not bind and verify it later. Let λ and ψ be the Lagrangian multipliers on (RC-2) and (DIC-2). Then the first order conditions with respect to n_L and n_H for Problem 8 are

$$|\mathcal{V}_n(\hat{y}_L, \hat{n}_L, b_L)| \leq (1 + \lambda) + \frac{\psi}{f_L} |\mathcal{V}_n(\hat{y}_L, \hat{n}_L, b_H)|, \tag{48}$$

$$|\mathcal{V}_n(\hat{y}_H, \hat{n}_H, b_H)| \leq (1 + \lambda) \left(\frac{f_H}{f_H + \psi} \right). \tag{49}$$

Lemma 13. *$\hat{n}_L < b_L$ and $\hat{n}_H < b_H$.*

Proof. Suppose that $\hat{n}_L = b_L$. Then $|\mathcal{V}_n(\hat{y}_L, \hat{n}_L, b_L)| = [\hat{y}_L/(b_L - \hat{n}_L)]^{1/2} = \infty$, while $|\mathcal{V}_n(\hat{y}_L, \hat{n}_L, b_H)| = \max\{[\hat{y}_L/(b_H - \hat{n}_L)]^{1/2}, 1\} < \infty$ due to $\hat{n}_L = b_L < b_H$. Since (48) cannot hold, $\hat{n}_L < b_L$. Next, $\hat{n}_L < b_H$ and (DIC-2) imply $\mathcal{V}(\hat{y}_H, \hat{n}_H, b_H) > 0$. Therefore, $\hat{n}_H < b_H$ must hold. □

Therefore, (48) and (49) hold as equalities.

Lemma 14. *$\psi > 0$.*

Proof. Suppose that $\psi = 0$. Then, from (48) and (49), we have $|\mathcal{V}_n(\hat{y}_L, \hat{n}_L, b_L)| = |\mathcal{V}_n(\hat{y}_H, \hat{n}_H, b_H)|$. Moreover, Lemma 12 implies that both $b_L < \hat{n}_L + \hat{y}_L$ and $b_H < \hat{n}_H + \hat{y}_H$ must hold. Therefore, $|\mathcal{V}_n(\hat{y}_L, \hat{n}_L, b_L)| = |\mathcal{V}_n(\hat{y}_H, \hat{n}_H, b_H)|$ can be written as

$$\frac{\hat{y}_L}{b_L - \hat{n}_L} = \frac{\hat{y}_H}{b_H - \hat{n}_H}.$$

From this and (RC-2), we have

$$\hat{n}_H - \hat{n}_L = \frac{(y_1 b_H - y_2 b_L) + (y_2 - y_1)(f_L k_1 + f_H k_2)}{f_L y_1 + f_H y_2},$$

which is greater than $y_2 - y_1$ if and only if

$$b_H - \frac{y_2}{y_1} b_L > \frac{(y_2 - y_1)[f_L(y_1 - k_1) + f_H(y_2 - k_2)]}{y_1}.$$

Since $f_L(y_1 - k_1) + f_H(y_2 - k_2) < y_2 - k_1$, $\hat{n}_H - \hat{n}_L > y_2 - y_1$ holds if part (iii) of Assumption 4 holds. $\hat{n}_H - \hat{n}_L > y_2 - y_1$ implies $\mathcal{V}(\hat{y}_H, \hat{n}_H, b) < \mathcal{V}(\hat{y}_L, \hat{n}_L, b)$ for some $b \geq \max\{\hat{n}_L + \hat{y}_L, \hat{n}_H + \hat{y}_H\} > b_H$. Therefore, by (SCD), $\mathcal{V}(\hat{y}_H, \hat{n}_H, b_H) < \mathcal{V}(\hat{y}_L, \hat{n}_L, b_H)$ must hold as well, leading to a contradiction. □

Lemma 15. *$\mathcal{V}(\hat{y}_H, \hat{n}_H, b_H) = \mathcal{V}(\hat{y}_L, \hat{n}_L, b_H)$, $\mathcal{V}(\hat{y}_L, \hat{n}_L, b_L) \geq \mathcal{V}(\hat{y}_H, \hat{n}_H, b_L)$, $|\mathcal{V}_n(\hat{y}_H, \hat{n}_H, b_H)| \geq |\mathcal{V}_n(\hat{y}_L, \hat{n}_L, b_H)|$, and $\hat{n}_L < \hat{n}_H$.*

Proof. $\mathcal{V}(\hat{y}_H, \hat{n}_H, b_H) = \mathcal{V}(\hat{y}_L, \hat{n}_L, b_H)$ holds due to $\psi > 0$. It also implies $\hat{n}_L < \hat{n}_H$, since $\mathcal{V}(y, n, b)$ is strictly increasing in y and strictly decreasing in n. $\mathcal{V}(\hat{y}_L, \hat{n}_L, b_L) \geq \mathcal{V}(\hat{y}_H, \hat{n}_H, b_L)$ follows from $\mathcal{V}(\hat{y}_H, \hat{n}_H, b_H) = \mathcal{V}(\hat{y}_L, \hat{n}_L, b_H)$ and (SCD). $|\mathcal{V}_n(\hat{y}_H, \hat{n}_H, b_H)| \geq |\mathcal{V}_n(\hat{y}_L, \hat{n}_L, b_H)|$ holds due to $|\mathcal{V}_n(y, n, b)| = \max\{y/(b-n), 1\}$, $\hat{y}_H > \hat{y}_L$, and $\hat{n}_H > \hat{n}_L$. □

Lemma 16. *$\lambda > 0$.*

Proof. Suppose that $\lambda = 0$. Then, from (49), Lemma 13, and Lemma 14, we have $|\mathcal{V}_n(\hat{y}_H, \hat{n}_H, b_H)| = f_H/(f_H + \psi) < 1$, which contradicts $|\mathcal{V}_n(y, n, b)| \geq 1$ for all (y, n, b) such that $n \leq b$. □

Therefore, $(\hat{n}_L, \hat{n}_H)$ solves

$$\mathcal{V}(\hat{y}_L, \hat{n}_L, b_H) - \mathcal{V}(\hat{y}_H, \hat{n}_H, b_H) = 0, \tag{50}$$

$$f_L[k(\hat{y}_L) - \hat{n}_L] + f_H[k(\hat{y}_H) - \hat{n}_H] = 0. \tag{51}$$

Lemma 17. *$k(\hat{y}_L) - \hat{n}_L > 0 > k(\hat{y}_H) - \hat{n}_H$.*

Proof. First, notice that $k(\hat{y}_L) - \hat{n}_L$ and $k(\hat{y}_H) - \hat{n}_H$ have opposite signs, as (51) implies

$$k(\hat{y}_L) - \hat{n}_L = -\frac{f_H}{f_L}[k(\hat{y}_H) - \hat{n}_H]. \tag{52}$$

Suppose that $k(\hat{y}_L) - \hat{n}_L \leq k(\hat{y}_H) - \hat{n}_H$, which implies $k(\hat{y}_L) - \hat{n}_L \leq 0 \leq k(\hat{y}_H) - \hat{n}_H$. Then

$$\begin{aligned}
\mathcal{V}(\hat{y}_H, \hat{n}_H, b_H) - \mathcal{V}(\hat{y}_L, \hat{n}_L, b_H) \geq& \mathcal{V}(\hat{y}_H, k(\hat{y}_H), b_H) - \mathcal{V}(\hat{y}_L, k(\hat{y}_L), b_H) \\
=& \mathcal{V}(y_2, k_2, b_H) - \mathcal{V}(y_1, k_1, b_H) \\
>& 0,
\end{aligned}$$

where the first inequality holds because $\mathcal{V}(y, n, b)$ is strictly decreasing in n and the second inequality holds due to $b_H > \underline{b}$ and Lemma 10. Because $\mathcal{V}(\hat{y}_H, \hat{n}_H, b_H) > \mathcal{V}(\hat{y}_L, \hat{n}_L, b_H)$ contradicts Lemma 14, $k(\hat{y}_L) - \hat{n}_L > k(\hat{y}_H) - \hat{n}_H$ must hold. □

This lemma implies that although (50) suggests that the type $i = H$ is indifferent between $(\hat{y}_L, \hat{n}_L)$ and $(\hat{y}_H, \hat{n}_H)$, assigning $(\hat{y}_H, \hat{n}_H)$ to the type $i = H$ leads to strictly lower aggregate distortions, as it is cheaper for the social planner. Therefore, any randomization between $(\hat{y}_L, \hat{n}_L)$ and $(\hat{y}_H, \hat{n}_H)$ for the type $i = H$ is not socially optimal. This proves that the assumption that the choice probabilities are either zero or one is not restrictive, which makes Problem 8 identical to Problem 2 (for a single group). The solution to Problem 2, $(\hat{\boldsymbol{g}}, (\hat{\boldsymbol{p}}(b_i))_{i\in\mathcal{I}})$, can be constructed from $(\hat{y}_i, \hat{n}_i)_{i\in\mathcal{I}}$ by setting $\hat{g}_L = k(\hat{y}_L) - \hat{n}_L$, $\hat{g}_H = k(\hat{y}_H) - \hat{n}_H$, and $\hat{p}_1(b_L) = \hat{p}_2(b_H) = 1$. Then it is easy to see that part (i) of Proposition 2 holds.

To prove part (ii) of Proposition 2, consider an alternative set of parameters for the distribution of parental transfers (f'_L, f'_H, b'_L, b'_H) such that (i) $b_{L'} \geq b_L$, $b_{H'} \geq b_H$, and $f_{H'} \geq f_H$, and (ii) Assumption 4 is satisfied. These parameters can be thought of as those for a higher parental wealth group. For such (f'_L, f'_H, b'_L, b'_H), let $(\hat{\boldsymbol{g}}', (\hat{\boldsymbol{p}}'(b_i))_{i\in\mathcal{I}})$ and $(\hat{y}'_i, \hat{n}'_i)_{i\in\mathcal{I}}$ be the solutions to Problems 2 and 8, respectively. From (50) and (51), it is obvious that the parameter b_L does not affect the solution. Therefore, it suffices to consider the effects of b_H and f_H.

Lemma 18. *If $b'_H > b_H$ and $f'_H = f_H$, then $\hat{n}_L \geq \hat{n}'_L$.*

Proof. By differentiating (50) and (51) with respect to b_H, we get

$$\begin{aligned} |\mathcal{V}_n(\hat{y}_L, \hat{n}_L, b_H)| \left(1 - \frac{\partial \hat{n}_L}{\partial b_H}\right) &= |\mathcal{V}_n(\hat{y}_H, \hat{n}_H, b_H)| \left(1 - \frac{\partial \hat{n}_H}{\partial b_H}\right), \\ \frac{\partial \hat{n}_L}{\partial b_H} &= -\frac{f_H}{1-f_H}\frac{\partial \hat{n}_H}{\partial b_H}, \end{aligned}$$

where $\mathcal{V}_b(y, n, b) = |\mathcal{V}_n(y, n, b)|$ is used. From these, we have

$$\frac{\partial \hat{n}_L}{\partial b_H} = \frac{|\mathcal{V}_n(\hat{y}_L, \hat{n}_L, b_H)| - |\mathcal{V}_n(\hat{y}_H, \hat{n}_H, b_H)|}{|\mathcal{V}_n(\hat{y}_L, \hat{n}_L, b_H)| + \frac{1-f_H}{f_H}|\mathcal{V}_n(\hat{y}_H, \hat{n}_H, b_H)|} \leq 0.$$

□

Lemma 19. *If $f'_H > f_H$ and $b'_H = b_H$, then $\hat{n}_L \geq \hat{n}'_L$.*

Proof. By differentiating (50) and (51) with respect to f_H, we get

$$\begin{aligned} |\mathcal{V}_n(\hat{y}_L, \hat{n}_L, b_H)| \frac{\partial \hat{n}_L}{\partial f_H} &= |\mathcal{V}_n(\hat{y}_H, \hat{n}_H, b_H)| \frac{\partial \hat{n}_H}{\partial f_H}, \\ \frac{\partial \hat{n}_L}{\partial f_H} &= \frac{k(\hat{y}_H) - \hat{n}_H}{(1-f_H)^2} - \frac{f_H}{1-f_H}\frac{\partial \hat{n}_H}{\partial f_H}, \end{aligned}$$

which can be combined to give

$$\frac{\partial \hat{n}_L}{\partial f_H} = \frac{|\mathcal{V}_n(\hat{y}_H, \hat{n}_H, b_H)| \frac{k(\hat{y}_H) - \hat{n}_H}{(1-f_H)^2}}{|\mathcal{V}_n(\hat{y}_L, \hat{n}_L, b_H)| \frac{f_H}{1-f_H} + |\mathcal{V}_n(\hat{y}_H, \hat{n}_H, b_H)|} < 0.$$

□

B.14 Optimal Policy with Unobservable Heterogeneity in Ability

Consider individuals who are endowed with an identical parental transfer, $b \in \mathbb{R}_+$, and who differ only in ability, a, that cannot be observed by the social planner. Let $\mathcal{A}$ be a totally ordered set of ability levels and $y(k,a)$ be post-schooling earnings for an individual with ability $a \in \mathcal{A}$ and a monetary investment in schooling $k \in \mathbb{R}_+$. Denote the partial derivatives of the earnings function as follows: $y_k(k,a) = \partial y(k,a)/\partial k$, $y_a(k,a) = \partial y(k,a)/\partial a$, and $y_{ka}(k,a) = \partial^2 y(k,a)/\partial k \partial a$. The following set of assumptions defines the relationship between investment, ability, and earnings.

Assumption 6. *(i) $y(\cdot,\cdot)$ is twice continuously differentiable and strictly increasing; (ii) for all $a \in \mathcal{A}$, $y(\cdot,a)$ is strictly concave and $\lim_{k\downarrow 0} y_k(k,a) = \infty$; and (iii) $y_{ka}(k,a)y(k,a) > y_k(k,a)y_a(k,a)$ for all $(k,a) \in \mathbb{R}_+ \times \mathcal{A}$.*

Both ability and investment increase earnings. The marginal return to investment is very high at low investment levels, and it diminishes as investment increases. The marginal return to investment is increasing in ability, reflecting the complementarity between ability and investment in producing human capital. In particular, part (iii) of Assumption 6 assumes that such complementary is strong in the sense that the Hicksian elasticity of substitution between k and a, $y_k(k,a)y_a(k,a)/[y_{ka}(k,a)y(k,a)]$, is strictly lower than the elasticity of intertemporal substitution, 1.[52] As discussed by Lochner and Monge-Naranjo (2011) and Caucutt, Lochner, and Park (2017), this assumption ensures that higher ability individuals invest more, even when they are borrowing constrained.

Similar to the proof of Proposition 1, I consider designing a direct revelation mechanism, where both schooling investment k and net payment $n \in \mathbb{R}$ are determined based on individuals' truthful reports about their ability. Therefore, the difference between the schooling cost and net payment, $k-n$, is the amount of subsidy.

Let $\mathcal{V}(k,n,a)$ be the money-metric indirect utility of a bundle (k,n) for an individual with ability a:

$$\mathcal{V}(k,n,a) = \begin{cases} 2\left[(b-n)y(k,a)\right]^{\frac{1}{2}}, & \text{for } b < n + y(k,a), \\ y(k,a) + b - n, & \text{for } b \geq n + y(k,a). \end{cases}$$

It is easy to see that under Assumption 6, $\mathcal{V}(k,n,a)$ is strictly increasing in (k,a) (unless $b = n$), strictly decreasing in n, and quasiconcave and continuously differentiable in (k,n). Importantly, the following *strict* Mirrlees-Spence condition is satisfied: for all $a' > a$,

$$\frac{\mathcal{V}_k(k,n,a)}{|\mathcal{V}_n(k,n,a)|} < \frac{\mathcal{V}_k(k,n,a')}{|\mathcal{V}_n(k,n,a')|}, \tag{SMS}$$

where $\mathcal{V}_k(k,n,a) \equiv \partial\mathcal{V}(k,n,a)/\partial k = \min\{[(b-n)/y(k,a)]^{1/2}, 1\}y_k(k,a)$ and $|\mathcal{V}_n(k,n,a)| \equiv |\partial\mathcal{V}(k,n,a)/\partial n| = \max\{[y(k,a)/(b-n)]^{1/2}, 1\} > 0$ are calculated by the envelope theorem. As

[52] When the earnings function is multiplicably separable in ability, as commonly assumed in empirical studies, the Hicksian elasticity of substitution is 1 and the Mirrlees-Spence condition holds only weakly. Although the case with the weak Mirrlees-Spence condition can be handled in the same way as in the proof of Proposition 1, I impose a stronger assumption that gives a strict Mirrlees-Spence condition to simplify analysis.

shown by Edlin and Shannon (1998), (SMS) implies the following *strict* single-crossing differences condition: for all $k' > k$ and $a' > a$,

$$\mathcal{V}(k', n', a) \geq \mathcal{V}(k, n, a) \Rightarrow \mathcal{V}(k', n', a') > \mathcal{V}(k, n, a'). \tag{SSCD}$$

Therefore, by Milgrom and Shannon (1994), a positive relationship between schooling investment and ability holds under any incentive-compatible allocations.

A direct mechanism consists of a message space, $\mathcal{I}$, and an allocation function, $(k, n) : \mathcal{I} \to \mathbb{R}_+ \times \mathbb{R}$, that assigns an allocation (k_i, n_i) to those reporting to have ability $a_i \in \mathcal{A}$. Without loss of generality, assume that $\mathcal{I} = \{1, 2, \ldots, I\}$ with $I > 2$ and $a_{i-1} < a_i$ for all $i \in \mathcal{I} \setminus \{1\}$. Let $\mathcal{T}(k, n, a_i) \equiv y_i^* - k_i^* + b + k - n - \mathcal{V}(k, n, a_i)$ be the distortion of type i with (k, n), where $k_i^* \equiv \text{argmax}_{k \in \mathbb{R}_+}\{y(k, a_i) - k\}$, and $y_i^* \equiv y(k_i^*, a_i)$.

The planning problem is as follows:

Problem 9.

$$\min_{(k_i, n_i)_{i \in \mathcal{I}}} \sum_{i \in \mathcal{I}} f_i \mathcal{T}(k_i, n_i, a_i)$$

$$\textit{subject to} \sum_{i \in \mathcal{I}} f_i (k_i - n_i) \leq G, \tag{RC-3}$$

$$\mathcal{V}(k_i, n_i, a_i) \geq \mathcal{V}(k_{i'}, n_{i'}, a_i), \quad \forall (i, i') \in \mathcal{I} \times \mathcal{I}, \tag{IC-3}$$

$$k_i \geq 0, \quad \forall i \in \mathcal{I}, \tag{53}$$

$$n_i \leq b, \quad \forall i \in \mathcal{I}. \tag{54}$$

As before, I consider the non-trivial case where (RC-3) binds. The remainder of this subsection provides the proof of the following proposition.

Proposition 5. *Suppose that Assumption 6 holds and $(\hat{k}_i, \hat{n}_i)_{i \in \mathcal{I}}$ solves Problem 9. Then, for all $i \in \mathcal{I} \setminus \{1\}$, $\hat{n}_i \geq \hat{n}_{i-1}$, $\hat{k}_i \geq \hat{k}_{i-1}$ and $\hat{k}_i - \hat{n}_i \geq \hat{k}_{i-1} - \hat{n}_{i-1}$. Moreover, $\hat{k}_i - \hat{n}_i > \hat{k}_{i-1} - \hat{n}_{i-1}$ if $\hat{k}_i > \hat{k}_{i-1}$.*

For $i \neq i'$, (IC-3) can be broken down into downward and upward incentive compatibility constraints:

$$\mathcal{V}(k_i, n_i, a_i) \geq \mathcal{V}(k_{i'}, n_{i'}, a_i), \quad \forall (i, i') \in \mathcal{I} \times \mathcal{I} \text{ such that } i > i', \tag{DIC-3}$$

$$\mathcal{V}(k_i, n_i, a_i) \geq \mathcal{V}(k_{i'}, n_{i'}, a_i), \quad \forall (i, i') \in \mathcal{I} \times \mathcal{I} \text{ such that } i < i'. \tag{UIC-3}$$

Instead of Problem 9, I consider a problem where (DIC-3) is replaced by (MC-3):

Problem 10.

$$\min_{(k_i,n_i)_{i\in\mathcal{I}}} \sum_{i\in\mathcal{I}} f_i \mathcal{T}(k_i,n_i,a_i)$$

subject to (RC-3), (UIC-3), (53), (54), *and*

$$k_i \geq k_{i'}, \quad \forall (i,i') \in \mathcal{I}\times\mathcal{I} \text{ such that } i \geq i'. \tag{MC-3}$$

The next two lemmas show that Problem 10 is identical to Problem 9.

Lemma 20. *A solution to Problem 9 satisfies* (MC-3).

Proof. Suppose that $(k_i,n_i)_{i\in\mathcal{I}}$ solves Problem 9 and there exists $i > i'$ such that $k_i < k_{i'}$. From (IC-3), we have $\mathcal{V}(k_{i'},n_{i'},a_{i'}) \geq \mathcal{V}(k_i,n_i,a_{i'})$. This implies, by (SSCD), $\mathcal{V}(k_{i'},n_{i'},a_i) > \mathcal{V}(k_i,n_i,a_i)$, which violates (IC-3). Since this contradicts the assumption that $(k_i,n_i)_{i\in\mathcal{I}}$ solves Problem 9, (MC-3) must be satisfied. □

Lemma 21. *A solution to Problem 10 satisfies* (DIC-3).

Proof. Consider the following adjacent incentive compatibility constraints:

$$\mathcal{V}(k_i,n_i,a_i) \geq \mathcal{V}(k_{i-1},n_{i-1},a_i), \quad \forall i \in \mathcal{I}\setminus\{1\}, \tag{ADIC-3}$$

$$\mathcal{V}(k_i,n_i,a_i) \geq \mathcal{V}(k_{i+1},n_{i+1},a_i), \quad \forall i \in \mathcal{I}\setminus\{I\}. \tag{AUIC-3}$$

Based on earlier results in Claims 7 and 17, which rely only on (SCD), I establish the following results without proof.

Claim 20. *(i)* (MC-3) *and* (AUIC-3) *imply* (UIC-3)*; (ii)* (MC-3) *and* (ADIC-3) *imply* (DIC-3).

Therefore, (UIC-3) in Problem 10 can be replaced by (AUIC-3). Moreover, (MC-3) and (AUIC-3) imply that n_i is also increasing in i.

Claim 21. (MC-3) *and* (AUIC-3) *imply that* $n_i \leq n_{i+1}$ *for all* $i \in \mathcal{I}\setminus\{I\}$.

Proof. Suppose that $k_i \leq k_{i+1}$ and $n_i > n_{i+1}$. Then $\mathcal{V}(k_i,n_i,a_i) \leq \mathcal{V}(k_{i+1},n_i,a_i) < \mathcal{V}(k_{i+1},n_{i+1},a_i)$ holds, because $\mathcal{V}(k,n,a)$ is strictly decreasing in n. Therefore, (AUIC-3) is violated. □

Therefore, (54) can be reduced to a single constraint, $n_I \leq b$. Moreover, (MC-3) implies that it suffices to consider only $k_1 \geq 0$ instead of (53). With reduced constraints, the Lagrangian for Problem 10 is

$$\sum_{i\in\mathcal{I}} f_i \Big\{\mathcal{V}(k_i,n_i,a_i) - (1+\lambda)(k_i - n_i)\Big\} + \lambda G - \sum_{i\in\mathcal{I}} f_i (b + y_i^* - k_i^*)$$
$$+ \sum_{i\in\mathcal{I}\setminus\{I\}} \Big\{\psi_i [\mathcal{V}(k_i,n_i,a_i) - \mathcal{V}(k_{i+1},n_{i+1},a_i)] - \varphi_i (k_i - k_{i+1})\Big\} + \zeta(b - n_I) + \xi k_1,$$

where λ, φ_i, ψ_i, ζ, and ξ are Lagrangian multipliers on (RC-3), (MC-3), (AUIC-3), $n_I \leq b$, and $k_1 \geq 0$, respectively.

In the remainder of this subsection, let $(\hat{k}_i, \hat{n}_i)_{i\in\mathcal{I}}$ be a solution to Problem 10. The first order conditions are, for all $i \in \mathcal{I}$,

$$(f_i + \psi_i)\mathcal{V}_k(\hat{k}_i, \hat{n}_i, a_i) - (1+\lambda)f_i - \psi_{i-1}\mathcal{V}_k(\hat{k}_i, \hat{n}_i, a_{i-1}) \leq \varphi_i - \varphi_{i-1}, \tag{55}$$

$$(f_i + \psi_i)|\mathcal{V}_n(\hat{k}_i, \hat{n}_i, a_i)| - (1+\lambda)f_i - \psi_{i-1}|\mathcal{V}_n(\hat{k}_i, \hat{n}_i, a_{i-1})| \leq 0, \tag{56}$$

where $\psi_i = \varphi_i = 0$ for $i = I$ and $\psi_{i-1} = \varphi_{i-1} = 0$ for $i = 1$. (55) holds as equality if $i > 1$ or $\hat{k}_1 > 0$, while (56) holds as equality if $i < I$ or $\hat{n}_I < b$. When (56) holds as equality, we can combine (55) and (56) to get

$$\begin{aligned}\varphi_i - \varphi_{i-1} \geq &(1+\lambda)f_i\left(\frac{\mathcal{V}_k(\hat{k}_i, \hat{n}_i, a_i)}{|\mathcal{V}_n(\hat{k}_i, \hat{n}_i, a_i)|} - 1\right)\\ &+ \psi_{i-1}\left(\frac{\mathcal{V}_k(\hat{k}_i, \hat{n}_i, a_i)}{|\mathcal{V}_n(\hat{k}_i, \hat{n}_i, a_i)|}|\mathcal{V}_n(\hat{k}_i, \hat{n}_i, a_{i-1})| - \mathcal{V}_k(\hat{k}_i, \hat{n}_i, a_{i-1})\right).\end{aligned} \tag{57}$$

Claim 22. *If $\hat{k}_i = \hat{k}_{i+1}$ for $i < I$, then $\hat{n}_i = \hat{n}_{i+1}$.*

Proof. Suppose that $\hat{k}_i = \hat{k}_{i+1}$ and $\hat{n}_i < \hat{n}_{i+1}$ for some $i < I$. Then $\mathcal{V}(\hat{k}_i, \hat{n}_i, a_i) > \mathcal{V}(\hat{k}_{i+1}, \hat{n}_{i+1}, a_i)$, which implies $\psi_i = 0$. Then (56) for $i+1$ and i can be written as

$$\begin{aligned}|\mathcal{V}_n(\hat{k}_{i+1}, \hat{n}_{i+1}, a_{i+1})| \leq &\frac{(1+\lambda)f_{i+1}}{f_{i+1} + \psi_{i+1}},\\ |\mathcal{V}_n(\hat{k}_i, \hat{n}_i, a_i)| = &(1+\lambda) + \frac{\psi_{i-1}}{f_i}|\mathcal{V}_n(\hat{k}_i, \hat{n}_i, a_{i-1})|,\end{aligned}$$

which imply

$$|\mathcal{V}_n(\hat{k}_{i+1}, \hat{n}_{i+1}, a_{i+1})| \leq 1 + \lambda \leq |\mathcal{V}_n(\hat{k}_i, \hat{n}_i, a_i)|. \tag{58}$$

From $\lambda > 0$, $|\mathcal{V}_n(\hat{k}_i, \hat{n}_i, a_i)| > 1$ holds. Since $|\mathcal{V}_n(k, n, a)|$ is strictly increasing in n and a when $|\mathcal{V}_n(k, n, a)| > 1$, $\hat{k}_i = \hat{k}_{i+1}$ and $\hat{n}_i < \hat{n}_{i+1}$ imply $|\mathcal{V}_n(\hat{k}_{i+1}, \hat{n}_{i+1}, a_{i+1})| > |\mathcal{V}_n(\hat{k}_i, \hat{n}_i, a_i)|$, which contradicts (58). Therefore, $\hat{n}_i = \hat{n}_{i+1}$ must hold. □

As before, the most efficient way to deliver a given level of utility v to type i is defined as follows: for $v \in \mathbb{R}_+$,

$$(k_i^*(v), n_i^*(v)) \equiv \operatorname*{argmin}_{(k,n)\in\mathbb{R}_+\times\mathbb{R}} \left\{k - n | \mathcal{V}(k, n, a_i) \geq v, n \leq b\right\}.$$

Claim 23. *For $v = 0$, $n_i^*(v) = b$ and $k_i^*(v) = 0$. For $v \in (0, 2y_i^*)$, $k_i^*(v) \in (0, k_i^*)$ and $n_i^*(v) \in (b - y_i^*, b)$. For $v \geq 2y_i^*$, $k_i^*(v) = k_i^*$ and $n_i^*(v) = y_i^* + b - v \leq b - y_i^*$.*

Proof. Since the case with $v = 0$ is trivial, consider $v > 0$. Since $n_i^*(v) = b$ implies $\mathcal{V}(k_i^*(v), n_i^*(v), a_i) =$

$0 < v, n_i^*(v) < b$ must hold. In this case, the first order conditions imply

$$\frac{\mathcal{V}_k(k_i^*(v), n_i^*(v), a_i)}{|\mathcal{V}_n(k_i^*(v), n_i^*(v), a_i)|} = 1. \tag{59}$$

Suppose that $n_i^*(v) + y(k_i^*(v), a_i) \leq b$. Then $\mathcal{V}(k_i^*(v), n_i^*(v), a_i) = y(k_i^*(v), a_i) - n_i^*(v) + b = v$ and (59) implies $k_i^*(v) = k_i^*$. Moreover, $n_i^*(v) + y(k_i^*(v), a_i) \leq b$ also implies $v = y_i^* - n_i^*(v) + b \geq y_i^* - (b - y_i^*) + b = 2y_i^*$. Therefore, for $v < 2y_i^*$, $y(k_i^*(v), a_i) > b - n_i^*(v) > 0$ must hold and (59) gives

$$y_k(k_i^*, a_i) = 1 = \frac{\mathcal{V}_k(k_i^*(v), n_i^*(v), a_i)}{|\mathcal{V}_n(k_i^*(v), n_i^*(v), a_i)|} = \frac{b - n_i^*(v)}{y(k_i^*(v), a_i)} y_k(k_i^*(v), a_i) < y_k(k_i^*(v), a_i), \tag{60}$$

which implies $k_i^*(v) < k_i^*$. □

Claim 24. *$k_i^*(v)$ and $n_i^*(v)$ are continuous in v.*

Proof. First, notice that $k_i^*(v)$ and $n_i^*(v)$ are continuous in v for $v \geq 2y_i^*$. For $v \in (0, 2y_i^*)$, $(k_i^*(v), n_i^*(v))$ solves $b - n_i^*(v) = y_k(k_i^*(v), a_i) y(k_i^*(v), a_i)$ and $v = 2\{[b - n_i^*(v)] y(k_i^*(v), a_i)\}^{\frac{1}{2}}$. Therefore, $k_i^*(v)$ solves $v = 2y(k_i^*(v), a_i)[y_k(k_i^*(v), a_i)]^{\frac{1}{2}}$. Since $y(\cdot, a_i)$ and $y_k(\cdot, a_i)$ are continuous, $k_i^*(v)$ and $n_i^*(v) = b - y_k(k_i^*(v), a_i) y(k_i^*(v), a_i)$ are continuous in v over $(0, 2y_i^*)$. Moreover, $\lim_{v \downarrow 0}(k_i^*(v), n_i^*(v)) = (0, b)$ and $\lim_{v \uparrow 2y_i^*}(k_i^*(v), n_i^*(v)) = (k_i^*, b - y_i^*)$. □

Let $\hat{v}_i \equiv \mathcal{V}(\hat{k}_i, \hat{n}_i, a_i)$ and consider $i \in \mathcal{I}$ with $\hat{v}_i > 0$. When $\hat{k}_i > k_i^*(\hat{v}_i)$ or $\hat{n}_i > n_i^*(\hat{v}_i)$, both $\hat{k}_i > k_i^*(\hat{v}_i)$ and $\hat{n}_i > n_i^*(\hat{v}_i)$ must hold, because $\mathcal{V}(\hat{k}_i, \hat{n}_i, a_i) = \mathcal{V}(k_i^*(\hat{v}_i), n_i^*(\hat{v}_i), a_i)$ and $\mathcal{V}(k, n, a)$ is strictly increasing in k and strictly decreasing in n. Since $\mathcal{V}_k(k, n, a)/|\mathcal{V}_n(k, n, a)|$ is decreasing in (k, n), $\hat{k}_i > k_i^*(\hat{v}_i)$ and $\hat{n}_i > n_i^*(\hat{v}_i)$ imply

$$\frac{\mathcal{V}_k(\hat{k}_i, \hat{n}_i, a_i)}{|\mathcal{V}_n(\hat{k}_i, \hat{n}_i, a_i)|} < 1. \tag{61}$$

Similarly, when $\hat{k}_i < k_i^*(\hat{v}_i)$ or $\hat{n}_i < n_i^*(\hat{v}_i)$, both $\hat{k}_i < k_i^*(\hat{v}_i)$ and $\hat{n}_i < n_i^*(\hat{v}_i)$ hold and

$$\frac{\mathcal{V}_k(\hat{k}_i, \hat{n}_i, a_i)}{|\mathcal{V}_n(\hat{k}_i, \hat{n}_i, a_i)|} > 1. \tag{62}$$

Claim 25. *$\hat{k}_i \geq k_i^*(\hat{v}_i)$ and $\hat{n}_i \geq n_i^*(\hat{v}_i)$ for all $i \in \mathcal{I}$.*

Proof. First, consider the case with $\hat{n}_i = b$. Since $\hat{v}_i = 0$, $\hat{k}_i \geq k_i^*(0) = 0$ and $\hat{n}_i = n_i^*(0) = b$ hold. Next, consider the case $\hat{n}_i < b$. Suppose that there exists $i \in \mathcal{I}$ such that $\hat{k}_i < k_i^*(\hat{v}_i)$ or $\hat{n}_i < n_i^*(\hat{v}_i)$. Therefore, (62) holds for such $i \in \mathcal{I}$. Define $i' \geq i$ to be the largest type such that $\hat{k}_{i'} = \hat{k}_i$. Then $\hat{n}_{i'} = \hat{n}_i$ holds due to Claim 22, and (62) for i' also holds:

$$\frac{\mathcal{V}_k(\hat{k}_{i'}, \hat{n}_{i'}, a_{i'})}{|\mathcal{V}_n(\hat{k}_{i'}, \hat{n}_{i'}, a_{i'})|} \geq \frac{\mathcal{V}_k(\hat{k}_{i'}, \hat{n}_{i'}, a_i)}{|\mathcal{V}_n(\hat{k}_{i'}, \hat{n}_{i'}, a_i)|} = \frac{\mathcal{V}_k(\hat{k}_i, \hat{n}_i, a_i)}{|\mathcal{V}_n(\hat{k}_i, \hat{n}_i, a_i)|} > 1, \tag{63}$$

where the first inequality holds due to (SMS).

For $i' \in \mathcal{I} \setminus \{1\}$, the first order condition (57) is

$$\begin{aligned}
&\varphi_{i'} - \varphi_{i'-1} \\
=&(1+\lambda) f_{i'} \left(\frac{\mathcal{V}_k(\hat{k}_{i'}, \hat{n}_{i'}, a_{i'})}{|\mathcal{V}_n(\hat{k}_{i'}, \hat{n}_{i'}, a_{i'})|} - 1 \right) + \psi_{i'-1} \left(\frac{\mathcal{V}_k(\hat{k}_{i'}, \hat{n}_{i'}, a_{i'})}{|\mathcal{V}_n(\hat{k}_{i'}, \hat{n}_{i'}, a_{i'})|} |\mathcal{V}_n(\hat{k}_{i'}, \hat{n}_{i'}, a_{i'-1})| - \mathcal{V}_k(\hat{k}_{i'}, \hat{n}_{i'}, a_{i'-1}) \right) \\
>&0,
\end{aligned}$$

where the inequality follows from (63) and (SMS). Therefore, $\varphi_{i'-1} < \varphi_{i'}$ holds. However, by the definition of i', $\hat{k}_{i'+1} > \hat{k}_{i'}$, so (MC-3) does not bind and $\varphi_{i'} = 0$. Since $\varphi_{i'-1} \geq 0$, this leads to a contradiction.

Similarly, for $i' = 1$, $\varphi_{i'-1} = \psi_{i'-1} = 0$, and we also have $\varphi_{i'} > 0$, which contradicts $\hat{k}_{i'+1} > \hat{k}_{i'}$. Therefore, for all $i \in \mathcal{I}$, $\hat{k}_i \geq k_i^*(\hat{v}_i)$ and $\hat{n}_i \geq n_i^*(\hat{v}_i)$ must hold. □

Claim 26. *$\hat{k}_i = k_i^*(\hat{v}_i)$ and $\hat{n}_i = n_i^*(\hat{v}_i)$ for $i = 1$ or $i \in \mathcal{I} \setminus \{1\}$ with $\psi_{i-1} = \varphi_{i-1} = 0$.*

Proof. First, consider the case $\hat{n}_i = b$. Since $\hat{v}_i = 0$ and $n_i^*(0) = b$, it remains to show that $\hat{k}_i = k_i^*(0) = 0$. Suppose that $\hat{k}_i > 0$. Then, for $i = 1$ or $i \in \mathcal{I} \setminus \{1\}$ with $\psi_{i-1} = \varphi_{i-1} = 0$, (55) is

$$(f_i + \psi_i)\mathcal{V}_k(\hat{k}_i, \hat{n}_i, a_i) = \varphi_i + (1+\lambda) f_i.$$

Since $\mathcal{V}_k(\hat{k}_i, \hat{n}_i, a_i) = [(b - \hat{n}_i)/y(\hat{k}_i, a_i)]^{1/2} y_k(\hat{k}_i, a_i) = 0$ and $(1+\lambda) f_i > 0$, this leads to a contradiction. Therefore, $\hat{k}_i = 0$ must hold.

Next, consider the case $\hat{n}_i < b$. For $i = 1$ or $i \in \mathcal{I} \setminus \{1\}$ with $\psi_{i-1} = \varphi_{i-1} = 0$, (57) implies

$$\varphi_i = (1+\lambda) f_i \left(\frac{\mathcal{V}_k(\hat{k}_i, \hat{n}_i, a_i)}{|\mathcal{V}_n(\hat{k}_i, \hat{n}_i, a_i)|} - 1 \right) \leq 0,$$

where the inequality follows from Claim 25. Since $\varphi_i \geq 0$, $\varphi_i = 0$ must hold and the above equation implies $\hat{k}_i = k_i^*(\hat{v}_i)$ and $\hat{n}_i = n_i^*(\hat{v}_i)$. □

Therefore, private information distorts allocation upward from efficiency, except for the lowest type. This reflects that individuals have incentives to over-report their type. That is, all constraints of (AUIC-3) hold as equalities.

Claim 27. *For all $i \in \mathcal{I} \setminus \{I\}$, $\mathcal{V}(\hat{k}_i, \hat{n}_i, a_i) = \mathcal{V}(\hat{k}_{i+1}, \hat{n}_{i+1}, a_i)$.*

Proof. Suppose that there exists $i \in \mathcal{I} \setminus \{I\}$ such that $\mathcal{V}(\hat{k}_i, \hat{n}_i, a_i) > \mathcal{V}(\hat{k}_{i+1}, \hat{n}_{i+1}, a_i)$. Because (AUIC-3) for i is slack, $\psi_i = 0$, which gives (58). Moreover, by Claim 22, $\hat{k}_i < \hat{k}_{i+1}$ must hold, because $\hat{k}_i = \hat{k}_{i+1}$ would imply $\hat{n}_i = \hat{n}_{i+1}$ and $\mathcal{V}(\hat{k}_i, \hat{n}_i, a_i) = \mathcal{V}(\hat{k}_{i+1}, \hat{n}_{i+1}, a_i)$. Therefore, we have $\psi_i = \varphi_i = 0$, which implies $\hat{k}_{i+1} = k_{i+1}^*(\hat{v}_{i+1})$ and $\hat{n}_{i+1} = n_{i+1}^*(\hat{v}_{i+1})$ by Claim 26. $\hat{k}_i < \hat{k}_{i+1}$ and $\mathcal{V}(\hat{k}_i, \hat{n}_i, a_i) > \mathcal{V}(\hat{k}_{i+1}, \hat{n}_{i+1}, a_i)$ also imply $n_i < n_{i+1}$.

Suppose that $\hat{n}_{i+1} = b$. Then $(\hat{k}_{i+1}, \hat{n}_{i+1}) = (k_{i+1}^*(\hat{v}_{i+1}), n_{i+1}^*(\hat{v}_{i+1})) = (k_{i+1}^*(0), n_{i+1}^*(0))$. Moreover, $|\mathcal{V}_n(k, n, a)|$ is continuous in (k, n, a), and $k_i^*(v)$ and $n_i^*(v)$ are continuous in v by Claim 24.

Therefore,

$$
\begin{aligned}
|\mathcal{V}_n(\hat{k}_{i+1}, \hat{n}_{i+1}, a_{i+1})| =& \lim_{v \downarrow 0} |\mathcal{V}_n(k^*_{i+1}(v), n^*_{i+1}(v), a_{i+1})| \\
=& \lim_{v \downarrow 0} \left(\frac{y(k^*_{i+1}(v), a_{i+1})}{b - n^*_{i+1}(v)} \right)^{\frac{1}{2}} \\
=& \lim_{v \downarrow 0} \left(y_k(k^*_{i+1}(v), a_{i+1}) \right)^{\frac{1}{2}} \\
=& \infty,
\end{aligned}
$$

where the third equality is due to (59) and the last equality follows from $\lim_{v \downarrow 0} k^*_{i+1}(v) = 0$. Therefore, (58) cannot be satisfied, as $n_i < n_{i+1} = b$ implies $|\mathcal{V}_n(\hat{k}_i, \hat{n}_i, a_i)| < \infty$.

Next, suppose that $\hat{n}_{i+1} < b$. If $b \geq \hat{n}_{i+1} + y(\hat{k}_{i+1}, a_{i+1})$, then $b \geq \hat{n}_i + y(\hat{k}_i, a_i)$ also holds, because $\hat{k}_i < \hat{k}_{i+1}$ and $\hat{n}_i < \hat{n}_{i+1}$. Therefore, $|\mathcal{V}_n(\hat{k}_i, \hat{n}_i, a_i)| = 1$, which contradicts (58). Thus, $b < \hat{n}_{i+1} + y(\hat{k}_{i+1}, a_{i+1})$ must hold. In this case, $|\mathcal{V}_n(\hat{k}_{i+1}, \hat{n}_{i+1}, a_{i+1})| > 1$, which also implies $|\mathcal{V}_n(\hat{k}_i, \hat{n}_i, a_i)| > 1$ due to (58). Therefore, $b < \hat{n}_i + y(\hat{k}_i, a_i)$. Because $b < \hat{n}_{i+1} + y(\hat{k}_{i+1}, a_{i+1})$ and $b < \hat{n}_i + y(\hat{k}_i, a_i)$, (58) can be written as

$$
\left(\frac{y(\hat{k}_{i+1}, a_{i+1})}{b - \hat{n}_{i+1}} \right)^{\frac{1}{2}} \leq 1 + \lambda \leq \left(\frac{y(\hat{k}_i, a_i)}{b - \hat{n}_i} \right)^{\frac{1}{2}}, \tag{64}
$$

which cannot hold due to $\hat{k}_i < \hat{k}_{i+1}$ and $\hat{n}_i < \hat{n}_{i+1}$. □

The fact that the constraints (AUIC-3) hold as equalities implies that (DIC-3) is satisfied. □

Therefore, Problems 9 and 10 are identical. I continue to characterize the solution to Problem 10, $(\hat{k}_i, \hat{n}_i)_{i \in \mathcal{I}}$.

Lemma 22. *$\hat{k}_i - \hat{n}_i < \hat{k}_{i+1} - \hat{n}_{i+1}$ for all $i \in \mathcal{I} \setminus \{I\}$ such that $\hat{k}_i < \hat{k}_{i+1}$.*

Proof. Consider $i \in \mathcal{I} \setminus \{I\}$ such that $\hat{k}_i < \hat{k}_{i+1}$. Since $\hat{k}_i - \hat{n}_i < \hat{k}_{i+1} - \hat{n}_{i+1}$ holds trivially if $\hat{n}_i = \hat{n}_{i+1}$, consider the case $\hat{n}_i < \hat{n}_{i+1}$. Then, by Claim 27, $\hat{v}_i = \mathcal{V}(\hat{k}_i, \hat{n}_i, a_i) = \mathcal{V}(\hat{k}_{i+1}, \hat{n}_{i+1}, a_i) > 0$, where the inequality follows from $b \geq \hat{n}_{i+1} > \hat{n}_i$. Define $\mathcal{N}(k, v, a)$ to be a function satisfying $v = \mathcal{V}(k, \mathcal{N}(k, v, a), a)$ and notice that $\mathcal{N}_k(k, v, a) \equiv \partial \mathcal{N}(k, v, a)/\partial k = \mathcal{V}_k(k, \mathcal{N}(k, v, a), a)/|\mathcal{V}_n(k, \mathcal{N}(k, v, a), a)|$. Then $\hat{n}_i = \mathcal{N}(\hat{k}_i, \hat{v}_i, a_i)$ and $\hat{n}_{i+1} = \mathcal{N}(\hat{k}_{i+1}, \hat{v}_i, a_i)$ hold and, by the fundamental theorem of calculus, we have

$$
\begin{aligned}
\hat{n}_{i+1} - \hat{n}_i =& \int_{\hat{k}_i}^{\hat{k}_{i+1}} \mathcal{N}_k(k, \hat{v}_i, a_i) dk \\
=& \int_{\hat{k}_i}^{\hat{k}_{i+1}} \frac{\mathcal{V}_k(k, \mathcal{N}(k, \hat{v}_i, a_i), a_i)}{|\mathcal{V}_n(k, \mathcal{N}(k, \hat{v}_i, a_i), a_i)|} dk \\
<& \int_{\hat{k}_i}^{\hat{k}_{i+1}} \frac{\mathcal{V}_k(\hat{k}_i, \hat{n}_i, a_i)}{|\mathcal{V}_n(\hat{k}_i, \hat{n}_i, a_i)|} dk \\
\leq& \hat{k}_{i+1} - \hat{k}_i.
\end{aligned}
$$

The first inequality holds because $\mathcal{V}_k(k,n,a)/|\mathcal{V}_n(k,n,a)|$ is decreasing in n and strictly decreasing in k (if $n < b$), and $\mathcal{N}(k,v,a)$ is strictly increasing in k (if $v > 0$). The second inequality holds due to Claim 25. □

Combining Lemma 20, Claim 22, and Lemma 22 gives the result.

B.15 Proof of Proposition 3

I first calculate several derivatives that are useful to derive the first order conditions. By differentiating $p_j(b;\boldsymbol{g})$ with respect to g_j for $j \in \mathcal{J}(b;\boldsymbol{g})$, we get

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

where I used the fact that a small change in g_j does not affect the set of feasible options.

Similarly, the derivative of $p_j(b;\boldsymbol{g})$ with respect to $g_{j'}$ for $j' \in \mathcal{J}(b;\boldsymbol{g}) \setminus \{j\}$ is

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

The derivative of aggregate spending with respect to g_j is

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

Finally, the derivative of $V(b;\boldsymbol{g})$ with respect to g_j is

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

The part of the Lagrangian for Problem 3 that depends on $\boldsymbol{g}_h$ is

$$\sum_{h\in\mathcal{H}} f_h \left\{ V(b_h;\boldsymbol{g}_h) - (1+\lambda) \sum_{j\in\mathcal{J}(b_h;\boldsymbol{g}_h)} p_j(b_h;\boldsymbol{g}_h) g_{j,h} \right\}.$$

Part (ii): The first order condition for $g_{j,h}$ is

$$p_j(b_h;\hat{\boldsymbol{g}}_h)V_j'(b_h+\hat{g}_{j,h}) = (1+\lambda)p_j(b_h;\hat{\boldsymbol{g}}_h)\left[1+\left(\hat{g}_{j,h} - \sum_{j'\in\mathcal{J}(b_h;\hat{\boldsymbol{g}}_h)} p_{j'}(b_h;\hat{\boldsymbol{g}}_h)\hat{g}_{j',h}\right)V_j'(b_h+\hat{g}_{j,h})\right].$$

By rearranging the above condition, we get (14).

Part (i): Consider $(j,j')\in\mathcal{J}(b_h;\hat{\boldsymbol{g}}_h)\times\mathcal{J}(b_h;\hat{\boldsymbol{g}}_h)$ such that $j'>j$. From (14), we have

$$\hat{g}_{j,h} + \frac{1}{V_j'(b_h+\hat{g}_{j,h})} = \hat{g}_{j',h} + \frac{1}{V_{j'}'(b_h+\hat{g}_{j',h})} \le \hat{g}_{j',h} + \frac{1}{V_j'(b_h+\hat{g}_{j',h})},$$

where the inequality holds because Lemma 2 suggests that $V_j(b)$ is increasing in j. Since Lemma 2 also implies that $V_j'(b)$ is decreasing in b, we have $\hat{g}_{j',h}\ge\hat{g}_{j,h}$.

B.16 Optimality of Full Redistribution Across Groups

Full redistribution across groups according to (17) may or may not be optimal. As explained in Footnote 22, it is because of the non-concavities in the indirect utility function induced by discrete schooling choice. This subsection shows that sufficient heterogeneity in returns to schooling eliminates such con-concavities in the indirect utility function. This idea was first suggested by Gomes, Greenwood, and Rebelo (2001).

As described in Section 4, let $\sigma\in\mathbb{R}_{++}$ be the scale parameter for the distribution of returns to schooling that governs the degree of heterogeneity. Then the schooling choice problem is

$$\max_{j\in\mathcal{J}(b;\boldsymbol{g})} \left\{V_j(b+g_j)+\sigma\varepsilon_j\right\},$$

which gives

$$p_j(b;\boldsymbol{g},\sigma) = \frac{\exp\left(V_j(b+g_j)/\sigma\right)}{\sum_{j'\in\mathcal{J}(b;\boldsymbol{g})}\exp\left(V_{j'}(b+g_j)/\sigma\right)}, \quad \forall j\in\mathcal{J}(b;\boldsymbol{g}),$$

$$V(b;\boldsymbol{g},\sigma) = \sigma\ln\left(\sum_{j\in\mathcal{J}(b;\boldsymbol{g})}\exp\left(V_j(b+g_j)/\sigma\right)\right)+\sigma\rho.$$

The following provides a condition for the indirect utility function $V(\cdot,\boldsymbol{g},\sigma)$ to be concave.

Condition 1. *For all $b\in\mathbb{R}_+$ where $V_j'(\cdot)$ is differentiable,*

$$-\sigma E(V_j''(b)|b,\sigma) \ge Var\left(V_j'(b)|b,\sigma\right),$$

where the conditional mean and variance are taken over $j \in \mathcal{J}(b;\mathbf{0})$, using probabilities $(p_j(b;\mathbf{0},\sigma))_{j\in\mathcal{J}(b;\mathbf{0})}$.

This condition is satisfied for a large value of σ because $p_j(b;\mathbf{0},\sigma) \to 1/|\mathcal{J}(b;\mathbf{0})|$ as $\sigma \to \infty$. The concavity of the indirect utility function implies that it is optimal to fully compensate for the differences in parental transfers across groups.

Corollary 2. *Suppose that Assumption 5 and Condition 1 hold. Then $(\hat{\boldsymbol{g}}_h)_{h\in\mathcal{H}}$ that satisfies* (13), (14)*, and* (17) *solves Problem 3.*

Proof. I first show that $V(b;\mathbf{0},\sigma)$ is concave in b if Condition 1 holds. The first derivative of $V(b;\mathbf{0},\sigma)$ with respect to b is

$$V'(b;\mathbf{0},\sigma) = \sum_{j\in\mathcal{J}(b;\mathbf{0})} p_j(b;\mathbf{0},\sigma)V_j'(b),$$

and the second derivative (where $V_j(b)$ is differentiable) is

$$\begin{aligned}
&V''(b;\mathbf{0},\sigma)\\
=&\sum_{j\in\mathcal{J}(b;\mathbf{0})} p_j'(b;\mathbf{0},\sigma)V_j'(b) + p_j(b;\mathbf{0},\sigma)V_j''(b)\\
=&\frac{1}{\sigma}\left\{\sum_{j\in\mathcal{J}(b;\mathbf{0})} p_j(b;\mathbf{0},\sigma)[V_j'(b)]^2 - \left[\sum_{j\in\mathcal{J}(b;\mathbf{0})} p_j(b;\mathbf{0},\sigma)V_j'(b)\right]^2\right\} + \sum_{j\in\mathcal{J}(b;\mathbf{0})} p_j(b;\mathbf{0},\sigma)V_j''(b)\\
=&\frac{\mathrm{Var}\,(V_j'(b)|b,\sigma)}{\sigma} + \mathrm{E}(V_j''(b)|b,\sigma),
\end{aligned}$$

which is negative if and only if Condition 1 holds. Because $V(\cdot;\mathbf{0},\sigma)$ is continuously differentiable and has a negative second derivative almost everywhere, it is concave.

Since $V(b;\mathbf{0},\sigma)$ is concave in b, Jensen's inequality implies that it is optimal to fully redistribute parental transfers across groups. Then, after the redistribution, all individuals will have an identical parental transfer that is equal to the average amount, $\tilde{b} \equiv \sum_{h\in\mathcal{H}} f_h b_h$. Since the redistribution makes all groups identical, the planning problem amounts to setting a single subsidy schedule, $\tilde{g} \equiv (\tilde{g}_j)_{j\in\mathcal{J}}$. The solution to this two-step optimization problem can be implemented by setting a subsidy schedule such that $\hat{g}_{j,h} = \tilde{g}_j + (\tilde{b} - b_h)$ holds. Therefore, (17) is satisfied. □

B.17 Proof of Corollary 1

The part of the Lagrangian for Problem 3 that depends on $\boldsymbol{g}_h$ is

$$\sum_{h\in\mathcal{H}}\sum_{i\in\mathcal{I}_h} f_i \left\{V(b_i;\boldsymbol{g}_h) - (1+\lambda)\sum_{j\in\mathcal{J}} p_j(b_i;\boldsymbol{g}_h)g_{j,h}\right\}.$$

The first order condition for $g_{j,h}$ is

$$\sum_{i\in\mathcal{I}_h} f_i p_j(b_i;\hat{\boldsymbol{g}}_h)\left\{V_j'(b_i+\hat{g}_{j,h})-(1+\lambda)\left[1+\left(\hat{g}_{j,h}-\sum_{j'\in\mathcal{J}} p_{j'}(b_i;\hat{\boldsymbol{g}}_h)\hat{g}_{j',h}\right)V_j'(b_i+\hat{g}_{j,h})\right]\right\}=0.$$

Rearranging it gives (18).

B.18 Proof of Lemma 4

When (21) does not bind, $V_j(z;a)=y_j(a)-k_j+z$. Therefore, $V_j'(z;a)=1$ holds. When (21) binds,

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

where the inequality holds if and only if

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

This holds because the annual consumption during schooling is lower than the annual consumption after schooling when (21) binds.

Since $u(c)$ is increasing and $u'(c)$ is decreasing in c, $V_j'(z;a)$ is increasing in a if and only if $U_j(z;a)$ is increasing in a (when (21) binds), which is equivalent to the following condition:

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

where the inequality holds because $y_j(a)$ is increasing in a.

Finally, when (21) does not bind, the amount of borrowing can be written as

$$\sum_{t=1}^{j}R^{-t}c_t+k_j-z=\sum_{t=1}^{j}R^{-t}\frac{y_j(a)-k_j+z}{\sum_{t=1}^{T_k}R^{-t}}+k_j-z, \tag{65}$$

which is increasing in a. This completes the proof that $V_j'(z;a)$ is increasing in a.

B.19 Proof of Proposition 4

First, note that (22) and (23) are given by

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

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

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

$$\frac{\partial p_j(\boldsymbol{b};\boldsymbol{g},a)}{\partial b_j} = p_j(\boldsymbol{b};\boldsymbol{g},a)\left[1-p_j(\boldsymbol{b};\boldsymbol{g},a)\right]\frac{V_j'(b_j+g_j;a)}{\sigma}.$$

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

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

The derivative of average parental transfer with respect to b_j is

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

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

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

Therefore, the first order condition for an interior solution is

$$(1-\delta)v'(\hat{c}_p)\left\{1+\left(\hat{b}_j-\sum_{j'\in\mathcal{J}(\hat{\boldsymbol{b}};\boldsymbol{g})}p_{j'}(\hat{\boldsymbol{b}};\boldsymbol{g},a)\hat{b}_{j'}\right)\frac{V_j'(\hat{b}_j+g_j;a)}{\sigma}\right\} = \delta v'(\hat{c}_k)V_j'(\hat{b}_j+g_j;a).$$

By rearranging the above condition, we get (27).

B.20 Distortions in Quantitative Model

Conditional on $(\boldsymbol{b},\boldsymbol{g},a)$, the average consumption distortion is

$$\tau_c(\boldsymbol{b};\boldsymbol{g},a) \equiv \sum_{j\in\mathcal{J}(\boldsymbol{b};\boldsymbol{g})}p_j(\boldsymbol{b};\boldsymbol{g},a)\left[y_j(a)-k_j+b_j+g_j-V_j(b_j+g_j;a)\right]$$

and the average schooling distortion is

$$\tau_s(\boldsymbol{b};\boldsymbol{g},a) = \int \max_{j\in\mathcal{J}} \left\{y_j(a) - k_j + \mu_j + \sigma\varepsilon_j\right\} dF(\boldsymbol{\varepsilon}) - \\ \sum_{j\in\mathcal{J}} \int \mathbb{I}_{j\in\operatorname{argmax}_{j'\in\mathcal{J}(\boldsymbol{b};\boldsymbol{g})}\{V_{j'}(b_{j'}+g_{j'};a)+\mu_{j'}+\sigma\varepsilon_{j'}\}} \left[y_j(a) - k_j + \mu_j + \sigma\varepsilon_j\right] dF(\boldsymbol{\varepsilon}). \quad (66)$$

When the social budget constraint binds, the social planner's objective function is

$$\begin{aligned}
&\sum_{h\in\mathcal{H}} \sum_{i\in\mathcal{I}_h} f_i \tau(\boldsymbol{b}_i;\boldsymbol{g}_h,a_i) \\
=&\sum_{h\in\mathcal{H}} \sum_{i\in\mathcal{I}_h} f_i \left[\int \max_{j\in\mathcal{J}} \left\{y_j(a_i) - k_j + \mu_j + \sigma\varepsilon_j\right\} dF(\boldsymbol{\varepsilon}) + \sum_{j\in\mathcal{J}} p_j(\boldsymbol{b}_i;\boldsymbol{g}_h,a_i)(b_{j,i}+g_{j,h}) - V(\boldsymbol{b}_i;\boldsymbol{g}_h,a_i)\right] \\
=&G + \sum_{h\in\mathcal{H}} \sum_{i\in\mathcal{I}_h} f_i \left[\int \max_{j\in\mathcal{J}} \left\{y_j(a_i) - k_j + \mu_j + \sigma\varepsilon_j\right\} dF(\boldsymbol{\varepsilon}) + \sum_{j\in\mathcal{J}} p_j(\boldsymbol{b}_i;\boldsymbol{g}_h,a_i)b_{j,i} - V(\boldsymbol{b}_i;\boldsymbol{g}_h,a_i)\right].
\end{aligned}$$

Therefore, minimizing aggregate distortions is equivalent to maximizing

$$\sum_{h\in\mathcal{H}} \sum_{i\in\mathcal{I}_h} f_i \left[w_i - \sum_{j\in\mathcal{J}} p_j(\boldsymbol{b}_i;\boldsymbol{g}_h,a_i)b_{j,i} + V(\boldsymbol{b}_i;\boldsymbol{g}_h,a_i)\right].$$

From the definitions of c_p and c_k in (24) and (25), we have

$$w - \sum_{j\in\mathcal{J}} p_j(\boldsymbol{b}_i;\boldsymbol{g}_h,a_i)b_j + V(\boldsymbol{b}_i;\boldsymbol{g}_h,a_i) = \sum_{t=1}^{T_p} R^{-t} c_{p,i} + \sum_{t=1}^{T_k} R^{-t} c_{k,i}.$$

C Details on Calibration

C.1 Monetary Returns to, and Costs of, Schooling

Table C.1: OLS Estimates of the Earnings Function

	Log Annual Earnings
Highest year of college completed:	
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

Notes: Standard errors in parentheses.

* Significant at the 5% level.

Table C.2: Present Discounted Value of Lifetime Earnings at Age 17 ($)

Schooling Choice	$y_j(a)$ 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 C.3: Average Annual Amounts for Each Year of College ($)

Year	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 C.4: Present Discounted Values at Age 17 ($)

Schooling Choice	Cost (k_j)	Subsidy (g_j) for Income Quartiles				Borrowing Limit ($\overline{d}_j$)
		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

C.2 Discussion: Sources of Inequality in Parental Transfers

While the heterogeneity in parental altruism is a modeling device to capture any unobserved differences across families that lead to differences in parental transfers, there is some evidence that is consistent with it. Brown, Scholz, and Seshadri (2012) provide direct evidence based on self-reported measures designed to elicit parents' altruism toward their children. The HRS respondents are asked whether they would be willing to give 5% of their own family income to their child if the income of the child were a certain fraction (1/3, 1/2, or 3/4) of their own income. While 9% of the respondents say they would not give under any circumstances, 62% say they would give when the child's income is 3/4 of their own income. As the transfer amounts are expressed as fractions of income in the questionnaire, these differences in answers are likely to reflect differences in altruism rather than income. Consistent with the prediction of their model with borrowing constraints, they also find that an exogenous increase in financial aid positively affects children's education outcomes only for parents with low willingness to give (i.e., low measured altruism).

Evidence presented by Sallie Mae (2012) also suggests preference heterogeneity across families regarding who should pay for children's college education. American college students and their parents are asked about who they think should be responsible for financing the students' college education. While higher-income families tend to believe that parents are more responsible than children, the answers vary substantially even among families with similar income. For low-income families making less than $35,000, 7% think parents should be entirely responsible, while 18% think it is children who should be entirely responsible. Among high-income families making more than $100,000, 14% think parents should be entirely responsible, while 9% think children should be entirely responsible.

Of course, there are other explanations for heterogeneous parental support conditional on EFC presented in Section 2. It might reflect unobserved differences in family resources because the EFC may not correctly measure the true financial situation of the family. For example, a single year's parental income might be a poor measure of lifetime income and, as mentioned earlier, the EFC calculation does not include housing and retirement wealth of parents.[53] The differences in parental support might also represent differences in loans rather than gifts provided by parents. While one-sided altruism is a common assumption in the literature, it is possible that altruism is two-sided so that children also care about their parents. In this case, there might be equilibria where children support parents later in return for parental support during schooling. Although these are also plausible explanations, to the extent that parental support for college education helps reduce the distortions due to borrowing constraints, the exact nature and origin of differences in parental support are unlikely to have significant enough policy implications.

[53]Furthermore, the EFC formula may overstate the importance of some variables. For example, students with siblings enrolled in college in the same year have lower EFC than others without the overlap. Brown, Scholz, and Seshadri (2012) use this variation in potential financial aid induced by birth spacing of siblings to test the implications of borrowing constraints.

C.3 Parental Transfers from the NLSY97 Data

The NLSY97 survey contains two sections in which questions about parental transfers are asked. The "college experience" section collects information about how much money youth received from parents to pay for college and the "income" section asks about the money youth received as part of their income in each year.

C.3.1 College Experience Section

In every interview, the college experience section starts with question YSCH-24991, which asks if the respondent attended college since the date of last interview (DLI). For each college and each term the respondent attended since the DLI, the survey asks, among other things, for the start date of the term (YSCH-20400) and how the respondent paid for college. The structure of questions on college financing differs depending on whether it is the first time the respondent is asked about the college in each round (YSCH-22004). For the first term of each college in each round, the respondent is first asked whether she received financial aid from biological parents, mother (and stepfather), father (and stepmother), grandparents, and other relatives and friends (YSCH-23900) and, if so, the amount of financial aid received from each of them in the form of gift (YSCH-24600) and loan (YSCH-24700). The exact wording of YSCH-24600 is as follows:

> Altogether, how much [have/has/did] your family and friends [give/given] you in gifts or other money you are not expected to repay to help pay for your attendance at this school/institution during this term?

and the wording for YSCH-24700 is as follows:

> Altogether, how much [have/has/did] your family and friends [loan/loaned] you to help pay for your attendance at this school/institution during this term?

From the second term for each college in each round, the respondent is asked whether there were any changes in college financing since the previous term (YSCH-22005), and no further questions are asked if there were no changes. In this case, I fill in this information using answers from previous terms. If there were changes, then the respondent is first asked whether she received any financial assistance from family (YSCH-22006), and if she did, the further questions on aid from family (YSCH-23900, YSCH-24600, and YSCH-24700) are asked.

In round 1, the structure of survey is a little bit different from what is described above. I do not use data from the first round of the survey because financial aid questions are not asked for each term. This is likely to have very little effect because only eight respondents attended college during the first round of the survey, which was administered in 1997.

C.3.2 Income Section

The income section of the survey collects wage and salary data for the past calendar year from all respondents. Those who are considered independent answer more extensive questions about other sources of income, including parental transfers. Since one of the criteria for independence is reaching the age of 18, I measure parental transfers youth received when they are more than or equal to 18 years old.[54]

For rounds 1–7 (survey years 1997–2003), respondents state the amount of money they received from (i) both parent figures, (ii) the mother figure, and (iii) the father figure. Youth who live with both parent figures (YINC-5600) are asked whether they received any money from them in the previous year (YINC-5700), and if so, how much (YINC-5800). In round 2, the exact wording of YINC-5700 is as follows:

> Other than an allowance, did your parents give [you/you or your spouse] any money during 1997? Please include any gifts in the form of cash or a check but do not include any loans from your parents.

The wording for YINC-5800 is as follows:

> How much did your parents give [you/you and your spouse] during 1997?

Those who do not know the answer or refuse to answer question YINC-5800 are again asked to provided answers in categories (YINC-5900).

Those who live with a mother/father figure or whose biological mother/father is alive (YINC-6400/YINC-7000), including those who live with both parent figures, are also asked similar questions about transfers received from their mother figure (YINC-6500, YINC-6600, and YINC-6700) and father figure (YINC-7100, YINC-7200, and YINC-7300).

For rounds 8–15 (survey years 2004–2011), youth are first asked whether they received money from family or friends (YINC-5700A), and if so, their amounts in categories (YINC-5900A). These questions are discontinued after round 15. Therefore, parental transfers can be measured only through calendar year 2010.

Finally, for all rounds, youth are asked whether they received money from inheritances (YINC-5200), and if so, their amounts in numbers (YINC-5300) or categories (YINC-5400). In round 2, the exact wording of YINC-5200 is as follows:

> During 1997, did [you/you or your spouse/partner] receive any property or money from any estates, trusts, annuities or inheritances?

And the wording of YINC-5300 is as follows:

> What was the total market value or amount that [you/you and your spouse/partner] received during 1997 from these sources?

C.3.3 Total Parental Transfers

For each calendar year, I construct annual parental transfers by adding all forms of college financial aid from all family members (YSCH-24600 and YSCH-24700) during all terms that started in that year to the amount of money received from family (YINC-5800/5900, YINC-6600/6700, YINC-7200/7300,

[54]Youth are also considered to be independent if they have had a child, are enrolled in a four-year college, are no longer enrolled in school, are not living with any parents or parent-figures, or have ever been married or are in a marriage-like relationship at the time of the survey.

and YINC-5900A) and inheritances (YINC-5300/5400). The variables provided in categories (YINC-5900, YINC-6700, YINC-7300, YINC-5400, and YINC-5900A) are turned into amounts by using the mid-point of each category or the minimum value for the highest category. Based on the annual parental transfers, I compute total parental transfers youth received between the years 1998 and 2010 and the ages 18 and 26, discounted back to age 17.

There are several issues related to constructing the total parental transfer variable. First, as discussed by Abbott et al. (forthcoming), it might be incorrect to sum all transfers from different sections of the survey, because they are not necessarily mutually exclusive categories. For example, respondents might consider the college financial aid from family (recorded in the college experience section) a part of their income received from family (reported in the income section). Therefore, one could take the maximum between the two instead of adding them in order to reduce the possibility of double counting. However, the two methods yield very similar results for the target statistics reported in Table 5c (results available upon request). Second, based on the discussion in Appendix C.2 that it is irrelevant whether youth repay the parental transfers later or not, I include loans from family (YSCH-24700) as well as gifts. Third, unlike Johnson (2013) and Abbott et al. (forthcoming), I do not include the monetary value of living with parents to parental transfers because it does not necessarily reflect parental support that helps youth attend college. Of course, students living at home while enrolled can save considerably on room and board costs, but this is useful only if there is a college that youth wish to attend within commuting distance from home.[55] Moreover, parents who are not willing give money for college might still let their children to stay with them because the additional cost of doing so would be small as a result of economies of scale. Therefore, it is possible that youth who do not live near a college stay at home because they do not have enough money to leave home and attend college elsewhere. Another reason for not including parental co-residence is that obtaining reliable information about it is difficult after round 6 because, as noted by Kaplan (2012), one must rely on the household roster that may refer to what youth consider their primary residence rather than their current residence. For example, college students who live away from home during the school year may still report to be living at home.

C.4 Details on Calibrating Preference Parameters

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, educational attainment and AFQT score of youth, and amount of parental transfer) are missing. To better fit the schooling distribution, the fractions choosing each schooling option are multiplied by 10, which is equivalent to putting 100 times higher weights on them compared with all other statistics.

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 child, which does not

[55] Do (2004) notes that about half of the high school sophomore class of 1980 in the US did not have a public university in their county of residence.

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 (one, two, or more than or equal to three) as well as the number of parents (more than one), assuming that these variables affect the levels of outcome, but not their gradients. Second, I assume a constant stream of parental income to compute the remaining discounted value of lifetime income of parents from 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.