From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Economics

2016

Essays in Public Economics

OLEA DE SOUZA E SILVA, PEDRO MIGUEL

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01wd375z71q

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

Essays in Public Economics

Pedro Miguel Olea de Souza e Silva

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Economics
Adviser: Professor Mikhail Golosov

May 2016

© Copyright by Pedro Miguel Olea de Souza e Silva, 2016.

All rights reserved.

Abstract

This dissertation is comprised of three separate chapters that study the optimal design of different government policies. In the first chapter, which is joint work with Christian Moser[1], I study the optimal design of retirement savings policies and how under-saving for retirement can be dealt with by the government when there is also considerable income inequality in the economy. I find that a redistributive government would find it optimal to provide more flexibility on savings of higher earners, thorugh the usage of special retirement savings accounts (401(k)s), and this government would find it optimal to offer social security benefits to make sure low earners have sufficient resources during retirement. In the second chapter, I study the optimal design of labor income taxes when there is advance information about the unobservable heterogeneity in labor earnings ability into the future. I find that the standard optimal taxation results are greatly affected, and in particular top income tax rates can differ significantly if the Pareto tail of the distribution of labor earnings ability differs across workers with different expectations. I find in the data from the Survey of Consumer Finances that there are significant differences on the Pareto tail of the distribution of earnings ability across different educational levels. Therefore, optimal income taxes across those groups differ significantly. Finally, in the third chapter, which is joint work with Michael Sockin[2], I study the optimal design of subsidies to higher education. I find that the optimal subsidy level satisfies a simple formula that depends on three key components: the redistributive motive; the extensive margin elasticity of education attainment with respect to the cost of education; and the elasticity of aggregate labor supply with respect to the cost of education.

[1]Princeton University.
[2]University of Texas at Austin.

Acknowledgements

I would like to thank Princeton University, the J. Wallace Ely Fund '32, the Griswold Center for Economic Policy Study, the International Economics Section at Princeton University, and the Department of Economics at Princeton University for the funding without which I would not be able to conduce my doctoral studies at Princeton.

Moreover, I would like to thank Mikhail Golosov for all the support, encouragement and generous advice during this process. I also would like to thank my examination and dissertation committees members Mark Aguiar, Roland Benabou, Alan Blinder and Nobuhiro Kiyotaki for their comments and suggestions. I also thank my collaborators Christian Moser and Michael Sockin for the great experience of working with them.

Also, I would like to thank my classmates at the doctoral program in Princeton University for the comments on my work and also for making the time at Princeton so enjoyable. In particular, I thank Nemanja Antic, Olivier Darmouni, Richard Katzwer, Ricardo Reyes-Heroles, Gabriel Tenorio, Juha Tolvanen, Sharon Traiberman and Dan Zeltzer. I also would like to thank Tom Winberry, Christian Vom Lehn, and other members of the Macroeconomics Student Lunch, the Public Finance Working Group and the Microeconomics Student Lunch for all the support during my dissertation. Furthermore, I want to thank Jen Bello and Laura Hedden for their incredible help during the job market process.

I would like to specially thank my parents, Sônia and Carlos, and my siblings Juliana, Fernando and Marcelo for the support during my entire life that was so essential for me. Finally, I want to thank Cinthia Konichi Paulo for all her support and love.

To my parents.

Contents

Abstract iii

Acknowledgements iv

List of Tables ix

List of Figures x

1 Paternalism vs Redistribution: Designing Retirement Savings Policies with Behavioral Agents **1**

1.1 Introduction 2

1.2 Optimal policy analysis 11

1.2.1 Example with 2×2 types 14

1.2.2 Main results 21

1.2.3 Comparison to most related results in the literature 25

1.3 Decentralization of optimal retirement savings policies 27

1.4 Quantitative exercise 30

1.4.1 Two-period calibration of current policy 30

1.4.2 Quantitative approximation to government problem 36

1.5 Extension: Behavioral and Pigouvian policies 44

1.5.1 Applications: Environmental and drug policies 47

1.6 Conclusion 49

2 Taxing a bright future? Advance information and optimal labor income taxation **52**

2.1 Introduction . . . 53

2.2 Model with linear taxation . . . 56

2.2.1 Example with two profiles . . . 57

2.2.2 Optimal linear taxes in general case . . . 61

2.3 Planner's problem and non-linear taxation . . . 63

2.3.1 Optimal non-linear labor wedges . . . 66

2.3.2 Top income tax rates . . . 68

2.3.3 Self-Selection and Taxation . . . 70

2.4 Empirical evidence on Pareto tails . . . 72

2.5 Conclusion . . . 76

3 How to Design Optimal Higher Education Subsidization Policy? When Schooling Creates Opportunities **78**

3.1 Introduction . . . 79

3.2 Education, Occupations and Inequality . . . 82

3.3 Model with Occupational Choice and Educational Level Choice . . . 86

3.3.1 Laissez-faire Economy . . . 87

3.3.2 Labor Income Taxation and Educational Level Choice . . . 89

3.4 Optimal Policy . . . 90

3.4.1 Extension: Non-Linear Labor Income Taxation and Education Subsidies . . . 95

3.5 Quantitative Example . . . 97

3.6 Conclusion . . . 98

A Appendix to Chapter 1 **99**

A.1 Proofs for Section 1.2 . . . 99

A.2 Extension: life-cycle model with hyperbolic preference shocks 107
A.3 Proofs of results for dynamic model with self-control shock 114
A.4 Proofs for Section 1.5 125
A.5 Life-cycle calibration with hyperbolic preferences 127

B Appendix to Chapter 2 **131**
B.1 Proofs 131
B.1.1 Proof of Theorem 7 131
B.1.2 Proof of Theorem 8 136
B.1.3 Proof of Theorem 9 139

C Appendix to Chapter 3 **141**
C.1 Empirical Analysis with Demographic Controls 141
C.2 Proofs 141
C.2.1 Proof of Theorem (13) 144
C.2.2 Optimal Income Tax 145
C.2.3 Optimal Educational Subsidies with Non-Linear Labor Income Taxes 146

Bibliography **148**

List of Tables

1.1 Calibration fit by life time earnings quantile. 34

1.2 Calibration results 35

1.3 Consumption equivalent welfare gains 43

2.1 Labor Income Distribution - Pareto Tail Estimates 73

2.2 Top Income Taxes for $\lim_{\theta\to\infty}\alpha(\theta)=0$ 75

3.1 Descriptive statistics for wage income in benchmark sample 83

3.2 Descriptive statistics of conditional labor income distribution 85

3.3 Educational Level by Occupational Category 86

3.4 Optimal Educational Subsidy Rates for $\tau=0.25$ 98

A.1 Life-Cycle Calibration Results 130

C.1 Descriptive statistics of conditional labor income distribution (with demographic controls 142

C.2 Liner Regression of Log-Wages on Demographics 142

List of Figures

1.1 Incentives with bunching. 16

1.2 Separation at high ability 17

1.3 Over consumption at retirement of low ability 19

1.4 Distribution of non-social security wealth at retirement ratio to lifetime earnings. 31

1.5 Proportion of retirement consumption for optimal allocations 40

1.6 Caps on retirement savings accounts 41

1.7 Tax rates on savings 42

1.8 Old-age benefits by labor earnings level 43

3.1 Distribution of labor income in benchmark sample 84

3.2 Income Distribution by Educational Level 85

Chapter 1

Paternalism vs Redistribution: Designing Retirement Savings Policies with Behavioral Agents

This paper develops a theory of optimal savings and redistributive policies when individuals under-save for retirement because of a behavioral bias. The two central features of our model are labor income inequality, arising from unobservable earnings ability differences, and heterogeneity in savings rates, due to unobservable degrees of present bias. The interaction between government's redistributive preferences and its paternalistic motive to correct savings leads to a novel insight: the optimal policy offers low income individuals a one-size-fits-all savings instrument, resembling social security, whereas it offers high income individuals a set of policies tailored to their heterogeneous preferences, similar to 401(k) and IRA accounts in the United States. The rationale for this policy is that the government uses flexibility at high earnings as a reward for generating income that can be taxed and used for redistribution. In a quantitative exercise, we use our normative model to evaluate the current U.S. social security and tax-transfer system. We find the current system to be inefficient, inde-

pendent of redistributive preferences. Relative to the utilitarian benchmark, current social security benefits are consistent with more progressive social preferences, while the tax-transfer system suggests lower progressivity. We explore the implications of our theory for other behavioral contexts as well as for non-behavioral Pigouvian tax policies. [1]

1.1 Introduction

With a budget of more than $700 billion dollars in 2014, the social security survivors and old-age benefits program is the largest U.S. federal government social policy. The public finance literature has suggested this program is a paternalistic policy that aims at helping individuals save for retirement [26, 64, 33]. The large body of evidence for present bias and heterogeneity in present bias reinforces the importance of corrective retirement savings policies, while the rise in labor income inequality in the U.S. has brought redistributive policies to the center of the policy discussion in recent years.[2]An under appreciated fact by both economists and policy makers is that savings policy and redistributive policy are likely to interact in non-trivial ways. For example, a policy that correct savings might reduce incentives for working as behavioral workers cannot allocate consumption the way they think is best, making

[1]This Chapter is based on a co-authored project with Christian Moser from Princeton University. We want to thank Mike Golosov for invaluable advice, guidance and encouragement throughout this project. We would also like to thank Dilip Abreu, Roland Bénabou, Olivier Darmouni, Henrique De Oliveira, Oleg Itskhoki, Cinthia Konichi Paulo, Ilyana Kuziemko, Ben Moll, Stephen Morris, Wolfgang Pesendorfer, Richard Rogerson, Juan Pablo Xandri and Sharon Traiberman. We also thank specially Mark Aguiar and Nobu Kiyotaki for detailed comments on the draft. We benefited from comments and suggestions by seminar participants at the Princeton Macro Student Workshop, the Princeton Microeconomic Theory Student Lunch Seminar, the Princeton Public Finance Working Group, Syracuse University Maxwell School Of Public Policy, Federal Reserve Bank of Atlanta, FGV-SP (EESP), FGV-Rio (EPGE), PUC-Rio, Insper, the University of São Paulo Economics Department, the 2014 Conference of German Economists Abroad at Kiel IfW.

[2]See [90], [76], [11], [57] and [18] for recent findings both on the presence of behavioral biases and its considerable heterogeneity. [80] and [9] document the increase in labor earnings and wealth inequality in the United States.

it harder for the government to redistribute. Thus understanding these interactions is of key importance to the public economics and behavioral economics literature.

In this paper, we develop a two-period model with unobservable heterogeneity in labor earnings ability and in the level of present bias to study the interaction between policies aiming at correcting savings choices and policies aiming at reducing economic inequality.[3] Our main finding concerns the shape of optimal retirement savings policies across different earnings levels when the government has a redistributive motive. At low earnings, policy forces individuals with the same ability but different present bias to save at the same rate. Thus optimal policy is paternalistic and leaves little flexibility for agents in savings decisions. At high earnings, optimal policy is less paternalistic and it offers individuals with the same ability but different present bias more flexibility on their savings choices. As a result, at low earnings optimal policy resembles forced savings through social security and at high earnings there are tailored subsidies or taxes on savings that resemble the availability of multiple individual retirement savings accounts, similar to 401(k)s or IRAs in the United States.

Three key forces in the model interact to generate our main finding. First, concern for present bias leads the government to help people increase their retirement savings. Second, the government uses all available tools to reward hard work so as to improve redistribution. There are two different ways to achieve this in our model: allowing individuals to keep some of their earnings and allowing individuals to indulge to their present bias. Third, as there is unobservable heterogeneity in present bias, the government wants to curtail flexibility on savings. By restricting the menu of savings options available to individuals, the government provides social insurance against the behavioral bias. This is true even if retirement savings are not at the first best level. Because of the interaction of those three forces, there is a push for paternalistic savings policies, whereby the government favors forced savings at low earnings, but

[3] In Appendix A.2 we extend our main results to a many period model with hyperbolic preferences.

at the same time a driving force for more flexibility so that high earnings individuals can indulge to their heterogeneous biases.

As usual in the public economics literature, there are different mechanisms the government can use to implement optimal policy, however we provide an implementation that uses three sets of policy tools. First, the government grants retirees a retirement benefit whose level depends on labor earnings during working life. In addition, young people cannot borrow against their retirement benefits. Second, there is both a regular savings account and multiple special retirement savings accounts. There is a cap on contributions to each special retirement savings account, and the cap depends on the labor earnings level of the person. Lastly, there are taxes on labor earnings and on savings returns. The labor income tax is non-linear in labor earnings and depends on which special retirement savings accounts the person contributes to. Taxes on the regular savings account are linear on savings but depend on earnings. Finally, taxes on special retirement savings accounts are lower than the tax on the regular savings account, and these taxes can also vary by the earnings level.

This implementation is particularly suitable for a comparison with current policy. The set of instruments we choose is very close to the set of instruments actually used by the U.S. government. For instance, retirement benefits directly translate into social security benefits and special retirement accounts resemble defined contribution plans such as a 401(k) account and individual retirement accounts (IRA). To the best of our knowledge, this is the first paper to highlight that a multitude of special retirement accounts can be used to implement optimal retirement savings policies when there is redistribution.

We find that an optimal retirement savings policy offers both social security old-age benefits and 401(k)-like retirement accounts to high earners, and only social security old-age benefits to low earnings individuals. Importantly, 401(k)-like retirement savings accounts are not available at low labor earnings. Indeed, low earnings

individuals rely heavily on retirement benefits. High earnings individuals that have less present bias use retirement savings accounts, but high earnings individuals that have severe present bias rely more on social security.

In order to compare current U.S. policy and the normative model prescriptions, we calibrate the distribution of discounting preferences and labor earnings ability in the data. We use current policies and data on wealth at retirement from the Health and Retirement Survey to calibrate for the joint distribution of discount factors and labor earnings. We find a considerable level of heterogeneity in discount factors and a small positive correlation with earnings.[4] These findings are consistent with estimations of heterogeneity in discount factors in the literature and also with a high heterogeneity in present bias found in the behavioral economics literature.[5]

Simulation of optimal policy in our model is complicated because of the government's two-dimensional screening problem. It is well known that results for multidimensional screening problems are hard to prove in closed form, it is also true that the presence of thousands of incentive constraints makes numerical solutions equally difficult ([58]). To this end, we develop a stable numerical solution algorithm to solve our model. Our algorithm searches for the smallest subset of incentive constraints that are relevant to the global solution of the problem.

Using calibrated parameters, one still needs to fix government preferences to solve for optimal policy. The government has a discount factor and a redistributive motive. The government discounts retirement consumption of individuals using its discount factor (paternalism) and assigns welfare weights to different individuals depending on their earnings ability (redistribution). We use two alternative procedures to choose the

[4]In Appendix A.5, we provide an alternative calibration with heterogeneity in present-bias in an incomplete markets life-cycle model with many periods and hyperbolic preferences. The level of heterogeneity we find in discounting of retirement savings is comparable to the one obtained in the benchmark calibration. Therefore, for the sake of not introducing a more complicated model only for the calibration section, in the main text we present only the calibration of the two-period model.

[5]See [3] and [4] for heterogeneity in discount factors. [76], [11] and [18] using experimental evidence find also considerable heterogeneity in present-bias.

government preferences. First, we develop a benchmark procedure in which we choose government preferences by approximating the normative model and the calibrated model allocation of consumption and earnings under current policy. This follows the approach in [50] and guarantees that the level of redistribution in the benchmark normative model we choose is similar to the one calibrated for the U.S. economy. In this benchmark, the government has a discount factor that is in the intermediary range of calibrated discount factors for individuals, so that there are both individuals that save too little and individuals that save too much from the government's perspective. Furthermore, the government redistributive motive is considerably less progressive than utilitarian. This is consistent with the findings in the literature.[6] However, this procedure does not guarantee a perfect approximation, accordingly we find that there is a gain available to the government of 18% in consumption equivalent terms at the benchmark normative model.

Moreover, we find in the benchmark normative model a striking difference between optimal retirement benefits and current social security benefits in the United States. We find that social security benefits are considerably lower than retirement benefits prescribed by the benchmark normative model, particularly so at high earnings. As the government is less progressive than utilitarian in the benchmark normative model, there is an important concern about the savings choices of higher earnings individuals. Therefore, the government finds it optimal to force a higher minimum level of savings at high earnings.

Motivated by the findings in our benchmark, in a second procedure we select government preferences by approximating the schedule of retirement benefits in the normative model to the current social security schedule in the United States. We find that this alternative procedure yields a planner with a redistributive motive that is more progressive than utilitarian and has a slightly smaller discount factor level

[6]See [50] and [51].

than in the benchmark. As a result of the considerable difference in the redistributive motive, in this procedure we find there are gains of 63% in consumption equivalent terms available to the government. These two exercises highlight that there is a sub-optimal mix of redistributive and savings policies in the current U.S. system according to our normative model. On one hand, in the first exercise we find the level of redistribution is largely inconsistent with the level of retirement benefits. On the other hand, in the second exercise we find that the level of retirement benefits is inconsistent with the current level of redistribution.

Our model can be generalized to explore government policies beyond savings. In section 1.5 we extend the approach in [77] and present a general model in which agents and the government disagree about the gains and costs of a generic action. The sources of disagreement can arise because of behavioral bias (e.g., present-bias), an externalities (e.g., pollution policy) or because of political economy considerations as in [2]. Importantly, there is heterogeneity in this disagreement, so that some agents agree with the government and others disagree. Furthermore, there is inequality in labor earnings ability so that redistribution is possibly a concern for the government. We demonstrate that as long as the government has a redistributive motive in addition to preferences for a particular action, optimal policy for the government will offer more flexibility in actions of high earnings individuals.

We explore two applications of this general model: drug policy and fuel efficiency policy. Of course, the drug policy debate is very complex and controversial. Our model focuses on the paternalistic nature of the policy and on its redistributive implications, with a caveat that it abstracts from other important considerations in this debate. Our findings imply that a paternalistic government will effectively make drug consumption prohibitive for low earnings individuals. However, if the possibility of consuming drugs pleases high earnings individuals enough to improve their willingness to work, then a government with a redistributive motive would set in place a

policy in which drug consumption is allowed at a high cost (e.g., using prescriptions). While redistribution demands more flexibility, if the government has no redistributive motive, we show that the government would be paternalistic toward all agents and would implement a prohibition on drug consumption.

In the second application we consider policies fostering the purchase of energy efficient vehicles. In the model, some people do internalize the cost of the extra pollution caused by an inefficient vehicles, but others do not internalize those costs and have preferences for less energy efficient vehicles. If there is no redistribution, optimal policy involves a strict regulation of the level of energy efficiency so that vehicles purchased by all agents would be energy efficient. However, again, if there is redistribution, the government is going to allow higher earnings individuals to purchase less efficient vehicles as long as that improves their willingness to work and contribute toward redistribution. In this case, policy can be implemented with income tax rebates on the purchase of vehicles depending on their energy efficiency level. At low earnings, there is a considerable tax rebate from the purchase of an efficient vehicle, but at high earnings, tax rebates are less generous. Therefore, there is progressivity in the schedule, a feature resembled by actual policy. Currently, there are tax rebates on federal income taxes for electric and plug-in hybrid cars in the United States. Therefore these rebates inherit the progressivity of the income tax schedule as the same rebate has different effects on total taxes paid by different taxpayers.

Moreover, our extension highlights that this paper contributes not only to the public finance literature on savings policies, but also to the intersection of the public finance and behavioral economics literature. Indeed, there are two strands of research in this intersection that closely relate to this paper. The first strand has considered how tax policy can be used to ameliorate behavioral biases. The second strand

has analyzed the provision of commitment devices for people with time inconsistent preferences.

There is a growing literature on optimal taxation with behavioral agents that focuses on optimal linear taxation without redistribution. Building upon the framework by [66], starting with [78], [45] and [79], this literature considers how linear taxes can be used to either prevent over consumption of some goods (e.g., fossil fuels, drugs) or to foster consumption of other goods (e.g., retirement savings). Recently, [30] find that if heterogeneity in behavioral biases is sufficiently high, quantity restrictions on consumption of the "behavioral" good for all agents fare better than linear taxes. Relative to these contributions, we consider the general optimal policy problem without a restriction to particular policy tools, and we show that optimal policy effectively restricts quantities at low earnings, but allows for a limited amount of flexibility on consumption at high earnings.

A few papers consider the taxation of behavioral agents with redistribution. [69] allow for non-linear labor earnings taxes and a linear tax on the good whose demand is affected by a behavioral bias. Conversely, we allow for a general tax structure both on labor earnings and on the good affected by a behavioral bias. Finally, [31] consider a model with a single level of present-bias.[7] We show that the introduction of heterogeneity in present bias allow us to characterize the level of flexibility that policies allow on retirement savings for agents at different earnings levels.

There is also a relatively small literature that has analyzed the optimal provision of commitment and flexibility in savings choices without redistribution. [5] develop a model in which there is heterogeneity in taste shocks toward future consumption but agents have a uniform level of present bias. They find that optimal policy is implemented by a minimum savings rule. In their model, there is both a value to offering flexibility and offering commitment even without redistribution. Recently,

[7] In their interpretation of the model disagreement is on the Pareto weight given to the future generation. However, this is isomorphic to present bias in a two-period model.

[37] extends this approach to a more general setting. In contrast to those contributions, in our model the value of flexibility is endogenous and it depends on how much flexibility can be used to improve redistribution. Indeed, the main insight on this paper arises from the endogenous variation of the value to flexibility across different earnings levels.

In concurrent work, [91] finds in a model with hyperbolic preferences that redistribution can be greatly enhanced by the planner. In fact, in his setup he finds that the first best can be implemented under certain conditions. The most important difference between his model and ours is on the timing that the social planner can obtain reports from agents. In his model, agents meet with the planner twice. Both before actually taking work and consumption decisions, and at the time they take those decisions. This is particularly important because agents have different preferences at those points in time and the same information set, thus the planner can extract informational rents from agents as emphasized in [23]. In contrast, in our two period model the planner only meets with agents when they actually have the behavioral bias, so that full surplus extraction is not possible. In Appendix A.3, we explore a multi-period version of our model with hyperbolic preference shocks and labor income shocks and show that the planner finds it optimal to offer more flexibility after high income shocks than after low income shocks.

Finally, we also broadly relate to two other strands of the literature. First, we relate to a large literature on the taxation of savings. [10] show that without behavioral biases and uncertainty on earnings ability, the government should not distort savings decisions for redistribution purposes. [88] and more recently [25] and [41] point out how a correlation between discount factors and labor earnings ability can break this result in models without a misalignment of individuals' and government's preferences. Second, we relate to a small literature on taxation with multi-dimensional characteristics. [8] emphasize that the lack of a general toolbox for the analysis of contracting

problems with multi-dimensional heterogeneity hindered their usage in applications.[8] When there is a misalignment in preferences, we are able to prove that allocations allow for an increasing level of flexibility when comparing low and high earnings levels, even without general tools to characterize the full problem's solution. Furthermore, we provide a numerical algorithm to fully solve the problem.

We organize this paper as follows. Section 1.2 develops a benchmark model for the joint analysis of paternalistic and redistributive policies. Then Section 1.3 shows that tax instruments similar to those already used in the U.S. can implement optimal retirement savings policies when there is a concern about behavioral biases in savings.. Section 1.4 quantitatively evaluates the benchmark model and shows that from the normative model perspective there is an inconsistency between current social security benefits and current redistribution policy in the United States. Section 1.5 shows that our main findings have broad implications for policies aiming at behavioral biases in other contexts as well as policies aiming at ameliorating externalities. Finally section 1.6 concludes the paper.

1.2 Optimal policy analysis

This section presents our benchmark model for the analysis of redistribution policies when there is a concern by the government about individuals undersaving for retirement. We develop a two-period model with unobservable heterogeneity in preferences toward retirement consumption and earnings ability. Following the approach in the optimal taxation literature, we first characterize properties of economic resource allocations that are efficient to the government. Then, we find our main insight about

[8] See [73], [7] and [81] for important contributions to the mechanism design literature with multi-dimensional types. However, there are important exceptions in public finance. [63] analyzes the taxation of couples and find that a higher earnings by a secondary earner generally reduce optimal taxes on the primary earner. [82] characterizes optimal income taxes when there is heterogeneity in earnings skills across different occupations.

government policy when there is a redistributive motive and the government disagrees with agents' preferences.

There is a unit mass of agents that live for two-periods and that differ in two unobservable attributes: earnings ability and preferences toward retirement consumption. We denote earnings ability by $\theta \in \Theta = \{\theta_1, \ldots, \theta_N\}$, where $\theta_N > \cdots > \theta_1 > 0$. The government discounts retirement consumption by δ, and agents discount retirement consumption by $\beta\delta$ where $\beta \in B = \{\beta_1, \ldots, \beta_M\}$ and $0 < \beta_1 < \cdots < \beta_M = 1$. We understand β as a placeholder for the disagreement between the government and the agent. We denote by (θ, β) an agent's type and denote its distribution by π which we assume has full support. The agents utility index at time of decision is

$$U(c_1, c_2, y; \theta, \beta) = u(c_1) - \frac{1}{\theta} v(y) + \beta\delta u(c_2)$$

where $u'(0) = +\infty$, $u''(\cdot) < 0$, $v(0) = 0$, $v'(0) = 0$, $v'(y) > 0$ for $y > 0$ and $v''(y) > 0$ for $y > 0$. Our choice of language for the disagreement between the government and the agents reflects the apparent impossibility to interpret β as present-bias in a two-period model. However, this is actually inconsequential as we can understand the government preference as the ex-ante preference of agents before they take decisions in the initial period.[9] Finally there is a storage technology with gross rate of return R for shifting resources into the second period.

We assume that the government does not observe agents' types, but it observes consumption levels and labor earnings, and it can choose the rules of the game to be played among agents. These rules can take an arbitrary form, but an important result in mechanism design theory, the Revelation Principle, guarantees that it is sufficient to consider a game in which the government asks agents their type, agents find it in their

[9] [5] argue in a similar two-period problem that the planner takes decisions before the initial period and that as a consequence the model indeed has three periods, even though agents only take decisions in two different periods. We show our results extend to a T-period model with behavioral biases in appendix A.2.

best interest to report their true type and the government assigns agents a resource allocation depending on their answer. We call this assignment rule and allocation and denote it by $\{c_1(\theta,\beta), c_2(\theta,\beta), y(\theta,\beta)\}_{(\theta,\beta)\in\Theta\times B}$. In the next section we show that the resulting government choice of rules is equivalent to a market economy with a sensible set of government policies in use. However, in this section our focus will be on this abstract resource allocation problem as it actually sheds light into the appropriate set of policy instruments that can implement optimal policy.

When choosing an allocation the government has to make sure it is consistent with the information and technological constraints of the economy. An allocation is said *incentive compatible* if

$$U(c_1(\theta,\beta), c_2(\theta,\beta), y(\theta,\beta); \theta, \beta) \geq U(c_1(\theta',\beta'), c_2(\theta',\beta'), y(\theta',\beta'); \theta, \beta) \quad (1.1)$$

for all $(\theta,\beta), (\theta',\beta') \in \Theta \times B$. If an allocation is incentive compatible, it is consistent with the unobservability of agents' types by the government. An allocation is consistent with the technology in the economy if it satisfies the resource constraint:

$$\sum_{(\theta,\beta)\in\Theta\times B} \pi(\theta,\beta)\left[y(\theta,\beta) - c_1(\theta,\beta) - \frac{c_2(\theta,\beta)}{R}\right] \geq 0 \quad (1.2)$$

so that earnings produced by workers can sustain their life-time consumption at the aggregate level.

The government evaluates individual payoffs according to

$$V(\theta,\beta) = u(c_1(\theta,\beta)) + \delta u(c_2(\theta,\beta)) - \frac{1}{\theta} v(y(\theta,\beta))$$

Therefore the government objective can be written as

$$W\left(\{c_1(\theta,\beta), c_2(\theta,\beta), y(\theta,\beta)\}_{(\theta,\beta)\in\Theta\times B}\right) = \sum_{(\theta,\beta)\in\Theta\times B} \pi(\theta,\beta)\lambda(\theta)V(\theta,\beta) \quad (1.3)$$

where $\lambda(\theta) \geq 0$ are Pareto weights the government assigns to agents. We assume weights depend only on earnings ability of the agent, not on its disagreement with the government. This is consistent with the government believing β is a bias that it ought to correct.[10] Without loss of generality, Pareto weights are normalized so that $\sum_{(\theta,\beta)\in\Theta\times B} \pi(\theta,\beta)\lambda(\theta) = 1$. The government problem is then to choose an allocation that maximizes (1.3) subject to both incentive constraints in (1.1) and the resource constraint in (1.2).

1.2.1 Example with 2×2 types

Before we state our theorems, it is useful to understand the government's problem in a simple example that highlights the key forces in the model. Assume there are two levels of disagreement and two levels of earnings ability. For simplicity, let $R\delta = 1$, $\beta \in \{\beta_L, 1\}$ with $\beta_L < 1$, $\theta \in \{\theta_L, \theta_H\}$ where $\theta_H > \theta_L = 0$, and assume the government is utilitarian, i.e., $\lambda(\theta_L) = \lambda(\theta_H) = 1$.

Bunching at low earnings – In the first best allocation without informational frictions and at a fixed ability type θ, the government bunches agents across their disagreement type β. As V is strictly concave, the government can always improve welfare by bunching the agents with a different disagreement level but the same ability level. In the second best with informational frictions, this same argument holds for low earnings ability agents. However, with unobservable characteristics it is necessary to satisfy incentive constraints. Since agents with low earnings ability cannot work as $\theta_L = 0$, the relevant incentive constraints are the ones preventing deviations of high

[10] Using the language of [66], we can think of two separate agents: self at period-0 and self at period-1. Self at period-1 has a present bias in saving for retirement ($\beta < 1$) and she takes all consumption and labor decisions. Conversely, self at period-0 has no present bias, but she cannot take period 1 decisions by herself. In this behavioral interpretation, the government preference is in agreement with preferences of self at period-0.

ability types into the allocations of low ability types:

$$u\left(c_1\left(\theta_H,\beta\right)\right)-\frac{1}{\theta_H}v\left(y_1\left(\theta_H,\beta\right)\right)+\beta\delta u\left(c_2\left(\theta_H,\beta\right)\right)\geq u\left(c_1\left(\theta_L,\beta'\right)\right)+\beta\delta u\left(c_2\left(\theta_L,\beta'\right)\right) \tag{1.4}$$

for $\beta,\beta'\in\{\beta_L,1\}$. Since incentive constraint (1.4) is linear in utility levels $u\left(c_1\left(\theta_L,\beta'\right)\right)$ and $u\left(c_2\left(\theta_L,\beta'\right)\right)$, then a convex combination of those utility levels satisfy incentive compatibility

$$u\left(c_1\left(\theta_H,\beta\right)\right)-\frac{1}{\theta_H}v\left(y_1\left(\theta_H,\beta\right)\right)+\beta\delta u\left(c_2\left(\theta_H,\beta\right)\right)\geq\bar{u}_1\left(\theta_L\right)+\beta\delta\bar{u}_2\left(\theta_L\right) \tag{1.5}$$

where $\bar{u}_t\left(\theta_L\right)=\sum_{\beta'}\pi\left(\beta'|\theta_L\right)u\left(c_1\left(\theta_L,\beta'\right)\right)$. Therefore, it is incentive compatible for the government to allocate $\bar{c}_t\left(\theta_L\right)=u^{-1}\left(\bar{u}_t\left(\theta_L\right)\right)$ for low ability agents. Further, the government objective does not change. However, since the utility index u is strictly concave, it follows that

$$\bar{c}_1\left(\theta_L\right)+\frac{1}{R}\bar{c}_2\left(\theta_L\right)<\sum_{\beta'}\pi\left(\beta'|\theta_L\right)\left[c_1\left(\theta_L,\beta'\right)+\frac{1}{R}c_2\left(\theta_L,\beta'\right)\right]$$

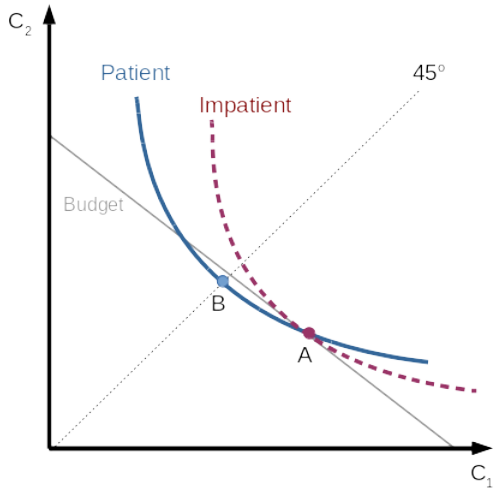
if the allocations of low ability agents are not all equal. We conclude the government bunches low ability agents. If that was not the case, the government would perturb the allocation as we propose and it would use the extra resources available to strictly increase its objective.

Separation at high earnings – Start with an allocation that bunches high ability agents, so that they receive the same consumption levels in both periods and produce the same labor earnings during working life. We show it is possible to improve the government's objective by changing such an allocation. There are two different cases to consider.

In the first case, there is bunching of high ability agents into an allocation with a higher consumption level in the initial period, that is, $c_1\left(\theta_H\right)>c_2\left(\theta_H\right)$. In Figure

1.1 we illustrate indifference curves on consumption allocations of individuals with the same level of labor earnings. At point A, we have $c_1(\theta_H) > c_2(\theta_H)$. At this point, the indifference curve of an impatient agent is steeper as a higher change in period two consumption is required by this agent to compensate for the same change in period one consumption.[11] While continuing to offer allocation A, the government can target patient agents by offering allocation B. However, since points on the 45^0 line minimize the cost of providing the same level of utility to patient agents ($R\delta = 1$), allocation B has a strictly lower cost in terms of resources. This is illustrated in the figure by allocation B being inside the budget of resources used to obtain allocation A. Therefore the government can obtain the same objective value while spending less, a contradiction.

Figure 1.1: Incentives with bunching.



In the second case, there is bunching of high ability agents into an allocation in which period two consumption is weakly higher that period one consumption, i.e., $c_2(\theta_H) \geq c_1(\theta_H)$. This is illustrated in Figure 1.2 by point D. In this case,

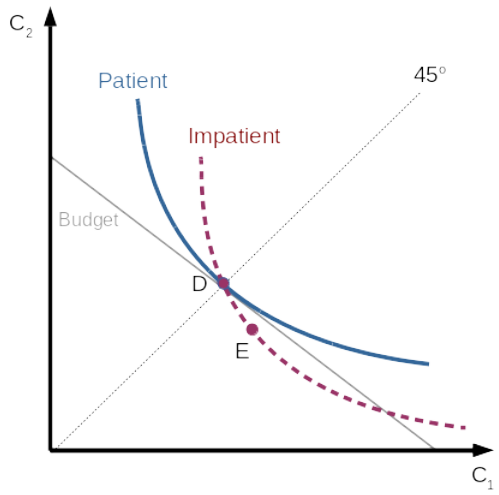
[11] The marginal rate of substitution between consumption in periods one and two is given by

$$\frac{\beta\delta u'(c_2)}{u'(c_1)}$$

which is monotone in β. Therefore there is single crossing in preferences conditional on a fixed level of labor earnings.

while keeping the availability of allocation D for patient agents, the government can target impatient agents by offering allocation E. While impatient agents are indifferent between allocations D and E, allocation E has a lower resource cost and it reduces the payoff of impatient agents from the government perspective (the government indifference curve coincides with that of that of the patient agent). Therefore, in this case, we need to show that the government gains in transferring those extra resources to low ability agents are sufficiently high so as to compensate for the loss with impatient high ability agents. But since ability is not observable and $y(\theta_H) > 0 = y(\theta_L)$, then from incentive compatibility it must be the case that either $c_1(\theta_H) > c_1(\theta_L)$ or $c_2(\theta_H) > c_2(\theta_L)$ or both. Therefore, for a utilitarian planner, the marginal gain in transferring resources to low ability agents will surpass the marginal loss of high ability impatient agents as long as E is sufficiently close to D. Finally, we conclude that the government finds it optimal to separate high ability agents.

Figure 1.2: Separation at high ability



Distortions at low ability – We already know that low ability agents are bunched. Now, assume that point A in Figure 1.1 represents low ability agents consumption allocation while the indifference curves represented are for high ability agents deviating into that allocation. Then it is possible to keep the government's goal fixed and save resources by offering point B instead to low ability agents. Therefore, we

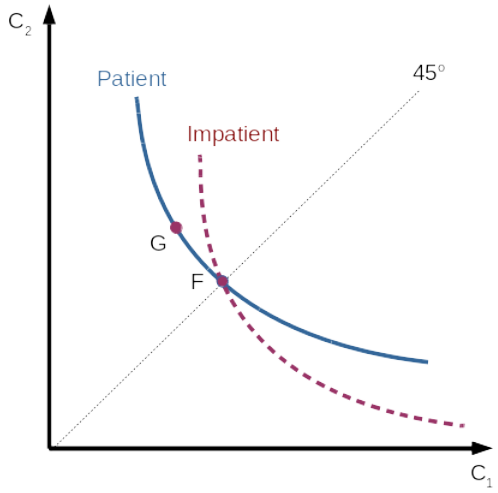
conclude $c_2(\theta_L) \geq c_1(\theta_L)$. Again, in Figure 1.3, we illustrate the consumption allocation of low ability agents shile the indifference curves of high ability agents. We see that if there is perfect consumption smoothing, represented by point F, then it is always possible to relax the incentive constraint of impatient high ability agents by perturbing the allocation in the direction of allocation G. Patient agents are made indifferent by the perturbation, but it strictly relaxes the incentive constraint of high ability impatient agents. It follows that at the solution to the government's problem we have $c_2(\theta_L) > c_1(\theta_L)$ as we will see that the incentive constraint of high ability impatient agents is binding.

Distortion at high ability and patient – If the government offers an allocation with $c_2(\theta_H, \beta_H) < c_1(\theta_H, \beta_H)$ for high ability patient agents, represented by point A in Figure 1.1, it is always possible to offer instead point B and save resources while keeping the government's objective constant. Thus $c_2(\theta_H, \beta_H) \geq c_1(\theta_H, \beta_H)$. Furthermore, the government would offer $c_2(\theta_H, \beta_H) > c_1(\theta_H, \beta_H)$ if and only if the incentive constraint of high ability impatient agents is binding with respect to the allocation of high ability patient agents. Next we show this is never the case in this example and therefore we conclude $c_2(\theta_H, \beta_H) = c_1(\theta_H, \beta_H)$. Therefore, the government finds it optimal not to distort the consumption allocation of patient agents.

Distortion at high ability and impatient – It immediately follow from separation and the distortion for high ability and patient agents that $c_1(\theta_H, \beta_L) > c_2(\theta_H, \beta_L)$. Therefore the government allows impatient high ability agents to consume proportionally less at retirement than what perfect consumption smoothing would predict.

Now we illustrate the pattern in which incentive constraints of high ability agents bind. At least one incentive constraint from high ability agents to low ability agents must bind as consumption is not equalized across agents with different ability.

Figure 1.3: Over consumption at retirement of low ability



High ability and impatient incentives – First, let's consider high ability impatient agents. Assume by way of contradiction that the incentive constraint of type (θ_H, β_L) is strictly slack with respect to the low ability agents' allocation. In this case we have $c_1(\theta_L) = c_2(\theta_L)$ as there is no gain from not providing perfect consumption smoothing to low ability agents. We have already shown that $c_2(\theta_H, \beta_H) \geq c_1(\theta_H, \beta_H)$, therefore $c_2(\theta_H, \beta_H) \geq c_2(\theta_L)$. Then a binding incentive constraint of type (θ_H, β_H) implies

$$u(c_1(\theta_H, \beta_H)) - \frac{1}{\theta_H} v(y_1(\theta_H, \beta_H)) + \beta_L \delta u(c_2(\theta_H, \beta_H)) \leq u(c_1(\theta_L)) + \beta_L \delta u(c_2(\theta_L))$$

and all incentive constraint of type (θ_H, β_L) are strictly slack, a contradiction. Therefore it must be the case that the incentive constraint of type (θ_H, β_L) is binding with respect to the allocation of low ability agents.

Second, the high ability impatient agent incentive constraint is slack with respect to the allocation of the high ability patient agent. Assume by way of contradiction that is not the case. Since the incentive constraint of high ability impatient agents deviating into low ability is binding, then we have

$$u(c_1(\theta_H, \beta_H)) - \frac{1}{\theta_H} v(y_1(\theta_H, \beta_H)) + \beta_L \delta u(c_2(\theta_H, \beta_H)) = u(c_1(\theta_L)) + \beta_L \delta u(c_2(\theta_L))$$

At the solution to the planner's problem we must have $c_2(\theta_H, \beta_H) > c_2(\theta_L)$.[12] Therefore we conclude

$$u(c_1(\theta_H, \beta_H)) - \frac{1}{\theta_H} v(y_1(\theta_H, \beta_H)) + \delta u(c_2(\theta_H, \beta_H)) > u(c_1(\theta_L)) + \delta u(c_2(\theta_L))$$

But since high ability agents are separated, we also must have $c_2(\theta_H, \beta_H) > c_2(\theta_H, \beta_L)$ and therefore all incentive constraint of high ability patient agents are strictly slack. But then, the government could improve redistribution in the economy, a contradiction. We conclude that the high ability impatient agent has a slack incentive constraint with respect to the allocation of the high ability patient agent.

High ability and patient incentives – Finally, for patient high ability agents the pattern in which incentives constraints bind is not straightforward. In particular, it will depend on several parameters such as the level of β_L and the redistributive motive. On one hand if $\beta_L \approx 1$, we have at the solution to the planner's problem $c_2(\theta_H, \beta_L) > c_2(\theta_L)$ so that

$$u(c_1(\theta_H, \beta_L)) - \frac{1}{\theta_H} v(y_1(\theta_H, \beta_L)) + \delta u(c_2(\theta_H, \beta_L)) > u(c_1(\theta_L)) + \delta u(c_2(\theta_L))$$

where we used our last result that the incentive constraint of (θ_H, β_L) is binding with respect to θ_L. In this case the patient high ability agent has a slack incentive constraint with respect to low ability agents. On the other hand, if $\beta_L \approx 0$ and $\lambda(\theta_H) \approx 0$, then $c_2(\theta_H, \beta_L) \approx 0 < c_2(\theta_L)$, and the incentive constraint of type (θ_H, β_H) with respect to low ability types will be binding. With many types, the characterization of which incentive constraints bind becomes an intractable problem.

[12] Assume by way of contradiction that $c_2(\theta_L) \geq c_2(\theta_H, \beta_H)$. Since there is separation of high ability agents, then $c_2(\theta_L) \geq c_2(\theta_H, \beta_H) > c_2(\theta_H, \beta_L)$. Therefore, the utilitarian government would like to redistribute period two consumption from low ability agents to high ability agents. This is always incentive compatible as low ability agents cannot work. We then obtain a contradiction and conclude that $c_2(\theta_H, \beta_H) > c_2(\theta_L)$.

However, it is still possible to provide key implications for optimal policy even without this characterization.

1.2.2 Main results

Now we turn back into our benchmark model with heterogeneity in disagreement and earnings ability. In the problem with two-dimensional heterogeneity and many types, the major difficulty is the characterization of the pattern in which incentive constraints bind. In a problem with one dimensional heterogeneity, such as in [75], under fairly general assumptions only incentive constraints of types close to each other can bind. That is not true in problems with multi-dimensional heterogeneity as pointed out by [7] and [81]. However, we are still able to characterize key properties of the solution to the government's problem at the bottom and top levels of earnings ability.[13] In Theorem 1 we obtain the main result in this paper and then in Theorem 2 we characterize distortions at top and bottom earnings at the solution to the planner's problem.

Theorem 1. *Assume $\lambda(\theta)$ is weakly decreasing in θ. Fix θ_N and $\{\beta_2, \ldots, \beta_M\}$ then there exists $\underline{\theta}_1 > 0$, $\bar{\theta}_{N-1} < \theta_N$ such that at the solution to the government's problem:*

1. *if $\underline{\theta}_1 > \theta_1 > 0$, then agents with types in $\{(\theta_1, \beta) : \beta \in B\}$ receive the same allocation, independently of their type β:*

$$c_t(\theta_1, \beta) = \bar{c}_t(\theta_1)$$
$$y(\theta_1, \beta) = \bar{y}(\theta_1)$$

for $t = 1, 2$ and all $\beta \in B$;

[13] In section 1.4 we find through numerical simulations that those properties are not particular to the end-points of the distribution of types, but that they are in fact a force present at all earnings levels.

2. *if* $\theta_1 < \cdots < \theta_{N-1} \leq \bar{\theta}_{N-1}$*, then agents with types in* $\{(\theta_N, \beta) : \beta \in B\}$ *do not receive the same allocation.*

Proof. See Appendix. □

On one hand, the government finds it optimal to offer different allocations for high earnings ability agents with type θ_N. Since agents preferences disagree with one another, they would like to receive different allocations. Therefore, the government can actually extract some extra resources from them by offering different allocations that agents find more attractive. Since Pareto weights are decreasing, the government would like to redistribute toward lower earnings ability individuals, hence extracting those extra resources has a positive marginal value. As a result the government will be willing to allow for different allocations at high earnings as to obtain extra resources from those agents.

On the other hand, the government finds it optimal to bunch the lowest earnings agents along the β dimension. It is not in the interest of the government to extract resources at low earnings ability. Therefore, the planner will find it optimal to bunch agents, effectively providing them full insurance against their preference heterogeneity β-types. In an interpretation of β as reflecting behavioral biases, that we explore fully in appendix A.2, the government here is insuring agents against the possibility they face time inconsistencies into the future.

Even though the government problem is a two-dimensional screening problem, we are able to obtain important insights about the optimal allocation. The assumption of a paternalistic government reduces the dimensionality of the objective function and allows for a partial characterization of optimal policy. It is clear that the government wants to bunch agents across the β dimension, therefore it is easier to understand how incentive constraints bind at the lowest and highest earnings levels.

Now we define wedges on decisions that represent distortions at the solution to the government's problem. The intertemporal wedge

$$\tau^{I}(\theta,\beta)=1-\frac{u'(c_1(\theta,\beta))}{R\beta\delta u'(c_2(\theta,\beta))}$$

allow us to understand distortions on intertemporal consumption decisions. A positive intertemporal wedge is alike a tax on future consumption so agents facing it are consuming relatively little in the future. We can also define the government's intertemporal wedge

$$\tau^{PI}(\theta,\beta)=1-\frac{u'(c_1(\theta,\beta))}{R\delta u'(c_2(\theta,\beta))}$$

which uses the government's preferences (i.e., $\beta=1$). The government's intertemporal wedge measures distortions on agents intertemporal decisions from the government's perspective. For example, if an agent faces no intertemporal wedge but has disagreement with the government $\beta<1$, then the government's intertemporal wedge is positive for that agent. In the next result we characterize the sign of intertemporal wedges at the solution to the government's problem.

Theorem 2. *Assume $\lambda(\theta)$ is weakly decreasing. Fix $\{\theta_2,\ldots,\theta_{N-1}\}$ and $\{\beta_2,\ldots,\beta_M\}$ then there exists $\underline{\theta}_1>0$ and $\bar{\theta}_{N-1}<\theta_N$ such that if $\theta_1\leq\underline{\theta}_1$, $\theta_1<\cdots<\theta_{N-1}\leq\bar{\theta}_{N-1}$ at the solution to the government's problem*

1. *for types with lowest earnings ability:*

   (a) *intertemporal wedge is negative: $\tau^{I}(\theta_1,\beta)<0$ for all $\beta<1$*

   (b) *planner's intertemporal wedge is weakly negative: $\tau^{PI}(\theta_1,\beta)\leq 0$ for all β*

2. *for types with higher earnings ability:*

   (a) *agents that agree with the government ($\beta_M=1$) face weakly negative intertemporal wedges $\tau^{I}(\theta_N,\beta_M)=\tau^{PI}(\theta_N,\beta_M)\leq 0$*

(b) agent that disagrees the most with the government ($\beta_1 < 1$) face a positive government's intertemporal wedge $\tau^{PI}(\theta_N, \beta_1) > 0$

Proof. See the appendix. □

In Theorem 2 we find that the intertemporal wedge of the lowest earnings ability individuals is negative for agents that disagree with the government. Further, from the planner's point of view it shows those agents weakly over-consume at retirement. In our numerical simulations we find a strictly positive government intertemporal wedge. By forcing low earnings ability individuals to consume relatively more at retirement than would be optimal in the first best, the government makes their allocation particularly undesirable for high earnings ability individuals that disagree with government preferences ($\beta < 1$). Therefore, the government relaxes their incentive constraints by doing so. But if the optimal allocation was to force all low earnings individuals into consuming at the first best marginal rate of substitution (a zero government's wedge), then the marginal cost to the planner of forcing them to consume relatively more at retirement is of second order. However, the gain from relaxing incentive constraints is of first order because of redistribution.

Finally Theorem 2 shows a dispersion in distortions at high earnings levels. For agents that agree with the government, there is no gain in under-consuming at retirement (in numerical simulations we find it is not optimal to over-consume at retirement as well). Distorting savings in this case generates both a lower government's objective and also reduces incentives for those agents to work. However, for individuals that disagree with the government ($\beta_1 < 1$) and also have high earnings, it is worth to the planner to allow them to consume in the way they like in exchange for extracting some more resources from those individuals in order to improve redistribution. As a result those agents consume too little at retirement relative to their working life consumption.

1.2.3 Comparison to most related results in the literature

In order to better understand the forces at work in our model, it is useful to consider two important particular cases. The first one is the classical result obtained by [10]. It shows that without disagreement, the government does not find it optimal to distort intertemporal consumption decisions for any redistributive motive. The second result is from [31] and it considers the case with disagreement but no heterogeneity in the level of disagreement across agents. They show that there is a distortion in intertemporal consumption decisions and that this distortion is monotonic on earnings.

It is easy to understand the intuition behind the result in [10] using our 2×2 types example. If $\beta = 1$ for all agents, there are only two types of agents with $\theta \in \{0, \theta_H\}$. The only incentive constraint in this problem is given by

$$u\left(c_1\left(\theta_H\right)\right) + \delta u\left(c_2\left(\theta_H\right)\right) - \frac{1}{\theta_H} v\left(y\left(\theta_H\right)\right) \geq u\left(c_1\left(0\right)\right) + \delta u\left(c_2\left(0\right)\right)$$

Notice that both types of agents and the government agree on how to value consumption over the life-time. Then at the optimum, it must be the planner is minimizing resource expenditures with consumption of each agent:

$$\begin{aligned} &\min_{c_1, c_2} \left\{ c_1 + \frac{1}{R} c_2 \right\} \\ &s.t. u\left(c_1\right) + \delta u\left(c_2\right) = \bar{U}\left(\theta\right) \end{aligned}$$

where $\bar{U}\left(\theta\right) = u\left(c_1\left(\theta\right)\right) + \delta u\left(c_2\left(\theta\right)\right)$. As only $\bar{U}\left(\theta\right)$ matters for the incentive constraint, not the consumption levels separately. But if we take the first order conditions of the problem above we obtain

$$u'\left(c_1\left(\theta\right)\right) = R\delta u'\left(c_1\left(\theta\right)\right)$$

and therefore the government should not distort agent's intertemporal decisions.

In order to understand the intuition behind the result in [31], consider the example when $\beta_L = \beta_H = \beta < 1$ and $\theta \in \{0, \theta_H\}$. The incentive constraint is given by

$$u\left(c_1\left(\theta_H\right)\right) + \beta\delta u\left(c_2\left(\theta_H\right)\right) - \frac{1}{\theta_H} v\left(y\left(\theta_H\right)\right) \geq u\left(c_1\left(0\right)\right) + \beta\delta u\left(c_2\left(0\right)\right)$$

Assume by way of contradiction that there are distortions to agents intertemporal decisions at the solution to the government's problem. Since $\lambda\left(0\right) = \lambda\left(\theta_H\right)$ and $y\left(\theta_H\right) > 0$, then either $c_1\left(\theta_H\right) > c_1\left(0\right)$ or $c_2\left(\theta_H\right) > c_2\left(0\right)$. Therefore the government wants to transfer more resources to agents with type $\theta = 0$, as a result the incentive constraint is binding and relaxing it would strictly improve the government's objective. However, under our assumption of no distortion, there are two different ways to relax the incentive constraint. On one hand, the government can increase $c_2\left(0\right)$ in exchange for a lower $c_1\left(0\right)$ while keeping the amount of resources used by agents with type $\theta = 0$ constant. This implies a second order welfare loss to the planner, but strictly relaxes the incentive constraint of type θ_H as $\beta < 1$. On the other hand, the government can increase $c_1\left(\theta_H\right)$ in exchange for a reduction of $c_2\left(\theta_H\right)$ while keeping resource usage constant. This perturbation will lead to a second order loss in the government's objective, but will strictly relax the incentive constraint. Finally, the government will optimally use both types of distortions and as a result agents with type θ_H under-consume in period two and agents with type $\theta = 0$ over-consume in period two.

The forces in [10] and in [31] are present in our model. As we can inspect in Theorem 2, the intertemporal distortions display a monotonic shape as in [31]. Also high earnings ability individuals that agree with the government ($\beta = 1$) are not distorted.

However, our results rely on a key third force into our model: heterogeneity in disagreement. With heterogeneity in disagreement, now the government also needs to worry about dispersion in allocations of agents with the same level of earnings ability. The interaction of this third force with redistribution then implies that the planner will be particularly worried about dispersion of allocations across low earnings ability agents, but not as much by the dispersion on allocations across high earnings ability agents. Thus, we find that the government will not only distort agents with disagreement as in [31] while not distorting some agents without disagreement as in [10], but it will also bunch them across disagreement types at low earnings ability.

1.3 Decentralization of optimal retirement savings policies

In this section we provide a decentralization of the solution to the government's problem as a competitive equilibrium in a market economy in which the government uses four sets of policy tools. All of those instruments are very similar to policy tools currently used in the U.S., so we are able to connect our results in the previous section with actual policy.

The first is an old-age benefit which is non-linear on labor earnings and which agents cannot borrow against during working life. Our previous results indicate that agents with a large level of disagreement (present bias) would like to borrow against those resources at period one, however as that would reduce the government objective it enforces by law that retirement benefits cannot be used as a guarantee on loans.[14]

The second set of policy tools includes multiple retirement savings accounts. Those accounts have caps on contributions and they have tax advantages as compared to a regular savings account. Both the caps and taxes on those account depend on

[14]In the U.S., the law prohibits the usage of social security benefits as collateral on loans.

the earnings level of individuals in the decentralization. In the U.S. 401(k) accounts are characterized by a differential income tax treatment, the possibility of employer matching contributions and a cap on contributions.

The third set of policy tools used by the government includes linear subsidies or taxes on both savings through retirement savings accounts and savings at the free market rate on bank accounts. Finally, the fourth policy tool used by the government is a non-linear labor earnings tax that depends on the set of retirement savings account used by the agent. This last property is present as well in the current U.S. tax code as contributions to 401(k) accounts or IRA accounts have a differential tax treatment.

Agents can save in a regular savings account with a gross rate of return $(1-\tau_{M-1}(y))R$ and in $m=1,\ldots,M-2$ retirement savings accounts with a rate of return $(1-\tau_m(y))R$ and contribution caps $\bar{a}_m(y)$. Both rates of return and caps depend on the earnings level y of the agent. We have $0 \geq \tau_{m+1}(y) \geq \tau_m(y)$ so that agents that use account $m+1$ also use accounts $1,\ldots,m$. Agents owe income taxes $T_m(y)$ to the government while young. We denote by $T_{M-1}(y)$ labor income taxes on the agent that saves on all retirement savings accounts and on the regular savings account, and by $T_0(y)$ the income tax of agents that do not save on top of their social security contribution. At retirement, agents receive a retirement benefit $b(y)$

that varies with the labor earnings level during working life. The agent's problem is

$$\begin{aligned}
&\max_{c_1,c_2,y,M_0} u(c_1) - \theta v(y) + \beta\delta u(c_2) \\
&\quad s.t. c_1 + a_s + \sum_{m=1}^{M_0} a_{r,m} = y - T_{M_0}(y) \\
&\quad c_2 = b(y) + Ra_s + R\sum_{m=1}^{M_0} (1 - \tau_m(y)) a_{r,m} \\
&\quad a_s \geq 0 \\
&\quad 0 \leq a_{r,m} \leq \bar{a}_m(y) \\
&\quad M_0 \in \{1, \ldots, M-2\}
\end{aligned}$$

We say a government policy $\left(\{\bar{a}_m(\cdot)\}_{m=1}^{M-2}, \{\tau_m(\cdot)\}_{m=1}^{M-1}, \{T_m(\cdot)\}_{m=0}^{M-1}, b(\cdot)\right)$ is feasible if agents' equilibrium choices given the pension scheme satisfy the economy-wide market clearing condition

$$\sum_{(\theta,\beta)\in\Theta\times B} \pi(\theta,\beta)\left[y(\theta,\beta) - c_1(\theta,\beta) - \frac{c_2(\theta,\beta)}{R}\right] = 0$$

It is straightforward that the solution to the government's problem can be decentralized by a government policy as just described. Assume $\lambda(\theta)$ is weakly decreasing in θ and $\lambda(\theta_N) > 0$. Fix θ_N and $\{\beta_2, \ldots, \beta_M\}$, then there exists $\bar{\theta}_1 > 0$, $\bar{\theta}_{N-1} < \theta_N$ and $\bar{\beta}_1 > 0$ such that if $\theta_1 < \bar{\theta}_1$, $\theta_1 < \cdots < \theta_{N-1} \leq \bar{\theta}_{N-1}$ and $\beta_1 < \bar{\beta}_1$, then the solution to the government's problem can be decentralized with a government policy as above and in which:

1. agents with types in $\{(\theta_1, \beta) : \beta \in B\}$ do not have access to retirement savings accounts and only receive social security benefits when old;

2. agents with types in $\{(\theta_N, \beta) : \beta \in B\}$ have access to retirement savings accounts and some save on top of social security contributions.

In Corollary 1.3 we show that a pension system with social security benefits and multiple retirement savings accounts in which low earnings individuals do not have access to the retirement savings accounts can be actually optimal for a paternalistic planner. At first, the nature of the pension system in the U.S. where high earners get disproportional incentives on savings for retirement through 401(k) retirement accounts might not sound efficient in an environment where redistribution is important. However, we find the somewhat surprising result that this system shares some characteristics of an efficient paternalistic policy.

1.4 Quantitative exercise

In this section we evaluate current redistributive and retirement savings policies in the U.S. through the lenses of our normative model. We first develop a two-period model with actual policy tools to calibrate for preference heterogeneity present in the data. Then we find planner preferences for redistribution and δ, the government's discount factor, such that we obtain a normative model that is comparable to actual policy in terms of both the levels of redistribution and savings distortions observed. By taking this parsimonious approach, we minimize the chances actual policy is inefficient from a normative perspective. Finally we compare the results of our normative model to actual policy.

1.4.1 Two-period calibration of current policy

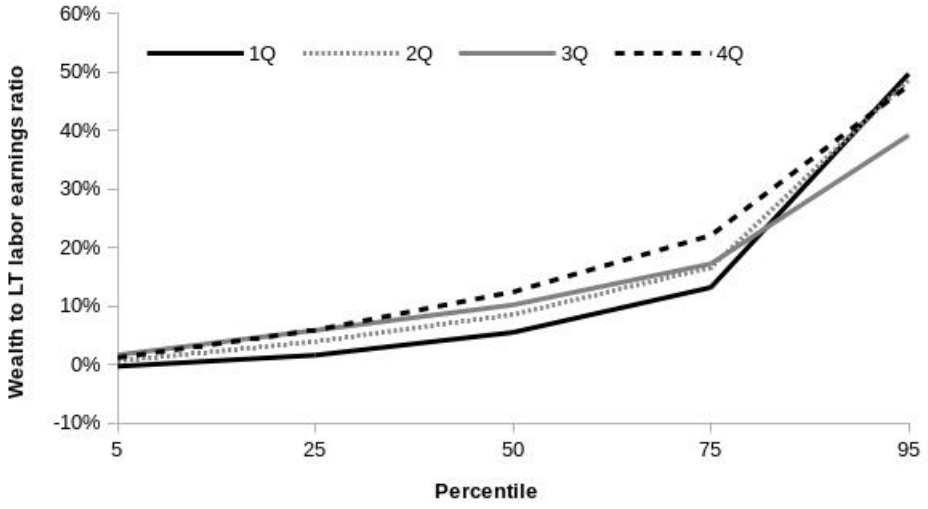
First we calibrate for heterogeneity in discount factors using a two-period model of retirement savings decisions under current government policies. We use data from the University of Michigan Health and Retirement Study (HRS) as provided in [28] on the ratio of wealth at retirement to life-time earnings of households.[15] This statistic

[15]The wealth measure used in [28] is at the household level and includes net liquid financial assets, half of the principal house equity, other real state equity, business equity, deposits in all

is informative about how people save controlling for their life-time labor earnings, the essential heterogeneity we would like to uncover. Therefore it is natural to calibrate the model to the shape of the cross sectional distribution of this statistic in the data.

There is considerable heterogeneity in retirement savings rates in the data and a small positive correlation with earnings. In Figure 1.4 we display the data on ratios of wealth at retirement to lifetime earnings that we use to match the model. Each line in the figure displays the percentiles of the distribution of the ratio of wealth at retirement to lifetime earnings within a lifetime earnings quantile. We see that at the different quantiles on lifetime earnings, the statistic under study has a very similar distribution although it is slightly larger at higher life-time earnings quantiles. This points to a small correlation between savings rates and labor earnings in the data.

Figure 1.4: Distribution of non-social security wealth at retirement ratio to life-time earnings.



Source: Each line represents the distribution within a different lifetime earnings quantile. Data from [28] using 1992 Health and Retirement Survey sample of households. Non social security household wealth includes all liquid wealth, deposits on retirement accounts, estimated defined benefit plans, business equity, other real estate equity and half of primary home value.

types of retirement accounts and estimates of defined contribution pensions benefits (excluding social security). Using the history of current and past reported earnings and estimates from Khitatrakun, Kitamura, and Scholz (2000), [28] obtain a measure of lifetime household earnings.

In our calibration, we parameterize households preferences by

$$U\left(c_1, c_2, y; \theta, \beta\right) = \frac{c_1^{1-1/\sigma} - 1}{1 - 1/\sigma} - \frac{\left(\frac{y}{\theta}\right)^{1+1/\gamma}}{1 + 1/\gamma} + \psi \frac{c_2^{1-1/\sigma}}{1 - 1/\sigma}$$

where we use a standard level of $\sigma = 0.5$ for the intertemporal elasticity of substitution and a value on $\gamma = 1$ for the Frisch elasticity of labor supply. Clearly, we cannot disentangle β from δ in the two-period model calibration.[16] Hence, we calibrate here for the joint distribution of $\psi = \beta\delta$ which is the discount factor of individuals at period one, and the distribution of θ which is labor earnings ability. In the next sub-section, we will use current policy to inform us about appropriate values for δ, the government's discount factor.[17] Further we assume that the joint distribution can be written as a Gaussian copula between the marginal distributions of ψ and θ with a correlation parameter ω that we calibrate.

For the marginal distribution of earnings ability θ, we follow [50] and [51] in using the actual observed labor income distribution. [50] argue for this approach as labor income and earnings are proportional to each other in logs after controlling for consumption in their model. In our model that is not entirely true, because of the complexity introduced by old-age benefits and retirement accounts that depend on earnings levels. However, we keep that assumption as it is a parsimonious first approximation. Finally, we use the percentiles of the labor income distribution across the age profile from the 2013 March CPS as the distribution of θ.[18]

[16] Separate identification of β and δ requires data on at least three periods in a model with hyperbolic discounting.

[17] In Appendix A.5 we provide an alternative calibration using a standard incomplete markets life cycle model in which agents have hyperbolic preferences. We find that the heterogeneity in discount factors is similar. As a result, the comparison of optimal policy and current policy has the same qualitative implications using both approaches to calibrate discount factors.

[18] We compute the percentiles within each age level from 25 to 60 years old for household heads. Then for each percentile we compute the average labor income across the life-cycle.

For the marginal distribution of ψ we assume a Beta(a, b) distribution with shape parameters $a > 0$ and $b > 0$.[19] The Beta distribution is an appropriate choice as it does not impose symmetry, it allows for fat tails on both directions (as well as a bell shape) and it is bounded in $[0, 1]$. In order to make ψ comparable to estimates of annualized discount factors in the literature, we report it in terms of an equivalent annualized value.[20]

In modeling the tax-transfer system, we follow [16], [17], [50]and [51] by using a parsimonious functional form calibrated to the level and progressivity of the net tax schedule that was introduced by [34]. In this formulation, we write net transfers T to individuals as function of taxable income Y :

$$T(Y; \lambda, \tau) = Y - \lambda Y^{1-\tau}$$

where taxable income include both labor earnings and taxable asset income. Estimating the functional parameters off PSID data for 2002–2006, [50] find that $\tau = 0.151$ and $\lambda = 0.836$ provide a good fit of the current U.S. tax and transfer system among households whose head worked more than 260 hours in the previous year. We adopt their estimates in our calibration.

In addition to the tax-transfer system applicable to individuals' working lives, we model the current retirement savings system with a sequence of accounts that are subject to subsidies and income-specific caps: first, we model the social security old-age benefit using the 2014 replacement ratios and the 2014 \$118,500 cap on eligible labor earnings. Second, we model a 401(k) account as allowing voluntary tax-deferred contributions up to \$18,000 plus 50% matching on these contributions

[19] In the numerical computations this distribution is discretized in 10 grid points with uniform probability.

[20] We assume that working life spans over 40 periods and that retirement lasts for 20 periods, then

$$\psi_{two-period} = \psi_{annual}^{40} \left(\frac{1 - \psi_{annual}^{21}}{1 - \psi_{annual}^{41}} \right)$$

through employers.[21] Third, we model an IRA account as allowing for voluntary tax-deferred contributions up to $5,500. Finally, we include in our model a regular savings account with a real rate of return of 3.44% following [43].

The model is able to match the data considerably well when we calibrate for (a, b, ω). The targets for the calibration are the 25, 50 and 75 percentiles of the distribution of wealth at retirement to lifetime earnings ratios at all four quantiles of the lifetime earnings distribution (12 targets in total). We report in Table 1.1 the fit of the model to the data. As we can see in the table the model is not perfectly able to match the data, but it gets the overall shape very well.

Table 1.1: Calibration fit by life time earnings quantile.

Statistic\Quantile	Data Q1	Model Q1	Data Q2	Model Q2
W/Y percentile 25%	0.0165	0.0163	0.0398	0.0367
W/Y percentile 50%	0.0554	0.0899	0.086	0.1021
W/Y percentile 75%	0.1322	0.1431	0.1664	0.1712

Statistic\Quantile	Data Q3	Model Q3	Data Q4	Model Q4
W/Y percentile 25%	0.059	0.0451	0.0593	0.653
W/Y percentile 50%	0.1024	0.1083	0.1248	0.1254
W/Y percentile 75%	0.1726	0.1765	0.2211	0.1839

Source: Author's computations. Objective on calibration is to minimize L2-norm between the model and data statistics.

Not surprisingly we find considerable heterogeneity in discount factors. In Table 1.2 we report the calibration results. We find an average discount factor of 0.985 which is slightly higher than the average of 0.96 found by [4]. We also find that the calibration points to considerable heterogeneity in discount factors, but with 90% of mass between 0.9 and 1.0 in annualized terms which is the interval considered in [4]. Finally, we find that the Gaussian copula correlation is negative, indicating that discount factors and ability have a slightly negative correlation. However, we do find a positive correlation between discount factors and labor income as endogenously individuals with a higher discount factor are also willing to work harder.

[21] This is the most common matching rate in the U.S. ([29])

Table 1.2: Calibration results

Parameter	Calibrated Value
a	0.855
b	1.795
ω	-0.107
$\mathbb{E}(\psi_{annual})$	0.985
90th percentile ψ_{annual}	0.999
10th percentile ψ_{annual}	0.905
$Corr(y, \psi)$	0.10

Source: Author's computations. Shape parameters (a, b) where $\psi \sim Beta(a, b)$ and for the correlation parameter ω between ψ and θ. Discount factor ψ is reported in terms of an equivalent annual discount factor where $\psi_{two-period} = \psi_{annual}^{40}\left(\frac{1-\psi_{annual}^{21}}{1-\psi_{annual}^{41}}\right)$.

Our calibration is broadly in line with other estimates in the literature. [4] find in PSID data a correlation of -0.02 between discount rates and the fixed effect on the labor income process on their estimation. However, they do find a correlation of -0.76 between the slope of the labor income profile and discount rates. In the behavioral economics literature, although [90] found no evidence of a correlation between present-bias and income levels in their experimental setting, [77] argues that different types of behavioral biases are more prevalent at the poor (e.g., addiction to cigarettes).

In Appendix A.5 we provide an alternative calibration using an incomplete markets life-cycle model with hyperbolic preferences. That model not only allows for a separate identification of β and δ, but also allows for a precautionary savings motive. We find that the calibrated distribution for $\psi = \beta\delta$ is similar in both calibrations. In fact, the heterogeneity in savings incentives (as measured by the effect time-inconsistency has on savings) is very similar on both models. Therefore, it seems that our benchmark calibration indeed produces a sensible calibration of the heterogeneity in discounting that is consistent with a behavioral explanation of the phenomena and parsimonious at the same time.

In fact, estimates in the behavioral literature for present-bias are largely consistent with the values of β we find in the life-cycle calibration. The distribution of β's we find are within the bounds estimated by [76] on the level of present bias in their

experimental data. Finally the average point estimate we find is consistent with the range of estimates in the behavioral economics literature. [67] find an estimate of 0.7 by calibrating a life-cycle model with hyperbolic preferences. More recently, [11] finds that on real effort tasks present bias is 0.88 in a laboratory experiment. [57] find a present bias of 0.34 in a field experiment providing commitment devices incentivizing savings of income tax returns in the United States.

1.4.2 Quantitative approximation to government problem

In this sub-section we compare the current policy in the U.S. with optimal policies arising from the normative model. To make an appropriate comparison, it is necessary to make sure that we choose a redistributive motive and a discount factor for the government that is consistent with the current levels of redistribution and with current savings policies. In that form, any inconsistency we find between current policy and the normative model prescription is not caused by the particular government preferences we choose.

We parameterize the Pareto weights the government assigns to agents by

$$\lambda(\theta) = \frac{\exp\{-\alpha\theta\}}{\sum_{\theta,\beta} \pi(\theta,\beta)\exp\{-\alpha\theta\}}$$

so that the parameter $\alpha \in \mathbb{R}$ represents the government's redistributive motive (again we follow [51] in using this functional form). If $\alpha = 0$ the government is utilitarian and if $\alpha > 0$ the government wants more redistribution than a utilitarian government would. The standard finding in the public finance literature is that the current level of income taxes in the U.S. is consistent with $\alpha < 0$ when ignoring savings policies.[22]

[22] See [50] and [51].

Numerical algorithm for solving two-dimensional screening problem

The literature has recognized that finding numerical solutions to two-dimensional screening problems is generally a difficult task. Many of the techniques that facilitate solving one-dimensional optimization problems fail in a multi-dimensional context. More specifically, we are no longer able to rely on the first-order approach, which reduces the number of relevant constraints to a small set of local incentive compatibility constraints. Instead, solving the problem generally requires the use of all incentive compatibility constraints in the solution routine. [58] point out that with the large number of incentive constraint there is a failure of the linear independence constraint qualification that leads conventional optimization routines to fail to find lagrange multipliers correctly.

We contribute to this literature by providing a stable and computationally efficient numerical algorithm to solve our problem even when naive solution approaches are unfeasible. Our algorithm reduces the complexity of the two-dimensional screening problem by searching for a subset of incentive constraints that are sufficient to obtain the solution to the global program. We solve a sequence of relaxed problems with high numerical accuracy for small subsets of incentive constraints.[23] If we find that the solution to this problem is globally incentive compatible, the convexity of the problem guarantees us this is the global solution.[24] This allows us to solve problems with a large number of incentive constraints.

The difficulty is then to find a suitable form of selecting subsets of incentive constraints. In fact, the number of subsets of a constraint set with many constraints is extremely large. Therefore, this whole strategy is only feasible if it is straightforward to converge to the correct set of constraints relatively fast.

[23] In our implementation we use IPOPT to solve each relaxed problem numerically.

[24] Writing the problem in terms of utility levels from consumption and disutility from work, it is easy to see the problem is convex.

It turns out a simple heuristic approach works very well in practice. We start with the first best problem with no constraints. Then we measure the relative violation of each incentive constraint in the original problem. We then add to the relaxed problem a fraction of the incentive constraints with the highest relative violation. We do so in a way that the total number of constraints in the problem does not grow too fast. In particular, we add only a fraction of the difference between the number of variables in the maximization problem and the number of constraints used in the last step. If this difference is very small, we add one constraint at a time.[25] After the initial step, we drop a random selection of slack constraints used in the relaxed problem.

This algorithm is guaranteed to converge with probability one to the set of incentive constraints that solve the global problem. Since the number of constraints is finite, always adding constraints guarantees convergence. Dropping constraints might generate a problem as it can generate cycles in the search for the correct subset of relevant incentive constraints, randomizing this selection allows us to prevent the algorithm from cycling forever (with probability one). In practice we find this algorithm converges relatively quickly (a few hours) in all our exercises with about 250k incentive constraints.[26]

Results

For any given level of α and $\delta \in (0, 1]$, we solve numerically the government's normative problem and compare the resulting allocation with the allocation arising in the calibrated economy with actual policies. In particular we find the values for the planner's preferences that minimize the L2-norm between the two allocations.[27] In

[25] In practice, this happens in intermediary steps when the initial set of constraints used is very different from the solution set.

[26] We have solved problems with up to 10 million constraints using this algorithm using the DELLA computer cluster at Princeton University.

[27] As a robustness check we also found planner's preferences minimizing the consumption equivalent welfare gain between the actual policy and the normative model. Results are qualitatively similar.

doing so, we follow the approach in [50, 51] in order to obtain a sensible level of redistribution and discounting by the government in our optimal policy computations.

We find that $\alpha = -0.60$ and $\delta_{annual} = 0.974$ better approximate the normative model allocation to the allocation in the calibrated model with current policies, and we find there are unexplored gains from the government's perspective according to the normative model.[28] Further, together with our calibration for ψ this implies that the disagreement of agents and the planner, $\beta = \frac{\psi}{\delta}$, is on average $\mathbb{E}(\beta) = 1.42$, that is to say, current policy is well rationalized if on average the government believes individuals are forward biased. Indeed, actual policy displays positive savings taxes and that is only consistent with the normative model if there is forward bias.[29] A consumption equivalent gain of 17.5% is available to the policymaker when we compare the allocation in the previously calibrated model and the approximated normative model.[30]

In Figure 1.5a we plot the proportion of retirement consumption ($\frac{c_2/R}{c_1+c_2/R}$) for the computed level of Pareto weights and discounting for the government in the benchmark normative model. We observe that at low earnings, there is little dispersion in retirement consumption whereas at high earnings there is a higher level of dispersion. In Figure 1.5b we plot the proportion of retirement consumption at the optimal allocation for a utilitarian planner with the same level of discounting.

The level of dispersion at low earnings level is very similar to the benchmark normative model, however there is considerably more dispersion at high earnings. A utilitarian planner is considerably more redistributive than the planner in the benchmark normative model, therefore it will offer considerably more flexibility on savings for high earnings individuals as that is a useful tool in improving redistribution. Conversely, the planner in the benchmark normative model cares considerably about

[28] We find $\delta = 0.22$ and here we report δ_{annual} which is the discount factor in annual terms.

[29] [76] indeed find a considerable level of forward bias present in their experiment.

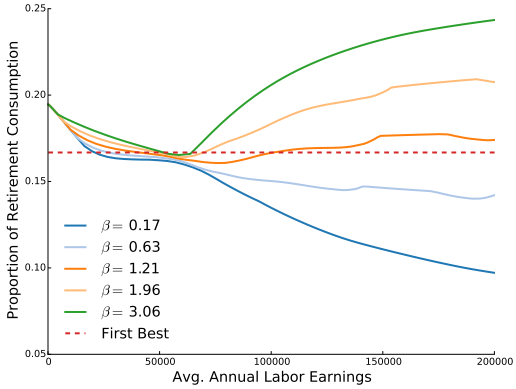
[30] In the robustness check where government gains from optimal policy are minimized, we still find a 13.5% gain available to the policymaker.

Figure 1.5: Proportion of retirement consumption for optimal allocations

(a) Benchmark normative model



(b) Utilitarian planner (same discounting as benchmark)



Source: Author's computations. Proportion of retirement consumption $\frac{c_2/R}{c_1+c_2/R}$ for optimal allocations. On the horizontal axis we have average labor earnings of the individual during its working life and on the vertical axis we have the proportion of retirement consumption at the optimal allocation. Each line in the graph represent the allocation of agents with a different level of β.

the payoff of high earnings individuals and therefore is particularly worried about their retirement consumption rates. As a result, in the benchmark normative model the planner offers an allocation with a much lower dispersion on the proportion of retirement consumption at high earnings.

In Figure 1.6 we plot the caps on retirement savings accounts arising from the benchmark normative model. Two accounts are used to a relevant extent and the caps are well approximated by piecewise-linear functions. Furthermore, low earnings

Figure 1.6: Caps on retirement savings accounts



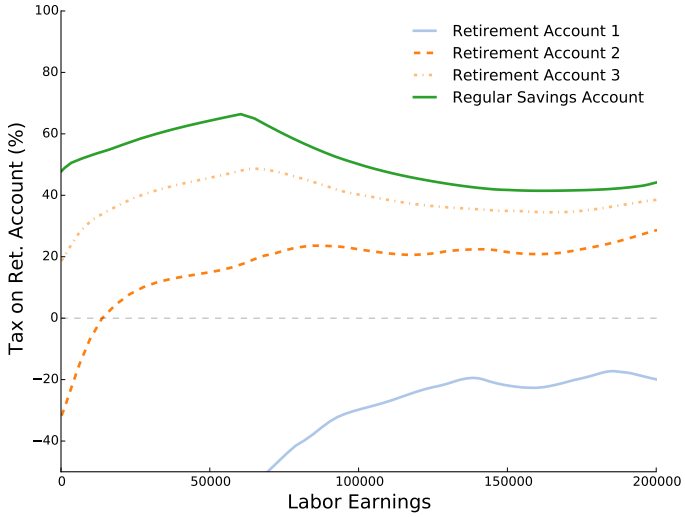
Source: Author's computations. Caps on retirement savings accounts on normative model that better approximate actual policy. On the horizontal axis we have average labor earnings of the individual during its working life and in the vertical axis we have the cap on the contribution to each of the 3 different retirement savings accounts.

individuals have very little access to these accounts. This points in the direction that optimal policy can be well approximated with simple policy tools.

In Figure 1.7 we plot the optimal levels of taxes (or subsidies if numbers are negative) on savings at the optimal allocation for the different retirement savings accounts. There are high taxes on savings as the planner in the benchmark normative model considers that some individuals are savings too much. There are also considerable tax breaks on the retirement savings accounts. Saving on retirement account 2 entails at least a 10% lower marginal tax rate on the return to savings as compared to savings on the regular savings account.

We now turn into the analysis of social security benefits. In order to understand the difference between current social security benefits and optimal old-age benefits implied by the normative model, it is important to understand how redistribution affects retirement benefits in our model. In Figure 1.8 we compare the current system of old-age benefits by earnings level with the normative model implication under the

Figure 1.7: Tax rates on savings



Source: Author's computations. On the horizontal axis we have average labor earnings of the individual during its working life and in the vertical axis we have the tax rate on savings. Each line represents a different retirement savings account or the regular savings account.

benchmark normative model ($\alpha = -0.6$) and for a utilitarian planner ($\alpha = 0$) with the same levels of δ. We see that a utilitarian government in our normative model would provide benefits with a qualitative shape more in line with the current system.

In order to check whether this wide difference between current retirement benefits and those arising from the normative model, we choose Pareto weights and discounting of the planner to match the current system of retirement benefits. In Table 1.3 we see that the redistribution motive indeed has to be close to utilitarian to explain current social security benefits. But if we take those as the government's preferences, then the available gain for the government of implementing optimal policy would be of around 63%. This highlights the difficulty in reconciling current social security benefits with current redistribution levels in the lenses of our normative model.

Finally, we check how much of the welfare gains available can be obtained with a simple policy change in the current system. We consider the following instruments:

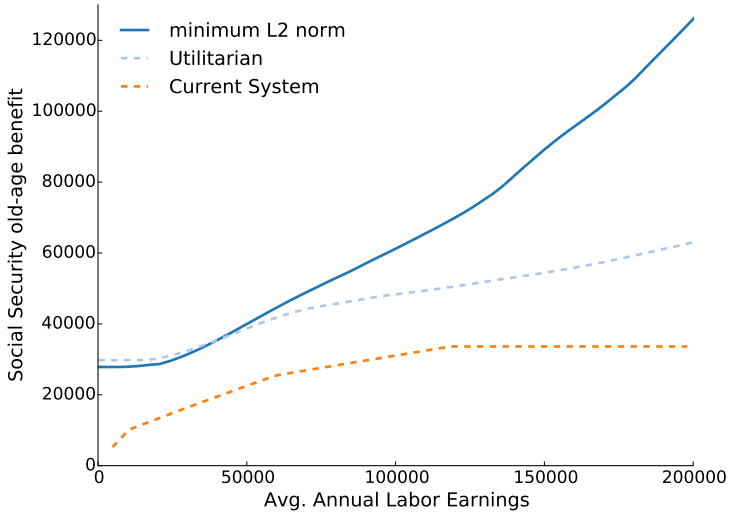


Figure 1.8: Old-age benefits by labor earnings level

Source: Author's computations. Optimal policy for benchmark normative model and utilitarian planner (same discounting level as benchmark model). On the horizontal axis we have average labor earnings of the individual during its working life and in the vertical axis we have the level of social security benefits at retirement. Each line represents the retirement benefits in a different system (current, benchmark normative model (L2 minimum norm) and utilitarian).

Table 1.3: Consumption equivalent welfare gains

Calibration	annual δ	α	Welfare Gain
Benchmark	0.974	-0.60	17.5%
Fitting SS benefits	0.940	0.15	62.9%

Source: Author's computations. Consumption equivalent welfare gains of moving to optimal policy for the different calibrations of the planner's preferences. The consumption equivalent gain is computed considering a uniform change in consumption during working life and retirement, but no change in labor supply.

tax schedule parameters λ and τ; abolishing the cap on social security benefits and allowing for a multiplier on current benefits by a parameter $\gamma > 0$ at all earnings levels; abolish IRA account and keep only one 401(k) account available with a minimum threshold on earnings for contributions, a cap that is a share of labor earnings and a match rate of 50%. Then using these policy instruments, we maximize the government's goal (with the same α and δ used in our approximation to current policy).

We find that with these instruments the government (with $\alpha = -0.6$ and $\delta = 0.974$) is able to obtain a sizable consumption equivalent welfare gain of 7%. In this exercise, we find much higher social security benefits ($\gamma = 1.94$) funded by higher income taxes ($\lambda = 0.74$ and $\tau = 0.41$). The minimum income threshold on the 401(k)-like account is \$34,775 average earnings and the cap on matching is 0.6% of labor earnings. As emphasized before, the main source of the gain for this regressive government is the increase in social security benefits to high earnings individuals. Of course, those results are for a particular government preference for redistribution and discounting as opposed to the previous exercise where we argued that there are no government preferences that can explain well both the current level of redistribution and the current level of retirement benefits. However, this final exercise highlights that simple instruments can achieve sizable welfare gains in this economy.

1.5 Extension: Behavioral and Pigouvian policies

Our main insights extend to a variety of behavioral and neoclassical models in which there is disagreement between agents and the government. The key element in the model is that at the time of decision individuals foresee heterogeneous costs (or gains) in their actions that are different than the costs (or gains) the government foresees. This wedge between government's objectives and agent's objectives is a common feature of several economic models.

mullainathan2012reduced uses an abstract "preference wedge" between the agents and the government to study a variety of models in behavioral economics as well as in classical economics. With behavioral agents a wedge arises from "internalities", which represent inefficient behavior that causes harm to the agent herself (e.g., present bias).[31]Conversely, in a classical model the wedge is caused by an externality in which

[31]In fact they show preference wedges arise naturally when agents have incorrect beliefs (over-optimism or over-pessimism) or inattention for example.

the agent cause harm (or benefit) to others. In both cases, agents and the government foresee different costs and gains from agents actions.

In this more general model, our insight implies that optimal policies involve an effective quantity restriction at low earnings whereas high earnings individuals are given more freedom in their choices if there is a redistributive motive. Furthermore, this more abstract setup allow us to show that without redistribution, a quantity restriction implements optimal policy.

There is a continuum of consumers with varying degrees of a generic "preference wedge." Action $a \in [0,1]$ has associated unit cost p and income-equivalent benefit b that depends on a as follows: $b(a) > 0$, $b'(a) > 0$, $b''(a) < 0$. In order to guarantee interior solutions, we also require $b'(0) > 0$ and $b'(1) < p$. Agents act according to their decision utility

$$\mathcal{U}^D(\theta, \alpha) = u(c) + \alpha\varepsilon(a) - \frac{1}{\theta}v(y)$$

where $\alpha\varepsilon(a)$ is a preference wedge and we assume $\varepsilon(a) > 0$, $\varepsilon'(a) > 0$, $\varepsilon''(a) < 0$ and $b(\varepsilon^{-1}(\cdot))$ is weakly concave.

However, experienced utility from consumption c and labor earnings y is given by

$$\mathcal{U}(\theta) = u(c) - \frac{1}{\theta}v(y)$$

where $u'(\cdot) > 0$, $u''(\cdot) < 0$, $v'(\cdot) > 0$, $, v''(\cdot) > 0$, $\lim_{y\to 0} v'(y) = 0$. Note that experienced utility depends on the action only through its effect on consumption. For $\alpha = 0$ we have $\mathcal{U}^D = \mathcal{U}$ and there is agreement between the government and the agent. For $\alpha \neq 0$, the agent's private cost (or gain) from action a differs from the social cost (or gain) from his action a. There is unobservable heterogeneity in $(\theta, \alpha) \in \Theta \times A$ distributed according to π with full support and where $A = \{\alpha_1, \ldots, \alpha_M\}$ and $\Theta = \{\theta_1, \ldots, \theta_N\}$ are finite sets as in our benchmark formulation. An allocation

$\{a(\theta,\alpha), c(\theta,\alpha), y(\theta,\alpha)\}_{(\theta,\alpha)\in\Theta\times A}$ is resource compatible if we have

$$\sum_{\theta,\alpha} \pi(\theta,\alpha)\left[c(\theta,\alpha) + pa(\theta,\alpha) - b(a(\theta,\alpha)) - y(\theta,\alpha)\right] \leq 0$$

In this model the first best level of the action a does not vary with the agent's type and is independent of redistribution.[32] In fact, at the first best all agent's choose a^* such that

$$b'(a^*) = 1$$

where $a^* \in (0,1)$ from our previous assumptions.

Laissez-Faire – Without government intervention, agents choose (a,c,y) to maximize $\mathcal{U}^D(\theta,\alpha)$ subject to the budget constraint $c+pa \leq y+b(a)$. In that case, agents with $\alpha \neq 0$ will generally choose $a^{LF}(\theta,\alpha) \neq a^*$ and we have $a^{LF}(\theta,\alpha) \neq a^{LF}(\theta,\alpha')$ for $\alpha \neq \alpha'$.

Quantity restriction - The implementation of a quantity restriction in the laissez-faire economy is efficient to the government that desires no redistribution. In fact, consider a quantity restriction in which the government forces $a = a^*$, but does not implement any redistribution policy so that agent's budget is still given by

$$c + pa^* \leq y + b(a^*)$$

It is then straightforward to show Lemma 3.

Lemma 3. *(Quantity restriction without redistribution) A quantity restriction of* $a(\theta,\alpha) = a^*$ *such that*

$$b'(a^*) = p$$

is efficient if it is implemented together with no redistribution policy.

[32] That was not the case in our benchmark model. In fact, in that model the planner would like an agent that receives more consumption in the initial period to also consume more in the second period.

Proof. In fact, let Pareto weights be given by $\lambda(\theta) = \frac{1}{u'(c^{LF*}(\theta))}$ where $c^{LF*}(\theta)$ is the level of consumption in the laissez-faire economy with the quantity restriction. □

The model in this section highlights the trade-off between paternalism and redistribution by disentangling the optimal first best level of a for the government from the redistributive motive. In fact, since the government believes preference wedges should be zero, then the first best level of the action a does not depend on the agent's type.

Theorem 4. *Fix θ_N and assume that $0 \in A$. If $\lambda(\theta)$ is weakly decreasing in θ, then there exists $\bar{\theta}_1 > 0$ such that at the solution to the planner's problem*

1. *if $\theta_1 < \bar{\theta}_1$, then all agents with types in $\{(\theta_1, \alpha) : \alpha \in A\}$ receive the same allocation (quantity restriction)*

2. *if $\theta_1 < \cdots < \theta_{N-1} \leq \bar{\theta}_N$, then agents with types in $\{(\theta_N, \alpha) : \alpha \in A\}$ do not receive all the same allocation.*

Theorem 4 shows that our main insight holds when there is over-consumption and when there is both over-consumption and under-consumption. Even though the signs of the distortions are undetermined in this case (so that Theorem 2 does not hold), Theorem 1 continues to hold and resulting optimal policies are very strict on low earnings individuals if there is a redistributive motive.

1.5.1 Applications: Environmental and drug policies

We now discuss two concrete applications of this general framework. In the first one, there is heterogeneity in how individuals value externalities generated by pollution of energy inefficient vehicles.[33] Some individuals do care considerably about pollution and buy more energy efficient cars, whereas some other individuals are not particularly

[33] Allcott201472 also studies energy policies that deal with externalities.

concerned with pollution. In the second, the government is concerned about drug usage and wills to reduce its consumption. Again, we assume there is considerable heterogeneity across individuals in preferences toward drug usage at all income levels.

Fuel Efficiency – In the case of energy efficiency, we can think of action $a \in [0,1]$ as the energy efficiency of the vehicle. There is a first best level of energy efficiency $a^* \in (0,1)$. The cost of a vehicle with energy efficiency $a \in (0,1)$ is given by pa. The benefit to society of allocating a vehicle to the agent is given by $b(a)$ in monetary terms which already take into account the cost of pollution. Agents have a preference wedge, in the sense they do not fully internalize the effect of pollution when purchasing a vehicle so that $\alpha < 0$ for some agents (and in general $\alpha \leq 0$).

If there is redistribution in this economy, optimal policy implies that low earnings individuals will purchase all at the same level of high energy efficiency according to the appropriate extension of Theorem 2 to this setup.[34] This policy can be implemented with income tax rebates on vehicle purchases that depend both on the level of energy efficiency on the car, but also on the earnings level of individuals. At low earnings levels, individuals receive a very high income tax rebate (or a tax credit) for a highly energy efficient car. At higher earnings levels, the government allows agents to enjoy energy inefficient vehicles in exchange for lower tax rebates (and hence an improvement in redistribution).

If there is no redistribution in this economy, then Lemma 3 implies that $a = a^*$ for all individuals. In this case, the government regulates fuel efficiency directly and forces all agents in the economy to purchase a vehicle with a level a^* of efficiency. We see in this example that the government is willing to trade-off a higher level of externalities from high earnings agents in exchange for an improvement in redistribution.

[34]Since we have $\alpha \leq 0$, that extension implies that at the lowest earnings level we have

$$b'(a(\theta_L, \alpha)) < 1$$

Drug Policy - In the case of drug policy, the action $a \in [0, 1]$ is drug consumption. The drug has a price of $p > 0$ per unit, and a social monetary benefit $b(a) = 0$ so that from the government's perspective, the optimal level of consumption is $a^* = 0$. However, some individuals disagree with the government and would like to consume drugs so that a $\alpha \geq 0$. Here we assume the government is actually able to control the sale of drugs, which is a rather strong assumption, however we think of it as an important benchmark.

With no redistribution, optimal drug policy is a quantity restriction $a = 0$, however with redistribution drug policy is considerably more heavy handed at low earnings individuals as compared to high earnings individuals. The government can implement optimal policy by offering prescriptions for drug usage whose cost varies with the earnings level of individuals and the quantity of drugs purchased. However, at low earnings levels those prescriptions are prohibitive so that agents are effectively prohibited to use drugs.[35]

1.6 Conclusion

In this paper we develop a normative model of paternalistic policies and show how redistribution plays a key role in shaping those policies. The main insight we find is that optimal policies are very restrictive on the behavior of low earnings individuals, but allow for more flexibility on the behavior of higher earnings individuals. This insight arises from a trade-off between paternalism and redistribution that is present in the model and that varies by earnings levels. Further, this insight implies that optimal retirement savings policies for behavioral agents will involve only social security (forced savings) at low earnings levels, but will involve offering a menu of retirement

[35] Another interpretation of such a policy at low earnings is drug testing of welfare recipients as those with the lowest level of earnings receive transfers.

savings accounts (similar to 401(k) and IRA accounts) to high earners on top of social security benefits.

We show that this insight is very general and hold in behavioral economics models with redistribution. In fact we construct one dynamic extension of the normative model in which agents have hyperbolic preferences and we show our insights hold true. Finally we show this insight holds generally when the government and agents disagree about the cost or gain of an action. This includes both behavioral economics models and also neoclassical models. Our theory implies quantity restrictions on low earnings individuals actions are a more efficient tool than linear taxes in achieving both better decisions from the government's perspective and improving redistribution. However, we show that it is efficient to relax quantity restrictions for high earnings individuals if there is a strong redistributive motive. In fact, by providing more flexibility to high earnings ability individuals, the government creates an extra incentive for them to work hard and actually obtain high earnings in the labor market.

In a quantitative evaluation, we show that actual policy has important differences to efficient policies arising from the normative model. In particular, we find that current social security benefits are consistent with a government that has close to utilitarian preferences, but at the same time the overall system of retirement savings policies and redistribution policies (income taxes and transfers) is better approximated by a more regressive government. This difference relies fundamentally on the heterogeneity in preferences and on two characteristics of the current income tax code, and the quantitative exercise just highlights those properties. First, there is widespread evidence for heterogeneity in discounting rates and in the level of present-bias. Second, the lack of progressivity in the U.S. tax code is a consensus in the public finance literature. In fact a utilitarian government would implement much higher top income tax rates than the currently implemented in the U.S.(see [86]). Further, [51] find that not only top income taxes are inconsistent with a progressive social planner,

but the overall tax structure in the U.S. is considerably more regressive than utilitarian. Finally, the cap on social security benefits for individuals with high earnings implies a considerable level of flexibility on their savings. Those three characteristics of the current economy are inconsistent with the normative model, exactly because a regressive planner is particularly worried about the behavior of high earnings agents and therefore would like to make sure they save on top of the current cap on social security benefits.

Finally, here we have focused our attention on paternalistic policies arising because of behavioral motives, due to the large body of evidence for present bias in the behavioral economics literature. However, there are important alternative explanations for paternalistic policies that might also play an important role. In particular, [53] and [22] find that altruistic behavior can be used to explain welfare policies. It would be interesting to check if our insight for paternalistic policies arising from behavioral biases would also hold in a setting where paternalistic policies arise for altruistic reasons and a Samaritan's dilemma.

Chapter 2

Taxing a bright future? Advance information and optimal labor income taxation

This paper studies the effect of heterogeneous expectations about future labor market outcomes on optimal labor income taxes. The recent literature on labor income processes ([46] and [49]) highlights that workers early in their careers already have considerable information about their future labor income. But there is not a full understanding of the effect this ex-ante information can have on optimal tax levels. I show that unobservable heterogeneity in expectations about future labor income fundamentally affects optimal top marginal income tax rates. In particular if those expectations have different Pareto tails, I show with quasi-linear preferences and an arbitrary level of redistribution that the top marginal income tax can reach 100%. Furthermore, I find that top marginal tax rates are lower for agents with a brighter future, and that in exchange those agents receive lower transfers [1]

[1]I want to thank Mike Golosov for invaluable advice, guidance and encouragement throughout this project. I would also like to thank Mark Aguiar, Benjamin Moll, Richard Rogerson and Juan Pablo Xandri and participants at the Princeton Macroeconomics Student Lunch Seminar for insightful comments and suggestions.

2.1 Introduction

There is an ongoing debate in the labor-economics and macroeconomics literature about the sources of uncertainty that people face on their labor market outcomes during the life-cycle. The predominant view during the last few decades has been that observable heterogeneity and uncertainty determine labor market outcomes ([70]). According to this view, the observable level of inequality in labor income is mostly due to luck over the life-cycle and not about initial conditions. The alternative view is that there is heterogeneity in initial conditions that explain a considerable portion of income inequality over the life-cycle ([68]). Recent findings in the empirical literature on labor earnings dynamics point to the importance of this alternative view.[2] Using data from the Panel Study of Income Dynamics, [47] finds that heterogeneity in life-cycle profiles can explain up to 75% of income inequality at age 55.

Furthermore, workers have a lot of information about their life-cycle profile early in their careers. [46] and [49] estimates that when entering the labor market workers have information that can explain more than 60% of the heterogeneity in life-cycle profiles. However, even with this considerable evidence for heterogeneity in expected labor market outcomes, the normative literature on optimal taxation has mostly followed the predominant view that labor income inequality arise mostly due to luck over the life-cycle (e.g., [32], [89] and [39]).

In this paper, I study optimal labor income taxation in a two-period model with heterogeneity in expected labor market outcomes. There are two sources of heterogeneity in the model: unobservable labor earnings ability in the labor market and unobservable ex-ante labor market profiles. People with a different profile face a different distribution of labor earnings ability into the future, but are otherwise equal in the current period. Intuitively, those people have different expectations about labor market outcomes.

[2] [54] and [48].

This paper makes four contributions to the taxation literature. First, it provides optimal tax formulas when there are heterogeneous labor earnings profiles using both non-linear and linear labor income taxes taxes. Second, it shows that optimal top marginal tax rates can reach 100% if there is heterogeneity on the Pareto tail of the labor ability distribution across different labor earnings profiles. Third, it provides a decentralization of optimal policy that explores self-selection of workers into different tax schedules. Lastly, in the case of linear taxes, it shows that workers with better earnings profiles will always choose tax schedules in which they face lower marginal taxes and receive lower transfers from the government as compared to those chosen by workers with worse labor earnings profiles.

By providing workers different tax schedules from which they can choose from beforehand, the government is able to diminish dead-weight losses from taxation. Workers with very high expectations about their future labor market outcomes are concerned about the marginal labor income tax rates they face, but not as much concerned about the transfers in case of a bad labor ability shock. Conversely, workers with low expectation about their future labor market outcomes are particularly worried about bad shocks and are willing to pay higher marginal income taxes in case of a high ability shock in exchange for higher transfers in case of low ability shocks. Therefore, the planner finds it optimal to offer different tax schedules for otherwise equal workers as a result of the heterogeneity in expectations.

I provide optimal tax formulas when there is heterogeneity in earnings profiles. I follow the approach in [27] and derive closed form solutions for taxes using quasi-linear preferences. The closed form expressions for tax rates allow me to study in detail the effect of earnings profile heterogeneity on top marginal income tax rates.

For the case without profile heterogeneity, [87] derived optimal top marginal income tax formulas. He shows that top marginal income taxes depend on a handful of sufficient statistics, such as the elasticity of labor supply at top incomes, and re-

distribution motives. One key intuition from his formula is that if the social planner assigns a very small welfare weight to high earnings individuals, then the optimal top marginal income tax is at the top of the "Laffer curve", i.e., it is at the level that maximizes tax revenue. Conversely, with unobservable heterogeneity in profiles, I find that top income tax rates can be substantially higher, approaching 100% for earnings profiles if the distributions across profiles have different Pareto tails.

The social planner offers individuals with relatively bad expectations about their future labor market outcomes an insurance scheme in which they receive high transfers, but they are taxed at extreme high marginal rates in case the labor market outcome is very good. Those tax rates prevent people that expect a very good future labor market outcome from pretending to have a worse earnings profile. As a result, the planner is able to improve redistribution and at the same time impose lower distortions on labor decisions of very high earners with an ex-ante better earnings profile.

The taxation literature focuses attention on taxation schemes that depend on observable outcomes, such as the labor earnings level, in [75] for instance, the saving level, in my job market paper, or even height, in [72] for example. In those schemes people are not given a choice of which tax scheme they would prefer. This is true even in the dynamic taxation literature where taxes are allowed to depend on the history of observable labor earnings. In our setup, people are absolutely identical ex-ante except for their unobservable heterogeneous profiles, any variation in the tax system across profiles has to come from self-selection and individual choice, and not from observable choices of labor supply or consumption.

In particular, the optimal tax formulas derived in this paper also hold in the case the government is constrained not to use some observable piece of information. For instance, the constitution might prevent the government from charging different tax levels from workers with different educational levels but the same amount of

labor income. Using data from the Survey of Consumer Finances, I find that there are significant differences in the Pareto tail of the labor income distribution across different educational levels. This implies that a self-selecting tax system would lead to large differences in top marginal income tax rates faced by workers at different educational levels.

This paper is organized as follows. Section 2.2 develops the model with linear taxation and obtain closed form solution for optimal taxes. Then Section 2.3 characterizes the solution to mechanism design problem the social planner faces, it provides a decentralization to the solution of the planner's problem, and it shows that top income taxes are greatly affected by heterogeneity in Pareto tails of the labor earnings ability distribution across different labor earnings profiles. Section 2.4 provides empirical evidence that Pareto tails can differ considerably across groups in the population with different educational levels, and it shows that this has important implications for taxation both in the case in which policy can be made directly conditional on the educational choice as well as in the case in which policy cannot be made conditional on educational choices. Finally, Section 2.5 concludes the paper.

2.2 Model with linear taxation

This section proposes a baseline model to consider the effect of advance information (heterogeneous profiles) into the design of optimal linear taxes and lump-sum transfers. There are three parts to this analysis. First, it describes the basic assumptions on preferences and on what the government observes. Second, it provides an example with only two-profiles in which I show how self-selection is useful to improve the efficiency of labor income taxation. Lastly, this section provides optimal tax formulas for linear tax systems that make usage of self-selection.

I start with the model description. There is a unit mass of workers that live for two periods. In the first period agents learn their labor market profile $p \in P = [\underline{p}, \bar{p}]$ whose probability distribution is denoted by $H(p)$ with a density $h(p)$. In the second period, agents learn their labor productivity type $\theta \in \left[\underline{\theta}, \bar{\theta}\right]$ with a distribution $F(\theta|p)$ where p is the profile of the agent realized in the first period. Distributions are ordered so that the distribution for a higher profile p dominate in first order stochastic terms distributions for lower profile levels. Agents work in the second period and their preferences are ordered by[3]

$$u(c, \ell) = c - \frac{\ell^{1+\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}}$$

where c is the level of consumption and ℓ is labor effort. Labor output of a worker is $y = \theta\ell$, so that it depends both on work effort and on the level of labor earnings ability of the worker. The government observes (c, y), but not (p, θ) or ℓ. Furthermore, the government and workers can communicate in both periods.

2.2.1 Example with two profiles

To gain intuition about the usefulness of self-selection, let's first consider an example with affine tax schedules and two labor earnings profiles ($p_H > p_L$). Workers receive a lump-sum transfer from the government T and pay a marginal tax τ on labor earnings.[4] Therefore a worker has a budget given by

$$c \leq T + (1-\tau)\, y$$

Then the indirect utility function in the second period is

$$\hat{U}(T, \tau; \theta) = T + \frac{1}{1+\varepsilon}\left[(1-\tau)\,\theta\right]^{1+\varepsilon}$$

[3] In Section 2.3, I relax the assumption of iso-elastic disutility from labor effort.

[4] Alternatively, one can think of T as a uniform deduction on income taxes.

and the policy function is $y^*(\theta;\tau) = (1-\tau)^{\varepsilon}\theta^{1+\varepsilon}$. In the first, period, the indirect utility of workers with an earnings profile p is

$$\hat{V}(T,\tau;p) = T + \frac{1}{1+\varepsilon}(1-\tau)^{1+\varepsilon}\,\mathbb{E}\left[\theta^{1+\varepsilon}|p\right]$$

Since $\mathbb{E}[\theta^{1+\varepsilon}|p]$ is monotonically increasing in the profile p, the "marginal rate of substitution" between transfers T and marginal taxes τ is different for different earnings profiles. Therefore, the indirect utility actually satisfies single-crossing in the earnings profile for the choice between the two forms of taxation. As a result, the government can separate workers with different profiles.

Therefore, the government in general can use a tax menu in which workers self-select into the optimal tax for themselves. Let's denote by $(T(p_H), \tau(p_H))$ and $(T(p_L), \tau(p_L))$ two of those tax schedules. In the following paragraphs, I show that the government finds it optimal to set $\tau(p_L) > \tau(p_H)$ when the redistributive motive is sufficiently strong. In this case, workers with a brighter future then face lower marginal taxes than workers that have low expectations about their labor market outcomes.

This result highlights the relevance of the discussion in the labor economics literature about the nature of the labor income process. On one hand, if workers have considerable information about their future labor market prospects, optimal taxation involves offering different tax schedules so that workers can self-select according to their expectations. On the other hand, if workers have homogeneous expectations and face only uncertainty, then optimal taxes involve offering a single tax schedule and there is no role for self-selection.

Assume that the redistributive motive is sufficiently high, so that

$$\left|\mathrm{Cov}\left(\alpha(\theta), \theta^{1+\varepsilon}|p\right)\right|$$

is strictly decreasing in the earnings profile. For instance, this happens when $\lim_{\theta\to\infty}\alpha(\theta)=0$ and high profile workers have a distribution of ability with sufficient higher weight on higher ability levels as compared to the distribution of ability of low profile workers.

Then, consider the government perturbs the tax schedule for workers with a high profile, but that it keeps the tax schedule for low earnings profile workers unchanged. The perturbation is such that the payoff of high profile workers in unchanged:

$$d\hat{V}\left(T\left(p_H\right),\tau\left(p_H\right);p_H\right)=dT\left(p_H\right)-\left(1-\tau\left(p_H\right)\right)^{\varepsilon}\mathbb{E}\left[\theta^{1+\varepsilon}|p_H\right]d\tau\left(p_H\right)=0$$

for $d\tau(p_H)<0$ and $dT(p_H)<0$. Then we have $d\hat{V}\left(T\left(p_H\right),\tau\left(p_H\right);p_L\right)<0$ so that workers with a low earnings profile dislike the new schedule and will keep using the one originally designed for them. However, high earnings profile workers are just indifferent between the two schedules and I assume they decide to use the new one.[5]

The change in net government spending with high profile workers is composed of the mechanical and behavioral effects, as in [87]: there is a change in tax revenue because of the direct change in the tax rate, taking as given workers choices (mechanical effect), and there is a change in tax revenue because workers choices change as a result of the tax change (behavioral effect). Computing those two effects in this example, we get

$$dR=-\varepsilon\tau\left(p_H\right)\left(1-\tau\left(p_H\right)\right)^{\varepsilon-1}h\left(p_H\right)\mathbb{E}\left[\theta^{1+\varepsilon}|p_H\right]d\tau\left(p_H\right)>0$$

Therefore, the lower tax revenue caused by lower marginal tax rates is more than compensated by the change in transfers. Intuitively, marginal taxes lead to deadweight losses and lump-sum transfers do not, so that relying relatively more on lump-

[5]It is possible to slightly change the perturbation to make high profile workers strictly better off, however it complicates the exposition considerably.

sum taxes (less on lump-sum transfers) for high profile workers lead to higher revenue to the government.

However, the change in the tax schedule also leads to a welfare effect. Indeed, there is a reduction in social insurance of earnings shocks across agents with a high profile. In order to compute the total change in welfare, we need to take into account both the value of extra funds obtained and the direct welfare effect on workers with a high earnings profile. Since preferences are quasi-linear, the social cost of funds is unitary in this economy. Therefore the total change in welfare is given by

$$\begin{aligned} d\mathcal{W} = & h\left(p_H\right)\left[\mathbb{E}\left[\alpha\left(\theta\right)|p_H\right]dT\left(p_H\right) - \left(1-\tau\left(p_H\right)\right)^{\varepsilon}\mathbb{E}\left[\alpha\left(\theta\right)\theta^{1+\varepsilon}|p_H\right]d\tau\left(p_H\right)\right] \\ & - \varepsilon\tau\left(p_H\right)\left(1-\tau\left(p_H\right)\right)^{\varepsilon-1}h\left(p_H\right)\mathbb{E}\left[\theta^{1+\varepsilon}|p_H\right]d\tau\left(p_H\right) \end{aligned}$$

where the first line in the expression represents the direct welfare change and the second line the gain in resources. It is useful to re-write this expression using the expression for $T\left(p_H\right)$ and rearranging:

$$\frac{d\mathcal{W}}{h\left(p_H\right)\left(1-\tau\left(p_H\right)\right)^{\varepsilon}} = -\left[\text{Cov}\left(\alpha\left(\theta\right),\theta^{1+\varepsilon}|p_H\right) + \varepsilon\frac{\tau\left(p_H\right)}{1-\tau\left(p_H\right)}\mathbb{E}\left[\theta^{1+\varepsilon}|p_H\right]\right]d\tau\left(p_H\right)$$

At the solution to the planner's problem it cannot be the case that $d\mathcal{W} > 0$, therefore from the condition above we obtain that it must be the case that

$$-\text{Cov}\left(\alpha\left(\theta\right),\theta^{1+\varepsilon}|p_H\right) \geq \varepsilon\frac{\tau\left(p_H\right)}{1-\tau\left(p_H\right)}\mathbb{E}\left[\theta^{1+\varepsilon}|p_H\right]$$

at the solution to the government's problem. The covariance term in the left hand side of this inequality captures the direct welfare loss as social insurance across workers with a high earnings profile is reduced. The left hand side captures the social gain from the extra resources made available with the tax change. As a result, optimal

taxes for high earnings profile workers must satisfy

$$-\frac{1}{\varepsilon}\frac{\text{Cov}\left(\alpha\left(\theta\right),\theta^{1+\varepsilon}|p_H\right)}{\mathbb{E}\left[\theta^{1+\varepsilon}|p_H\right]} \geq \frac{\tau\left(p_H\right)}{1-\tau\left(p_H\right)}$$

An analogous perturbation in the tax schedule of workers with a low profile shows that optimal taxes satisfy

$$\frac{\tau\left(p_L\right)}{1-\tau\left(p_L\right)} \geq -\frac{1}{\varepsilon}\frac{\text{Cov}\left(\alpha\left(\theta\right),\theta^{1+\varepsilon}|p_L\right)}{\mathbb{E}\left[\theta^{1+\varepsilon}|p_L\right]}$$

From the assumption that the government has a strong redistributive motive, we then have that $\tau\left(p_L\right) > \tau\left(p_H\right)$, so that the government finds it optimal to charge a lower marginal tax rate in the tax schedule offered to high earnings profile workers. Of course, this is in exchange for lower transfers that high earnings profile workers face.

2.2.2 Optimal linear taxes in general case

In this sub-section, I show that linear taxation menus offered to workers by the government always involve weakly lower marginal tax levels for workers with higher earnings profiles. Moreover, I characterize optimal taxes when the government fully separates all different earnings profiles into different tax schedules.

A tax menu $\left(T\left(p\right),\tau\left(p\right)\right)_{p\in P}$ is incentive compatible if

$$\hat{V}\left(T\left(p\right),\tau\left(p\right);p\right) \geq \hat{V}\left(T\left(p'\right),\tau\left(p'\right);p\right) \tag{2.1}$$

for all $p, p' \in P$. But notice that this implies

$$\left[\left(1-\tau\left(p\right)\right)^{1+\varepsilon} - \left(1-\tau\left(p'\right)\right)^{1+\varepsilon}\right]\left[\mathbb{E}\left[\theta^{1+\varepsilon}|p\right] - \mathbb{E}\left[\theta^{1+\varepsilon}|p'\right]\right] \geq 0$$

and therefore I just proved:

Lemma 5. *In any incentive compatible tax menu* $(T(p), \tau(p))_{p \in P}$*, the marginal labor tax* $\tau(p)$ *is weakly decreasing in the earnings profile, and the level of government transfers* $T(p)$ *is weakly decreasing in the profile as well.*

This result highlights that with affine taxes, it is impossible for the government to charge workers with a high profile a strictly higher marginal tax rate. The example in the last sub-section show however that the government might actually want to charge lower taxes for those high profile agents as that can improve tax revenue considerably by reducing the level of government transfers (or tax deductions).

Since indirect utility over tax schedules satisfy single crossing, it is possible to solve the government problem of finding the optimal tax menu $(T(p), \tau(p))_{p \in P}$ using standard techniques in contract theory. In particular, if we define $\hat{V}(p) = \hat{V}(T(p), \tau(p); p)$, then incentive constraint (2.1) is equivalent to the envelope condition

$$\frac{\partial \hat{V}(p)}{\partial p} = \frac{1}{1+\varepsilon} (1 - \tau(p))^{1+\varepsilon} \frac{\partial \mathbb{E}[\theta^{1+\varepsilon}|p]}{\partial p} \tag{2.2}$$

for all p and the requirement that $\tau(p)$ be weakly decreasing in p. If we relax the planner problem by ignoring the monotonicity constraint in marginal taxes, we obtain the following Theorem.

Theorem 6. *At the solution to the relaxed planner's problem with linear tax menus, taxes satisfy*

$$\frac{\tau(p)}{1-\tau(p)} = \frac{1}{\varepsilon} [D_0(p) - D_1(p)]$$

where

$$D_0(p) = -\frac{Cov(\alpha(\theta), \theta^{1+\varepsilon}|p)}{\mathbb{E}[\theta^{1+\varepsilon}|p]}$$

and

$$D_1(p) = \frac{\frac{\partial \mathbb{E}\left[\theta^{1+\varepsilon}|p\right]}{\partial p}}{\mathbb{E}\left[\theta^{1+\varepsilon}|p\right]} \left(\frac{1-H(p)}{h(p)}\right) \int_p^{\bar{p}} \left(\frac{h(p')}{1-H(p)}\right) (1 - \mathbb{E}(\alpha(\theta)|p')) \, dp'$$

for all $p \in [\underline{p}, \bar{p}]$. *If* $\tau(p)$ *defines in this way in monotonically decreasing, then* $\tau(p)$ *is the solution to the original government's problem.*

Proof. See Appendix. □

As in the previous example with two profiles, the first term in optimal taxes represent the redistribution motive within workers with an earnings profile p. If the covariance between welfare weights and labor ability is very negative, marginal taxes are higher so as to provide more social insurance. However, in the optimal tax formula, it is possible to inspect that taxes not only depend on social insurance within an earnings profile, but also on social insurance across earnings profiles. The second term depends on how different profiles are from each other in terms of their earnings ability distribution, and also on how average social welfare weights vary with the earnings profile.

2.3 Planner's problem and non-linear taxation

This section extends the results in the previous section by allowing for an arbitrary set of policy instruments for the government. There are three important contributions in this section to the general argument of the paper. First, I characterize the solution to the general social planner's problem in an economy in which workers have heterogeneous earnings profiles. Second, I show how the general solution to the social planner's problem can be decentralized using tax menus that rely on self-selection by workers. Lastly, I derive the implications for top income tax rates of heterogeneity in the upper tail of the earnings distribution across different labor earnings profiles.

There is a unit mass of agents that live for two periods. In the first period agents learn their labor market profile $p \in P = [\underline{p}, \bar{p}]$ whose probability distribution is denoted by $H(p)$ with a density $h(p)$. In the second period, agents learn their labor productivity type $\theta \in \left[\underline{\theta}, \bar{\theta}\right]$ with a distribution $F(\theta|p)$ where p is the profile of the agent realized in the first period. Distributions are ordered so that the distribution for a higher profile p dominate in first order stochastic terms distributions for lower profile levels. Agents work in the second period and their preferences are ordered by

$$U_2(c, y; \theta) = c - v\left(\frac{y}{\theta}\right)$$

where c is the level of consumption and y denotes labor earnings. Here I relax the assumption that the disutility from work effort is iso-elastic. The planner observes (c, y), but (p, θ) is not observable.

The planner is able to meet with agents in both periods so that it can design a policy that depends on decisions of agents in both periods. By the Revelation principle I can restrict attention to direct mechanisms $\{c(p, \theta), y(p, \theta)\}$. Further, I assume that $F(\cdot|p)$ has a density $f(\cdot|p)$ that has full support for all $p \in P$ so that I can restrict attention to mechanisms in which agents report their current type only.[6] An allocation $\{c(p, \theta), y(p, \theta)\}$ is incentive compatible if

$$c(p, \theta) - v\left(\frac{y(p, \theta)}{\theta}\right) \geq c(p, \theta') - v\left(\frac{y(p, \theta')}{\theta}\right) \tag{2.3}$$

for all $\theta, \theta' \in \left[\underline{\theta}, \bar{\theta}\right]$ and

$$\int_{\underline{\theta}}^{\bar{\theta}} \left[c(p, \theta) - v\left(\frac{y(p, \theta)}{\theta}\right)\right] f(\theta|p)\, d\theta \geq \int_{\underline{\theta}}^{\bar{\theta}} \left[c(p', \theta) - v\left(\frac{y(p', \theta)}{\theta}\right)\right] f(\theta|p)\, d\theta \tag{2.4}$$

[6] Otherwise it would be necessary to report the full history of types in the second period.

for all $p, p' \in P$. Define $U(\theta|p) = c(p,\theta) - v\left(\frac{y(p,\theta)}{\theta}\right)$. Following the standard approach in mechanism design theory, incentive constraint (2.3) is equivalent to the envelope condition

$$U(\theta|p) = U(\underline{\theta}|p) + \int_{\underline{\theta}}^{\theta} \frac{y\left(p,\tilde{\theta}\right)}{\tilde{\theta}^2} v'\left(\frac{y\left(p,\tilde{\theta}\right)}{\tilde{\theta}}\right) d\tilde{\theta} \tag{2.5}$$

for all $\theta \in \left[\underline{\theta}, \bar{\theta}\right]$ and $y(p,\cdot)$ weakly increasing in θ.[7]

I further relax the incentive problem by assuming also using the envelope condition for the report of the profile in the first period. In this case the envelope condition that arises from incentive constraint 2.4 is defined by

$$V(p) = V(\underline{p}) + \int_{\underline{p}}^{p} \left(\int_{\underline{\theta}}^{\bar{\theta}} U(\theta, p') \frac{\partial f(\theta|p')}{\partial p} d\theta \right) dp' \tag{2.6}$$

for all $p \in [\underline{p}, \bar{p}]$, where we define

$$V(p) = \int_{\underline{\theta}}^{\bar{\theta}} U(\theta, p) f(\theta|p) \, d\theta$$

for all $p \in [\underline{p}, \bar{p}]$.

The resource constraint in the economy is

$$\int_{\underline{p}}^{\bar{p}} h(p) \int_{\underline{\theta}}^{\bar{\theta}} \left[c(p,\theta) - y(p,\theta)\right] f(\theta|p) \, d\theta dp \leq R \tag{2.7}$$

where R is net resources available to the government.[8] The government evaluates welfare in this economy according to

$$\int_{\underline{p}}^{\bar{p}} h(p) \int_{\underline{\theta}}^{\bar{\theta}} \alpha(\theta) \left[c(p,\theta) - v\left(\frac{y(p,\theta)}{\theta}\right)\right] f(\theta|p) \, d\theta dp \tag{2.8}$$

[7] [74]

[8] In the usual case, R is given by other government spending and therefore $R < 0$.

where $\alpha(\theta) \geq 0$ is a set of Pareto weights normalized so that

$$\int_{\underline{p}}^{\bar{p}} h(p) \int_{\underline{\theta}}^{\bar{\theta}} \alpha(\theta) f(\theta|p) \, d\theta dp = 1$$

The planner's problem then is to maximize (2.8) subject to incentive constraints (2.4), (2.3), and resource constraint (2.7). In the following I consider the solution to the relaxed planner's problem where I substitute incentive constraints (2.4) and (2.3) by their local versions (2.6) and (2.5).

2.3.1 Optimal non-linear labor wedges

In this section, I characterize the optimal labor wedge at the solution to the social planner's problem without restrictions on tax instruments. I first show that labor wedges also depend both on redistribution within earnings profiles and redistribution across earnings profiles. Then I show that if Pareto tails of the distribution of ability differ by profile, then optimal top labor wedges can approach 100%. In the next section, I explore this result further.

Before characterizing the solution to this problem, it is useful to define a few objects. The labor wedge

$$\tau(p, \theta) = 1 - \frac{1}{\theta} v' \left(\frac{y(p, \theta)}{\theta} \right)$$

is a natural measure of the distortion on labor choices of an individual with type (p, θ). Without redistribution, the first order condition on the individual problem gives $\tau(p, \theta) = 0$. However, with a redistributive motive, the social planner distorts the individual allocation and that distortion can be measured by the labor wedge. Furthermore, there is a close link between labor wedges and marginal taxes on labor earnings. In fact, in the decentralization proposed in Section 2.3.3, the marginal labor tax faced by individuals equals the labor wedge they face. Finally, I denote by

$\varepsilon(p,\theta)$ the Frisch elasticity of labor supply for an individual of type (p,θ) at a given allocation.

Theorem 7. *At the solution to the relaxed planner's problem, the optimal labor wedge satisfies*

$$\frac{\tau(p,\theta)}{1-\tau(p,\theta)} = \left(1+\frac{1}{\varepsilon(p,\theta)}\right)\left[A(p,\theta)+B(p,\theta)\right]$$

where

$$A(p,\theta) = \underbrace{\frac{1-F(\theta|p)}{\theta f(\theta|p)}}_{A_1(p,\theta)}\underbrace{\left[1-\int_{\theta}^{\bar{\theta}}\alpha(\theta')\frac{f(\theta'|p)}{1-F(\theta|p)}d\theta'\right]}_{A_2(p,\theta)}$$

and

$$B(p,\theta) = \underbrace{\frac{\int_{\theta}^{\bar{\theta}}\frac{\partial f(\theta'|p)}{\partial p}d\theta'}{\theta f(\theta|p)}}_{B_1(p,\theta)}\underbrace{\left(\frac{1-H(p)}{h(p)}\right)}_{B_2(p)}\underbrace{\left[1-\int_{p}^{\bar{p}}\int_{\underline{\theta}}^{\bar{\theta}}\alpha(\theta')f(\theta'|p')d\theta'\frac{h(p')}{1-h(p)}dp'\right]}_{B_3(p)}$$

for all $\theta\in\left[\underline{\theta},\bar{\theta}\right]$ *and* $p\in[\underline{p},\bar{p}]$.

Proof. See Appendix. □

This tax formula has two components related to redistribution within an earnings profile and redistribution across earnings profiles. The term $A(\theta,p)$ is a standard term in tax formulas ([27] and [87]). It's first component, $A_1(p,\theta)$ captures the thickness of the tail of the distribution of labor ability at a given profile p. The second component, $A_2(p,\theta)$ captures the redistribution motive of the social planner within the income profile p. If social welfare weights converge rapidly to zero at high earnings ability, then $A_2(p,\theta)$ converges to zero as earnings ability increase so that the thickness of the tail of the distribution of labor ability and the elasticity of labor supply determine the contribution of this term for labor wedges at top earnings ability.

The term $B(p,\theta)$ in the formula captures the contribution of heterogeneity in profiles to optimal labor wedges. It depends on the following three components: (1)

how the distribution of earnings ability varies with earnings profiles; (2) the shape of the distribution of earnings profiles; and, (3) the redistributive motive of the planner across different earnings profiles.

The redistributive motive of the social planner across different earnings profiles is captures by the term $B_3(p)$. From our assumption that higher profiles have a distribution of ability that is better in first-order terms, then term $B_3(p)$ is positive and decreasing in the earnings profile level p as long as $\alpha'(\cdot) < 0$. The shape of the distribution of profiles is captured by term $B_2(p)$. This term has a positive sign. Notice that those two terms do not depend on earnings ability, therefore they have a uniform effect on the labor wedge across all earnings levels at a given profile.

Finally, $B_1(p, \theta)$ measures changes in the distribution of ability across different earnings profiles. This component has a positive sign as higher earnings profile have a distribution that dominates in first order terms lower earnings profiles distributions of ability. Further, it is very important to understand how this term depends on the Pareto tail of the earnings ability distribution. Assume that for high enough ability we have $1 - F(\theta|p) \propto \left(\frac{\theta_0}{\theta}\right)^{\rho(p)}$, then we have for high enough θ that

$$B_1(p, \theta) \propto \frac{\rho'(p)}{\rho(p)} \log\left(\frac{\theta_0}{\theta}\right)$$

If $\rho'(p) < 0$, we then obtain $\lim_{\theta\to\infty} B_1(p, \theta) = +\infty$. Thus in that case we have $\lim_{\theta\to\infty} \tau(p, \theta) = 1$. In the next section we explore further the insight that top income taxes can become arbitrarily large if Pareto tails are different across different earnings profiles.

2.3.2 Top income tax rates

In this section we make further assumptions on the distribution of profiles and we then characterize labor wedges for top earnings ability individuals. In particular, I

simplify the problem so that there is no need to rely on the first order approach in the reporting of profiles in the first period. I characterize top marginal income taxes in the case with two different profiles.

Let's first start with the benchmark case in the literature. Assume that $P = \{p_L, p_H\}$ and $\bar{\theta} = \infty$ so that there are only two earnings profiles and the distribution of earnings ability is unbounded. If $p_L = p_H$ and $\alpha'(\theta) < 0$, [27] and [87] show that

$$\lim_{\theta \to \infty} \left(\frac{\tau(\theta)}{1 - \tau(\theta)} \right) = \left(1 + \frac{1}{\varepsilon} \right) \frac{1}{\rho}$$

where $\varepsilon = \lim_{\theta \to \infty} \varepsilon(\theta)$ and $\rho = \lim_{\theta \to \infty} \frac{\theta f(\theta)}{1 - F(\theta)}$ is the thickness of the Pareto tail of the distribution F. This tax rate maximizes the revenue obtained from high earnings ability individuals, i.e., it is at the top of the Laffer curve. The next theorem considers the case

Now let's consider the case with heterogeneous earnings profiles. Denote by ρ_H and ρ_L the Pareto tail parameters of the distributions $F(\cdot|p_L)$ and $F(\cdot|p_H)$, respectively.

Theorem 8. *Assume that* $\rho_L = \rho_H$, $v(\ell) = \frac{\ell^{1+\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}}$ *and* $\alpha'(\theta) < 0$, *then*

$$\lim_{\theta \to \infty} \left(\frac{\tau(\theta, p_L)}{1 - \tau(\theta, p_L)} \right) = \left(1 + \frac{1}{\varepsilon} \right) \frac{1}{\rho} \int_{\underline{\theta}}^{\infty} \alpha(\theta) f(\theta|p_L) \, d\theta$$
$$\lim_{\theta \to \infty} \left(\frac{\tau(\theta, p_H)}{1 - \tau(\theta, p_H)} \right) = \left(1 + \frac{1}{\varepsilon} \right) \frac{1}{\rho} \int_{\underline{\theta}}^{\infty} \alpha(\theta) f(\theta|p_H) \, d\theta$$

and

$$1 > \lim_{\theta \to \infty} \tau(\theta, p_L) > \lim_{\theta \to \infty} \tau(\theta, p_H) > 0$$

at the solution to the relaxed planner's problem.

Proof. See Appendix. □

In Theorem 8, we find that the optimal top labor wedge is lower for individuals with a high earnings profile. Intuitively, the planner separates individuals with different earnings profiles by offering different distortions. Individuals with a high earnings profile are particularly worried about distortions at high earnings whereas individuals with a low earnings profile are not that worried about distortions at high earnings. As a result, the planner can separate those two types of profiles by offering higher distortion for low earnings profiles at high ability.

Theorem 9. *Assume that* $\rho_H < \rho_L$, $v(\ell) = \frac{\ell^{1+\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}}$ *and* $\alpha'(\theta) < 0$, *then*

$$\lim_{\theta\to\infty} \tau(\theta, p_L) = 1 > \lim_{\theta\to\infty} \tau(\theta, p_H) > 0$$

and $\tau(\theta, p_L) < 1$ for all $\theta > \underline{\theta}$.

Proof. See Appendix. □

The result in Theorem 9 is striking: individuals with a low earnings profile face labor wedges that converge asymptotically to 100%. Conversely, individuals with a high earnings profile face positive but smaller labor wedges. With different Pareto tails, the probability that an agent with very high ability had actually a low earnings profile becomes arbitrarily small as ability increases. As a result, the planner distorts considerably the allocation of those agents in order to prevent deviations from high earnings profile agents which have a much higher probability of having high ability.

2.3.3 Self-Selection and Taxation

In this section, I provide a decentralization of optimal policy that highlights self-selection as an indispensable policy tool. In Section 2.2, I showed with affine taxes how useful self-selection can be for optimal taxation. In this section, I build up on that insight and the results in Section 2.2 for labor wedges to show that non-linear

labor taxes should also use self-selection when workers have considerable information about their labor market prospects.

In the decentralization, the government offers workers a menu of non-linear tax schedules in the first period, denoted by $\mathcal{T} = \{T(\cdot|p) \,|p \in P\}$ where each $T \in \mathcal{T}$ is a different non-linear labor income tax that assigns a tax bill $T(y)$ for a worker with before taxes labor earnings y. Given a tax schedule T, in the second period, workers with earnings ability θ solve the problem

$$U(\theta, T) = \max_{c,y} c - v\left(\frac{y}{\theta}\right)$$
$$s.t. c \leq y - T(y)$$

where y denotes labor earnings and c denotes consumption. In the first period, workers decide which tax schedule they want to face later in life. Depending on their earnings profile, workers might choose a different tax schedule, so that their problem is

$$\max_{T \in \mathcal{T}} \mathbb{E}\left[U(\theta, T) \,|p\right]$$

Intuitively, the planner can explore differences in workers expectations to improve taxation. On one hand, workers that expect a very good labor market outcome are willing to forego social insurance of bad shocks for lower marginal labor taxes at high earnings. On the other hand, workers that expect a not so good labor market outcome are particularly worried about low ability shocks and might be willing to pay higher marginal taxes at high earnings levels. As a result, the government finds it optimal to offer workers a tax menu, instead of a single tax schedule.

I define a *competitive equilibrium with tax menus* as given by a tax menu $\mathcal{T}$ and an allocation

$$\{c(\theta, p), y(\theta, p), T(\cdot|p)\}_{(\theta,p,T(p)) \in \Theta \times P \times \mathcal{T}}$$

such that: (1) Firms pay workers their marginal product; (2) During working life the allocation $\{c(\theta,p), y(\theta,p)\}_{(\theta,p)\in\Theta\times P}$ is optimal for workers of type (θ,p) given the tax schedule $T(\cdot|p)$; (3) Early in life, agents with type p find it optimal to choose the tax schedule $T(\cdot|p)$; and, (4) The government budget balances.

Theorem 10. *An allocation that solves the social planner's problem can be decentralized as a competitive equilibrium with tax menus.*

Proof. See Appendix. □

We have seen in Sections 2.2 and 2.3.2 that allocations are generally different for agents with different earnings profiles. As there is no decision taken based on the profile other than the tax schedule chosen, we conclude self-selection is extremely important in implementing optimal policy in our environment.

There are a few tools currently in use in the U.S. that involve self-selection and some proposals to change income taxes that have a self-selection component. In retirement savings, IRAs and Roth-IRAs have a differential tax treatment and individuals can choose which one to use in saving for retirement and therefore allow for self-selection. Furthermore, there are a few proposals to change the current tax system by offering an alternative flat income tax rate and eliminate all deductions. Although in our framework the optimal tax rates are not necessarily flat, this type of policy has the type of self-selection component that the theory in this paper emphasizes.

2.4 Empirical evidence on Pareto tails

In the previous sections we have shown that different Pareto tails on the labor income distribution across workers with different expectations about their future labor income implies extremely different optimal labor income tax rates. Unfortunately, there is not much data that measures jointly expectations and labor income of workers, and at the same time has a high coverage of the top of the income distribution.

Table 2.1: Labor Income Distribution - Pareto Tail Estimates

Year	1989	1992	1995	1998	2001	2004	2007	2010	2013
College graduates									
Pareto Tail	2.10	1.94	1.94	2.09	1.89	1.88	1.83	1.62	1.74
s.e.	0.12	0.08	0.08	0.08	0.07	0.06	0.06	0.05	0.05
Non-College graduates									
Pareto Tail	3.18	3.23	2.94	3.38	3.22	3.51	3.75	3.17	3.13
s.e.	0.29	0.18	0.20	0.20	0.20	0.23	0.21	0.21	0.19

Source: Author's computations using data from the Survey of Consumer Finances for each corresponding year. Standard errors (s.e.) computed using replicate weights provided in the original data and method proposed by [84] to adjust for imputation. Fraction of missing information using Rubin's formula and imputations provided by the Federal Reserve Board. It represents an adjustment on the degrees of freedom of the computed statistics because of imputed data. The estimate for the Pareto tail is computed using the standard method proposed by [52] with a threshold at the 75% percentile of the labor income distribution for the upper tail of the distribution. All statistics computed using labor income of head of household. Sample is restricted to head of households between 25 and 60 years old, as well as labor income of at least $1,000.

In this section, we explore revealed preferences of young adults toward acquiring a higher education degree as a proxy for beliefs about better labor market outcomes. Of course, this is far from a perfect measure of beliefs, and therefore the evidence here should be taken with a grain of salt with respect to the implications about belief differences. However, from a theoretical perspective, those measures are tightly connected in standard models of labor market outcomes and human capital acquisition. Therefore, it is a useful first step to understand how the top of the labor income distribution differs across educational levels.

In Table 2.1, we provide evidence that Pareto tails of the labor income distribution differ markedly across educational levels in the Survey of Consumer Finances. Indeed, we find that Pareto tails are statistically different across educational levels at all years of the Survey of Consumer Finances. Our estimates for 2013 show that the labor income distribution across workers without a higher education degree is around 3.13 (s.e. 0.19), and of the labor income distribution across college graduates of 1.74 (s.e. 0.05).

This difference in Pareto tails of the labor income distribution between different educational levels implies large differences in optimal top taxes even if taxes can be made conditional on the educational level. Indeed, if $\alpha'(\theta) < 0$, then

$$\lim_{\theta \to \infty} \left(\frac{\tau(\theta)}{1 - \tau(\theta)} \right) = \left(1 + \frac{1}{\varepsilon} \right) \frac{1}{\rho}$$

for each observable group, where ε is the labor supply elasticity in that group and ρ is the Pareto tail of the distribution of labor ability. Our estimates for the Pareto tail of the labor income distribution have important implications for the distribution of labor ability in our model. If we assume that the labor income tax is approximately linear at high income levels, then the Pareto tail of the labor income distribution at top income levels is given by $\gamma = 1 + \frac{\rho}{1+\varepsilon}$. Therefore, we have $\rho = (1 + \varepsilon)(\gamma - 1)$. The top income tax rate satisfies then

$$\lim_{\theta \to \infty} \left(\frac{\tau(\theta)}{1 - \tau(\theta)} \right) = \frac{1}{\varepsilon(\gamma - 1)}$$

where γ is the Pareto tail on the labor income distribution.

Table 2.2: Top Income Taxes for $\lim_{\theta\to\infty}\alpha(\theta)=0$

Frish Elasticity	$\varepsilon=1.0$	$\varepsilon=0.75$	$\varepsilon=0.50$	$\varepsilon=0.25$	$\varepsilon=0.10$
College graduates					
Top income tax rate	57.4%	64.2%	72.9%	84.3%	93.1%
s.e.	2.0%	1.8%	1.6%	1.1%	0.5%
Non-College graduates					
Top income tax rate	31.9%	38.5%	48.4%	65.2%	82.4%
s.e.	1.9%	2.1%	2.2%	2.0%	1.3%

Source: Author's computations using data from the Survey of Consumer Finances for each corresponding year. Standard errors (s.e.) computed using replicate weights provided in the original data and method proposed by [84] to adjust for imputation. Fraction of missing information using Rubin's formula and imputations provided by the Federal Reserve Board. It represents an adjustment on the degrees of freedom of the computed statistics because of imputed data. The estimate for the Pareto tail is computed using the standard method proposed by [52] with a threshold at the 75% percentile of the labor income distribution for the upper tail of the distribution. All statistics computed using labor income of head of household. Sample is restricted to head of households between 25 and 60 years old, as well as labor income of at least $1,000.

In Table 2.2, we compute optimal tax rates for social welfare weights that converge to zero at top ability. This is a useful benchmark as it provides an upper bound on optimal top income tax rates. As we just saw, optimal income tax rates in this case depend both on the Frisch elasticity of labor supply (ε) and on the Pareto tail of the labor income distribution (γ).

There is a wide range of estimates in the literature for the Frisch elasticity of labor supply.[9] Therefore, in Table 2.2 I provide estimates of optimal top income tax rates at different levels of the labor supply elasticity. There is no straightforward way to integrate the uncertainty in the literature estimates of the labor supply elasticity into our estimates of Pareto tails. In particular, we don't have sufficient information

[9]See [21] for a detailed review of the results in the literature.

about the degree of correlation between the two estimators. Therefore, instead of taking that uncertainty directly into account, we provide different estimates of top income tax rates across a range of labor supply elasticities.

Across all levels of labor supply elasticities, we find that the optimal top income tax rate on college graduates is higher than the optimal top income tax rate on non-college graduates. Not only that difference is statistically significant, but it is economically sizable, ranging from level differences in tax rates of around 10% and 20%. Obviously, we only considered in these estimates upper bound to optimal tax rates, therefore a less redistributive social planner would implement lower tax levels at the top. However, the precision of the Pareto tail estimates implies that tax rates would be still be statistically different across a considerable range of redistributive motives.

Finally, if the government cannot condition taxes directly on educational decisions, then the problem reduces exactly to our optimal tax formulas with heterogeneous beliefs. In that case, the government will implement an optimal top income tax rate close to 100% for non-college graduates, and a lower level of top income tax rates for college graduates.

2.5 Conclusion

This paper shows the importance of the exact source of labor earnings heterogeneity for tax theory. I find that if there is advance information so that some agents have different profiles as compared to others, then optimal top marginal income taxes are greatly affected. Further, tax schemes that allow for self-selection play an important role for this type of heterogeneity as taxation on observable earnings levels is not enough to implement optimal policy.

A natural extension of this model is the consideration of human capital investment. [14] find that top marginal tax rates can have a significant effect on human capital investment. In this paper, I find that individuals with a high earnings profile face a lower optimal top marginal income tax rate, even without the distortions in human capital accumulation. Therefore, it seems likely that human capital investment considerations might even exacerbate those results. Finally, in joint ongoing work with Michael Sockin, I study the effects of advance information and college choice in a Roy model of occupational choice.

Chapter 3

How to Design Optimal Higher Education Subsidization Policy? When Schooling Creates Opportunities

In this paper we study the optimal design of higher education subsidies and the optimal design of income tax policies when schooling opens up new opportunities in the labor market. We consider a redistributive government using labor income taxes and subsidies of educational costs in a Roy model of occupational choice with an extensive margin educational decision. We provide a simple formula for the optimal subsidy on higher education that has three key components: the redistributive motive; the extensive margin elasticity of education attainment with respect to the cost of education; and the elasticity of aggregate labor supply with respect to the cost of education. [1]

[1]This chapter is based in joint work with Michael Sockin from the University of Texas at Austin. We thank Mikhail Golosov for invaluable advice.

3.1 Introduction

Income inequality has risen in the United States in the past half century at the same time as the labor income gap between those with a higher education degree and those without one has considerably risen ([80]). The strong evidence for skill based technological change reinforces the importance of higher education attainment for the study of trends in inequality. Concurrently, the recent empirical literature highlights the importance of occupational choice in understanding trends in inequality over the past 40 years in the United States.[2] Both educational and occupational choices are extremely important for labor market outcomes, and moreover those choices are entangled together as the set of skills required in different occupations often involve different educational levels. Therefore, a government concerned with the level of inequality in labor market outcomes needs to consider the intertwined nature of those choices in its design of both labor income taxes and educational policies.

In this paper, we develop a model for the study of income inequality and the design of optimal higher education subsidies with both an occupational choice problem of workers and an educational level choice problem of young adults. We achieve that by developing a Roy model with an extensive margin educational decision. Furthermore, higher education acts by creating new opportunities for workers in occupations they could not participate without higher education. A key assumption in the model is that young adults already have considerable information about the new opportunities that would arise if they were to obtain a higher education degree. Therefore, in the model young adults have a multi-dimensional vector of skills over which they can base their educational choice. This opens up the possibility for positive selection into higher education as only young adults that have a sufficiently high prior on their returns to higher education will decide to pay the fixed cost of obtaining a degree. Furthermore, in the model we allow for considerable flexibility on labor income taxation. The

[2] [13], [42], [12], [59], [1], and [36].

government has access to a uniform lump-sum transfer, a labor tax schedule that can either be linear or non-linear and is not assumed to be optimal, and a subsidy on the cost of higher education.

Our main theoretical result is a tax formula determining the optimal level of educational subsidies that a redistributive government optimally implements. A redistributive government could use labor income taxes and educational subsidies or taxes in order to improve equity in the economy. We find the government is going to use both policy instruments, and in particular that it might tax education. The fundamental reason for this result is the positive selection that occurs into high education. Only workers with a sufficiently high return on education will obtain a degree, by taxing them the planner is able to place a tax levy on exactly those workers that are on average better off. A strongly redistributive government therefore would be willing to implement such a tax.

We find empirical evidence of considerable selection of workers with high education degrees into particular occupations and also that workers on those occupations have better labor market outcomes on average. Using data from the 2013 Survey of Consumer Finances, we show that there is a concentration of workers with high education degrees into managerial and professional occupations. Moreover, the distribution of labor income across workers in those occupations is significantly different than in other occupations. First, there are higher average and median wages across workers in managerial and professional occupations. Second, there is a much heavier upper tail on the distribution among workers in the managerial and professional occupations. Indeed, we find that the Pareto tail of the labor income distribution is 1.73 (with a standard error of 0.06) among managers and professionals whereas among other occupations the estimate of the Pareto tail of the labor income distribution is 3.17 (with a standard error of 0.18).

There is a sizable literature that studies the optimal design of human capital policies.[3] [20] in a static model and [89] in a dynamic model consider the optimal design of educational policy when there is an intensive margin decision on human capital accumulation in a single occupation model. They find that if the wage elasticity with respect to unobservable ability is increasing in human capital, then it is optimal to offer less than full deductibility of educational costs on income taxes. Moreover, if the wage elasticity with respect to ability is decreasing on the stock of human capital accumulated by the worker, then the optimal policy grants workers more than full deductibility of educational costs on income taxes. Therefore, the level of complementarity of human capital and ability is key in determining whether optimal policy fosters or precludes human capital investment.

In contrast, in our model schooling has no direct effect on wages, it acts by opening up opportunities in occupations that workers could not otherwise access. As a result, the key determinant of educational subsidies in our model is the degree to which those with these new opportunities are most likely to end up at the top of the income distribution or not. Furthermore, we show that if workers with high education are most likely to be better off, then not only would optimal policy offers less than full deductibility of educational costs on income taxes, but actually high education would face positive taxation for a sufficiently redistributive government.

We also relate to a small literature that considers optimal taxation in a Roy model. [82] study redistributive policies in a Roy model with multidimensional unobservable ability. They focus on redistribution and the design of non-linear optimal taxation policies across different sectors of the economy. In contrast, we focus on redistribution and on the intertwined choice among different occupations and the educational level of workers.

[3]In static models, we have [20], [24], [71], [56] and [55]. In dynamic models, we have [19], [6], [44], [85], [61], [65], [62] and [89].

We organize the paper as follows. Section 3.2 studies the empirical relationship among labor market outcomes, occupations and education attainment using the 2013 Survey of Consumer Finances. Section 3.3 develops the theoretical model of educational and occupational choice. Then Section 3.4 solves the optimal policy problem and provides key intuitions about optimal educational policies. Finally, Section 3.6 concludes the paper.

3.2 Education, Occupations and Inequality

In this section, we study the empirical relationship of the distribution of labor market income with educational and occupational choices. There is widespread evidence in the literature for a positive correlation between education and income, as well as for the importance of occupational fixed effects in understanding the variability of labor income in the data. In contrast, we focus our attention on the shape of the income distribution at different occupations as well as different educational levels.

We use data from the 2013 Survey of Consumer Finances in our analysis. The original dataset is composed of 6,015 observations. The Federal Reserve Board provides replication weights as well as multiple imputation data so that inference on the survey can be done using bootstrapping techniques.[4] This data is particularly useful for the study of labor income inequality because of the high quality information it has on the upper tail of the income distribution as well as the richness of demographic information available in the survey.

Since the focus of this section is the study of the distribution of labor market income, we focus attention on households with a head that is no less than 25 years old and no more than 60 years old. Furthermore, we consider only the income of the head, and we restrict attention to the working population with wage incomes higher

[4]There are 999 replication weights available as well as 5 datasets that result from the multiple imputation scheme used.

than $1,000 in 2013. Those restrictions leave us with a benchmark sample of 3,028 observations which represent 50.3% of the total weight in the original sample.

There is a high level of labor income inequality in the data.[5] Table 3.1, provides basic descriptive statistics for labor income of the head of household in our benchmark sample. First, we notice from the table that the distribution is highly asymmetric, with mean wages much higher than median wages. Second, there is a high level of income inequality in the data, with a Gini index of 48%. Third, the tail of the labor income distribution is very heavy, with a point estimate of the Pareto tail of 1.99 (using the estimator proposed by [52]). All of those empirical findings are qualitatively consistent with the findings in the literature for labor income inequality.[6]

Table 3.1: Descriptive statistics for wage income in benchmark sample

Statistic	Value	s.e.	Fraction of Missing Information
Mean	$84,422	$1,889	7%
Median	$56,002	$909	33%
Percentile 25%	$32,465	$507	0%
Percentile 75%	$97,800	$1,639	15%
Percentile 95%	$211,631	$8,270	24%
Pareto Tail	1.99	0.07	5%
Gini Index	48.08	0.93	4%
Observations	3,028		

Source: Author's computations using 2013 Survey of Consumer Finances. Standard errors (s.e.) computed using replicate weights provided in the original data and method proposed by [84] to adjust for imputation. Fraction of missing information using Rubin's formula and imputations provided by the Federal Reserve Board. It represents an adjustment on the degrees of freedom of the computed statistics because of imputed data. The estimate for the Pareto tail is computed using the standard method proposed by [52] with a threshold of $100,000 for the upper tail. All statistics computed using labor income of head of household. Sample is restricted to head of households between 25 and 60 years old, as well as labor income of at least $1,000.

Figure 3.1a displays the estimated kernel density of the distribution of labor income (in levels). Figure 3.1b displays the relationship between the empirical percentiles of the distribution of labor income, with the logarithm of labor income. For

[5] In this Section we use raw data on labor market income in all analyses. In Appendix C.1 we show that our findings hold when we add a first stage to control for demographics such as age, sex and race of the head of household.

[6] See [80].

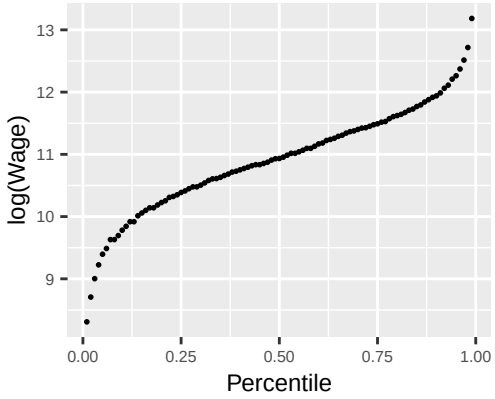
a Pareto distribution, this plot would show a straight line. Therefore, we see that the distribution is well approximated by a Pareto distribution in between percentiles 25% and 75%. Above percentile 75%, the distribution seems to have a tail that is even heavier than a Pareto tail (convex region).

Figure 3.1: Distribution of labor income in benchmark sample

(a) Kernel Density Estimate



(b) Log-quantiles of labor income distribution



Source: Author's computations using 2013 Survey of Consumer Finances. All statistics computed using labor income of head of household. Sample is restricted to head of households between 25 and 60 years old, as well as labor income of at least $1,000. For the kernel density estimate we used a bandwidth of $10,000.

Moreover, there is considerable heterogeneity in the labor income distribution across broad occupation categories and educational levels. Table 3.2 provides descriptive statistics about the distribution of labor income within two broad occupations and two broad educational categories. First, there is a strong effect of occupation and of education on average labor income. Second, we find significantly heavier upper tails for College graduates and for workers in the "Managerial/Professional" occupational category. This last point can be seen from the significantly lower Pareto Tail parameter estimates for those groups of workers. Finally, in Figure 3.2 we provide kernel density plots for the different occupations and for educational levels. We confirm in the plots that the labor income distribution has a much heavier tail for the "Man-

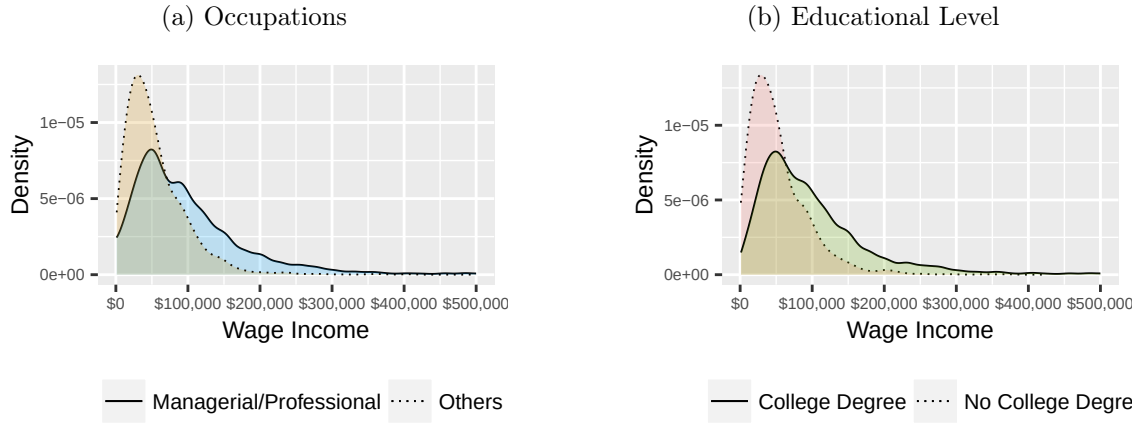
agerial/Professional" occupational category and also for workers that have a college degree.

Table 3.2: Descriptive statistics of conditional labor income distribution

		Wage Income			
Category	Freq.	Mean	s.e.	Pareto Tail	s.e.
Benchmark	100%	$84,422	$1,889	1.99	0.07
Occupation					
Managerial/Professional	43.3%	$121,150	$4,037	1.73	0.06
Others	56.7%	$56,436	$1,041	3.17	0.18
Education					
College	45.6%	$120,081	$3,699	1.78	0.06
No Finished College	54.4%	$54,573	$902	3.08	0.20

Source: Author's computations using 2013 Survey of Consumer Finances. Standard errors (s.e.) computed using replicate weights provided in the original data and method proposed by [84] to adjust for imputation. The estimate for the Pareto tail is computed using the standard method proposed by [52] with a threshold of $100,000 for the upper tail. All statistics computed using labor income of head of household. Sample is restricted to head of households between 25 and 60 years old, as well as labor income of at least $1,000.

Figure 3.2: Income Distribution by Educational Level



Source: Author's computations using 2013 Survey of Consumer Finances. All statistics computed using labor income of head of household. Sample is restricted to head of households between 25 and 60 years old, as well as labor income of at least $1,000. For the kernel density estimate we used a bandwidth of $10,000.

Table 3.3: Educational Level by Occupational Category

Occupation	Education			
	No Finished College	s.e.	College	s.e.
Managerial/Professional	27.1%	0.01	72.9%	0.01
Others	75.3%	0.01	24.7%	0.01

Source: Author's computations using 2013 Survey of Consumer Finances. All statistics computed using labor income of head of household. Sample is restricted to head of households between 25 and 60 years old, as well as labor income of at least $1,000.

There is a much higher concentration of college graduates in the "Managerial/Professional" occupation than in other occupations. In Table 3.3 we see that more about 73% of workers in the "Managerial/Professional" occupation category are college graduates. In contrast, only around 25% of workers in other occupations are college graduates in our benchmark sample. Therefore, not only there is an important difference on the labor income distribution across occupations and across educational levels, but there is also a strong selection of workers with higher education degrees in a particular set of occupations.

3.3 Model with Occupational Choice and Educational Level Choice

In this section, we propose a normative model for the study of high education subsidy policy and income taxation. In modeling this issue, we perceive that it is essential for the model to be consistent with three key ingredients that are present in the empirical evidence: first, there is high inequality in labor market outcomes; second, there is heterogeneity in the educational level of workers; third, there is a higher concentration of workers with high education degrees in particular occupations. Therefore, in the analysis of optimal policy we assume a simple framework where those conditions hold: a Roy model with a discrete choice educational decision and taxation.

Agents have preferences over consumption and labor effort ordered by a utility index $u(c, \ell)$, where c denotes consumption and $\ell \geq 0$ denotes work effort. Moreover, we assume that the utility index u is twice continuously differentiable with $u_c > 0$, $u_\ell < 0$, $u_{cc} \leq 0$, $u_{\ell\ell} < 0$, $u_{c\ell} < 0$, $\lim_{c \to 0^+} u_c(c, \ell) = +\infty$ for all $\ell \geq 0$, and $\lim_{\ell \to 0^+} u_\ell(c, \ell) = 0$ for all $c \geq 0$.

Following [83], there is an occupational choice in the model. We assume there is a set of occupations $\mathcal{O} = \{a, b\}$ in which agents can choose to work. Each individual draws a vector $\theta = (\theta_o)_{o \in \mathcal{O}}$ of labor productivity across occupations from a distribution with cumulative distribution function F with an associated probability density function denoted by f. An agent with labor productivity vector (θ_a, θ_b) that exerts effort ℓ in occupation a produces output of $y = \theta_a \ell$.

Furthermore, there is a discrete choice educational decision in the model. Agents can go to college for a fixed consumption cost $e > 0$ and learn skills that are necessary for working in occupation b. An agent with a college degree and labor productivity vector (θ_a, θ_b) that exerts effort ℓ in occupation b produces output of $y = \theta_b \ell$. Therefore, agents without a college degree can only "choose" to work in occupation a, and a college degree opens up a new opportunity for agents to choose to work in occupation b.

3.3.1 Laissez-faire Economy

Before we analyze the government problem in setting policy, it is very useful to understand outcomes in a benchmark without government policy (laissez-faire). To understand the objects of the model well, it is extremely useful to decompose the worker's problem into two: first the education and occupational choices; and, second, the labor supply conditional on the occupational choice and education. Conditional

on the choice of working on occupation a, agents solve

$$\max_{y \geq 0} u\left(y, \frac{y}{\theta_a}\right)$$

where we used that output is given by $y = \theta_a \ell$, and that agents would consume all output they produce. Let's denote this indirect utility by $U^a(\theta_a) = \max_{y \geq 0} u\left(y, \frac{y}{\theta_a}\right)$, and let's denote by $y^a(\theta_a)$ the optimal choice of output of the agent. It is straightforward to show that both U^a and y^a are strictly increasing. Analogously, we define the payoff of agents that obtain a high education degree and work in occupation b as

$$U^b(\theta_b) = \max_{y \geq 0} u\left(y - e, \frac{y}{\theta_b}\right)$$

As standard in models with fixed costs, the educational choice follows a simple threshold rule. Furthermore, the decision to attend a high education program depends on the comparative advantage of agents. Indeed, agents will choose to obtain a college degree if and only if

$$U^b(\theta_b) \geq U^a(\theta_a)$$

Since U^b is strictly increasing, then the agent's educational level choice will follow a threshold rule. Next Lemma summarizes this result.

Lemma 11. *In the laissez-faire economy, there is a function $\theta_b^*(\theta_a)$ defined by*

$$U^b(\theta_b^*(\theta_a)) = U^a(\theta_a)$$

such that a college degree is obtained if an only if

$$\theta_b \geq \theta_b^*(\theta_a)$$

for an agent with type (θ_a, θ_b).

3.3.2 Labor Income Taxation and Educational Level Choice

In this section we consider the effect of labor income taxation on educational level choices of agents in our model. For simplicity, we approximate the income tax system with only two components, a lump sum transfer T that is awarded to all agents and a marginal tax rate τ that is uniform across all labor income levels. Further, the government subsidizes educational costs at a rate s.

We can definite analogous objects as we defined in the laissez-faire economy. For instance, the indirect utility of agents conditional on working in occupation a is given by

$$U^a(\theta_a; \tau, T) = \max_{y \geq 0} \left\{ u\left(T + (1-\tau)\, y, \frac{y}{\theta_a} \right) \right\}$$

and for an agent that pursue a high education degree and works in occupation b, the indirect utility is given by

$$U^b(\theta_b; \tau, T, s) = \max_{y \geq 0} \left\{ u\left(T + (1-\tau)\, y - (1-s)\, e, \frac{y}{\theta_b} \right) \right\}$$

Agents will decide to obtain a higher education degree if and only if

$$U^b(\theta_b; \tau, T, s) \geq U^a(\theta_a; \tau, T)$$

for an agent with labor productivities (θ_a, θ_b). For ease of notation we define by $V(\theta_a, \theta_b; \tau, T, s)$ the overall value function of the agent:

$$V(\theta_a, \theta_b; \tau, T, s) = \max\left\{ U^a(\theta_a; \tau, T), U^b(\theta_b; \tau, T, s) \right\}$$

The optimal educational choice of the agent then follows a simple threshold strategy as in the previous case.

Proposition 12. *There is a function* $\theta_b^*(\theta_a;\tau,T,s)$ *defined by*

$$U^*\left(\theta_b^*\left(\theta_a;\tau,T,s\right);\tau,T\right)-(1-s)\,e=U^*\left(\theta_a;\tau,T\right)$$

such that agents will pursue a higher education degree if and only if

$$\theta_b\geq\theta_b^*\left(\theta_a;\tau,T,s\right)$$

for any pair of labor productivities (θ_a,θ_b)*.*

3.4 Optimal Policy

In this Section we provide an optimal formula for the computation of high education subsidies. The optimal level of the subsidy depends on a few characteristics of the economy and of the tax system. There are two major considerations on the design of high education policies that show in the optimal subsidy formula. First, there is an effect of high education policies on the revenue from labor income taxes and the higher this effect, the higher the level of optimal subsidies on high education. Second, there is a redistributive effect as workers with a high level degree are more likely to end up at the top of the labor income distribution. This second effect go against subsidizing high education. The optimal level of the high education subsidy will depend on how those two effects trade-off with one another.

The government is benevolent in our model. It chooses a tax system (τ,T,s) to maximize welfare ordered by

$$\mathcal{W}(\tau,T,s)=\int_{\theta_a}\int_{\theta_b}G\left(V\left(\theta_a,\theta_b;\tau,T,s\right)\right)dF\left(\theta_a,\theta_b\right)$$

where G measures the degree of distaste for inequality by the government. We assume G is concave and everywhere differentiable. The government's budget constraint is

$$\tau Y(\tau, T, s) - T - s \cdot e \cdot E(\tau, T, s) \geq 0$$

where $Y(\tau, T, s)$ is the level of aggregate output from labor in this economy, and $E(\tau, T, s)$ denotes the fraction of the population with a high education degree.

As we will see in Theorem 13, optimal education subsidies depend on several objects related to social preferences of the government and the behavioral elasticities of workers in this economy. Let's first introduce each of those important objects. For ease of notation we will omit the dependence of those objects on government policy, however it should be clear to the reader that those objects are in general endogenous and therefore depend on government choices. We denote by Ω the marginal welfare effect of the lump-sum transfer T

$$\Omega = \frac{\partial \mathcal{W}(\tau, T, s)}{\partial T}$$

and we denote by Ω_e the average marginal welfare effect of a transfer only from those with a high education degree

$$\Omega_e = \frac{1}{E(\tau, T, s)} \int_0^\infty \int_{\theta_b^*(\theta_a;\tau,s)}^\infty G'\left(U^b(\theta_b; \tau, T, s)\right) \frac{\partial U^b(\theta_b; \tau, T, s)}{\partial T} f(\theta_a, \theta_b)\, d\theta_b d\theta_a$$

We denote by ξ_z^X the elasticity of a variable X with respect to variable z. Furthermore, we denote by M the total income available for consumption by workers:

$$M = (1-\tau) Y(\tau, T, s) + T - (1-s) eE(\tau, T, s)$$

and by $R = \frac{eE(\tau,T,s)}{Y(\tau,T,s)}$ the fraction of educational expenses out of aggregate output.

Theorem 13. *At the solution to the government's problem we have*

$$\frac{s}{1-s} = \frac{1}{\xi^E_{1-s}}\left[1 - \left(\frac{\Omega_e}{\Omega}\right)\omega\right] + \frac{1}{R}\left(\frac{\tau}{1-s}\right)\frac{\xi^Y_{1-s}}{\xi^E_{1-s}} \tag{3.1}$$

where

$$\omega = 1 - \frac{\tau Y(\tau, T, s)}{M}\xi^Y_M + \frac{s \cdot e \cdot E(\tau, T, s)}{M}\xi^E_M$$

To gain intuition about optimal high education subsidy formula (3.1), it is useful to decompose the effect of a marginal change in the subsidy rate into the three components proposed by [86]: the welfare effect, the mechanical effect and the behavioral effect. A marginal increase in the subsidy rate is an income transfer for those workers that would pursue a high education degree even without the policy change. The marginal social benefit of that transfer is measured by Ω_e, the transfer is proportional to the cost of education e, and there are a fraction of $E(\tau, T, s)$ of those workers, so that the total welfare effect on those workers is $e \cdot \Omega_e \cdot E(\tau, T, s)$. The effect of the marginal policy change on the welfare of those workers that would not pursue a high education degree is null. In fact, only those workers that are just indifferent between attending school or not would change their choices, and since they are indifferent, a marginal change in subsidies has a second order effect on their payoffs. Therefore the welfare effect is given by

$$e \cdot \Omega_e \cdot E(\tau, T, s)$$

The mechanical effect of a marginal increase in the subsidy rate is $e \cdot E(\tau, T, s)\lambda$ where λ is the social cost of public resources in this economy.

There are two types of behavioral effects generated by the policy change. First, workers change their decisions to pursue or not a high education degree. The effect on the government budget of that change in decision is given by

$$s \cdot e \cdot \frac{\partial E(\tau, T, s)}{\partial s}\lambda$$

Second, workers might also change their labor supply decisions with the change in the subsidy rate. This later change might happen as a result of an income effect, as the subsidy acts as a transfer for those that would pursue a degree anyway, but it might also arise because of occupational switching and the fact workers have generally different labor abilities in different occupations. The revenue from income taxes resulting of the marginal change in educational subsidies is given by

$$\tau \cdot \frac{\partial Y\left(\tau, T, s\right)}{\partial s} \lambda$$

At the optimum, the benefits and costs of the change in the subsidy rate should be equalized so that

$$e \cdot \Omega_e \cdot E\left(\tau, T, s\right) = \lambda \left[e \cdot E\left(\tau, T, s\right) + s \cdot e \cdot \frac{\partial E\left(\tau, T, s\right)}{\partial s} - \tau \cdot \frac{\partial Y\left(\tau, T, s\right)}{\partial s} \right]$$

We obtain (3.1) from rearranging this equation and using the fact that $\lambda = \Omega/\omega$.

There are four important components that determine the optimal level of educational subsidies in equation (3.1). First, the direct response of education attainment to the policy change, which is measured by the elasticity ξ^E_{1-s}. The higher this elasticity, the lower is the need for a subsidy rate far from zero (or a tax rate far from zero). Second, the optimal level of subsidies depend on the relative welfare weight on those with a high level education degree. The term Ω_e/Ω measures exactly this relative welfare weight. Third, the level of optimal high level education subsidies depends on how the cost of public funds is affected by income effects. The term ω measures exactly this effect, if preferences are such that there are no income effects, then $\omega = 1$. As income effects become more important, the cost of public funds will be affected by the policy change as it becomes easier or harder to raise revenue using lump-sum taxation. Finally, the fourth component is the relative effect of behavioral changes to the expenses of the government on high level education subsidies, and on

the revenue of the government on labor income taxes. The term

$$\frac{1}{R}\left(\frac{\tau}{1-s}\right)\frac{\xi_{1-s}^{Y}}{\xi_{1-s}^{E}}$$

measures this relative effects and contributes to the level of the optimal subsidy on high level education.

We now turn into the optimal level of marginal income taxes and how that level is affected by the educational choice and subsidies on education in general. In the next result we obtain a tax formula for the marginal labor income tax rate.

Theorem 14. *At the solution to the government's problem we have*

$$\frac{\tau}{1-\tau}=\frac{1}{\xi_{1-\tau}^{Y}}\left(1-\frac{\Omega_{\tau}^{Y}}{Y\left(\tau,T,s\right)\Omega}\omega+\frac{s}{1-\tau}R\xi_{1-\tau}^{E}\right)$$

and Ω_{τ}^{Y} *is defined as follows*

$$\Omega_{\tau}^{Y}=\int_{0}^{\infty}\int_{0}^{\infty}G'\left(V\left(\theta_{a},\theta_{b};\tau,T,s\right)\right)y\left(\theta_{a},\theta_{b};\tau,T,s\right)f\left(\theta_{a},\theta_{b}\right)d\theta_{b}d\theta_{a}$$

The novel aspect of this tax formula is the term

$$\frac{s}{1-\tau}R\xi_{1-\tau}^{E}$$

Since $\xi_{1-\tau}^{E}\left(\tau,s\right)>0$, we see that education subsidies are in general associated with higher marginal income tax rates. On the other hand, educational taxes are associated with lower marginal income tax rates. Furthermore, we see that the current level of educational expenditures as a share of aggregate output play an important role as well. Indeed, any change in educational subsidies will affect not only workers that are just indifferent between going to college or not, but it will have an important effect on the problem of all infra-marginal workers.

This component arises because of the effect on the educational policy budget of a change in income taxes, due to the behavioral effect of marginal income taxes on educational choices of agents. As marginal income taxes rise, less agents find it in their best interest to obtain a higher education degree, as a result, if higher education is subsidized, the government needs less resources to fund educational policy. On the other hand, if agents face a tax on education, then an increase in the marginal income tax leads to less revenue for the government on such a tax due to the behavioral change in education.

3.4.1 Extension: Non-Linear Labor Income Taxation and Education Subsidies

In the previous section, we provide optimal tax formulas for the subsidy on high level education for a linear structure on labor income taxes. However, actual labor income taxes have a non-linear structure, with different tax brackets and deductions that make the actual tax schedule on labor income considerably non-linear.

In this section we extend our tax formula for optimal high level education subsidies conditional on a fixed non-linear schedule of labor income taxation. We do not assume that this schedule of labor income taxes is optimal, to the contrary, we provide the formula for optimal education subsidies conditional on a given tax schedule.

We assume that individual workers face a labor income tax schedule denoted by a differentiable function t with $t(0) = 0$ so that their after tax income is

$$y + T - t(y)$$

for a labor income of $y \geq 0$. Furthermore, we assume that t is such that after tax income is strictly increasing in y. Then workers in occupation a solve

$$U^a(\theta_a) = \max_{y \geq 0} u\left(y + T - t(y), \frac{y}{\theta_a}\right)$$

and workers in occupation b after obtaining a high level education degree solve

$$U^b(\theta_b; s) = \max_{y \geq 0} u\left(y + T - t(y) - (1-s)e, \frac{y}{\theta_b}\right)$$

Moreover, let's denote by $y^a(\theta_a)$ and $y^b(\theta_b; s)$ the resulting labor supply functions. In the next result we extend our tax formula for optimal education subsidies.

Theorem 15. *Conditional on a fixed labor income tax function t, the optimal level of high level education subsidies satisfies*

$$\frac{s}{1-s} = \frac{1}{\xi^E_{1-s}}\left(1 - \frac{\Omega_e}{\lambda}\right) + \frac{\bar{\tau}\xi^{\mathcal{I}}_{1-s}}{(1-s)R\xi^E_{1-s}}$$

where $\bar{\tau}$ denotes the average marginal income tax rate that workers face, $Y(s)$ is aggregate output, and λ is the social cost of resources to the government.

The only change as compared to optimal tax formula (3.1) is the term that involves the relative effect of the subsidy change on labor income tax revenue and subsidization expenses. Now the optimal subsidy rate depends on the level of average taxes in the economy and the elasticity of labor income tax revenue with respect to the subsidy rate. Therefore, the tax formula we obtain is robust to a non-linear labor income tax schedule with minor changes.

3.5 Quantitative Example

In this section, we study the quantitative implications of our optimal educational subsidy formula. For simplicity, we consider workers have quasi-linear preferences and iso-elastic labor supply

$$u\left(c,\ell\right)=c-\frac{\ell^{1+\psi}}{1+\psi}$$

These preferences are particularly tractable as the second component of the optimal educational subsidy formula simplifies and we obtain that

$$\frac{s}{1-s}=\frac{1}{\xi_{1-s}^{E}}\left(1-\frac{\Omega_{e}}{\Omega}\right)+\left(\frac{\tau}{1-\tau}\right)\left(1+\frac{1}{\psi}\right)$$

Therefore, the optimal educational subsidy rate, depends on the Frish elasticity of labor supply ($1/\psi$), on the marginal labor income tax rate (τ), on the elasticity of the high education attainment rate with respect to the subsidy rate (ξ_{1-s}^{E}) and on social preferences (Ω_e/Ω).

In Table 3.4 we compute the level of optimal subsidy rates for alternative values of those variables and a flat marginal labor income tax rate of 25%. We see that the optimal subsidy rate varies considerably depending on social preferences, the labor supply elasticity and the elasticity of high level education attainment. Furthermore, we see that a sufficiently redistributive government might find it optimal to impose a tax on high level education. It is important to notice here that in this exercise we take the level of the marginal labor income tax as given, and therefore we are not considering the fully optimal policy, but the educational policy that would be optimal conditional on a fixed level of the marginal labor income tax rate.[7]

[7]Lump-sum taxes can be adjusted by the government in this exercise.

Table 3.4: Optimal Educational Subsidy Rates for $\tau = 0.25$

	$1/\psi = 1$		$1/\psi = 0.5$	
	$\Omega_e/\Omega = 0.5$	$\Omega_e/\Omega = 0.9$	$\Omega_e/\Omega = 0.5$	$\Omega_e/\Omega = 0.9$
$\xi^E_{1-s} = -1.0$	$s = 0.14$	$s = 0.36$	$s = 0.00$	$s = 0.29$
$\xi^E_{1-s} = -0.9$	$s = 0.10$	$s = 0.36$	$s = -0.06$	$s = 0.28$
$\xi^E_{1-s} = -0.8$	$s = 0.04$	$s = 0.35$	$s = -0.14$	$s = 0.27$

3.6 Conclusion

In this paper, we study the optimal design of educational policies jointly with labor income taxes. We obtain an optimal tax formula for the design of optimal subsidies on high education. The formula depends on measurable elasticities and on the redistributive taste of the government. As in [86], this allows for a simple computation of optimal subsidy rates given estimates of a few elasticities.

We focused on high education policy and its redistributive aspects. However, if there are also credit constraints in the model, then there might be welfare gains in subsidizing high level education even if those imply a transfer to those at top wages. It would be interesting to generalize this model to embed both types of frictions and study the optimal design of high education policies in that case.

Appendix A

Appendix to Chapter 1

A.1 Proofs for Section 1.2

It is useful to restate the problem in terms of utility levels:

$$\begin{aligned} u_t(\theta,\beta) &= u(c_t(\theta,\beta)) \quad \text{for } t=1,2 \\ v(\theta,\beta) &= v(y(\theta,\beta)) \end{aligned}$$

and then $c_t(\theta,\beta) = C(u_t(\theta,\beta))$ where $C = u^{-1}$, and $y(\theta,\beta) = Y(v(\theta,\beta))$ where $Y = v^{-1}$. Then the planner's problem becomes

$$\max_{u_1,u_2,v} \sum_{\theta,\beta} \pi(\theta,\beta)\lambda(\theta)\left[u_1(\theta,\beta) - \frac{1}{\theta}v(\theta,\beta) + \delta u_2(\theta,\beta)\right]$$

$$s.t. \text{for all } (\theta,\beta) \text{ and } (\theta',\beta')$$

$$u_1(\theta,\beta) - \frac{1}{\theta}v(\theta,\beta) + \beta\delta u_2(\theta,\beta) \geq u_1(\theta',\beta') - \frac{1}{\theta}v(\theta',\beta') + \beta\delta u_2(\theta',\beta')$$

$$\sum_{\theta,\beta} \pi(\theta,\beta)\left[Y(v(\theta,\beta)) - C(u_1(\theta,\beta)) - \frac{1}{R}C(u_2(\theta,\beta))\right] \geq 0$$

Since $u(\cdot)$ is strictly concave and $v(\cdot)$ is strictly convex, then $C(\cdot)$ is strictly convex and $Y(\cdot)$ is strictly concave, which in turn tells us that the government's problem is

a convex problem. This property is used in our numerical solution algorithm, but is not very useful for the proofs below. Also define the payoff provided to individuals by

$$U(\theta,\beta) = u_1(\theta,\beta) - \frac{1}{\theta}v(\theta,\beta) + \beta\delta u_2(\theta,\beta)$$

and the payoff to the government of the individual allocation by

$$V(\theta,\beta) = u_1(\theta,\beta) - \frac{1}{\theta}v(\theta,\beta) + \delta u_2(\theta,\beta)$$

We begin by proving a few Lemmas that are going to be useful in the proofs of the main Theorems. Lemma 16 shows that if the lowest labor earnings ability is sufficiently low, then agents with lowest earnings ability will have all their incentive constraint with respect to higher ability types strictly slack. Then Corollary A.1 builds upon the proof of Lemma 16 to show that if the highest labor earnings type is sufficiently high as compared to all other labor earnings ability types, then incentive constraints of all agents with labor ability in $\{\theta_1,\ldots,\theta_{N-1}\}$ are strictly slack with respect to the allocation of types with ability θ_N.

Lemma 16. *Assume $\lambda(\theta)$ is weakly decreasing. Given $\{\theta_2,\ldots,\theta_N\}$, there is $\bar{\theta}_1 > 0$ such that if $0 < \theta_1 < \bar{\theta}_1$, then at the solution to the government's problem we have*

$$u_1(\theta_1,\beta) - \frac{1}{\theta_1}v(\theta_1,\beta) + \beta\delta u_2(\theta_1,\beta) > u_1(\theta',\beta') - \frac{1}{\theta_1}v(\theta',\beta') + \beta\delta u_2(\theta',\beta')$$

for $\theta' \in \{\theta_2,\ldots,\theta_N\}$.

Proof. In fact, consider $\theta_1 = 0$, then for any $v_1(\theta_1,\beta) > 0$ we would have $V(\theta_1,\beta) = -\infty$. Since π has full support and λ is weakly decreasing, then $\pi(\theta_1,\beta)\lambda(\theta_1) > 0$ and this cannot be optimal for the government. Therefore we have $v(\theta_1,\beta) = 0$ at the solution to the government's problem.

Now fix $\theta' > 0$ and $\beta' \in \{\beta_1, \ldots, \beta_M\}$. Assume by way of contradiction that $\mathrm{v}(\theta', \beta') = 0$. Since $Y'(0) = +\infty$, then this agent can produce enough resources to make all agents strictly better off in the economy, a contradiction.[1] Therefore $v(\theta', \beta') > 0$. We finally conclude that

$$u_1(\theta', \beta') - \frac{1}{\theta_1} v(\theta', \beta') + \beta\delta u_2(\theta', \beta') = -\infty$$

so that the result holds for $\theta_1 = 0$. The government's problem satisfy all conditions of the Maximum Theorem as we change θ_1, therefore the solution to the government's problem is continuous in θ_1 and there is $\bar{\theta}_1 > 0$ such that for all $0 \leq \theta_1 < \bar{\theta}_1$ the solution to the planner's problem satisfy

$$u_1(\theta_1, \beta) - \frac{1}{\theta_1} v(\theta_1, \beta) + \beta\delta u_2(\theta_1, \beta) > u_1(\theta', \beta') - \frac{1}{\theta_1} v(\theta', \beta') + \beta\delta u_2(\theta', \beta')$$

for all (θ', β') with $\theta' \in \{\theta_2, \ldots, \theta_N\}$. □

Assume $\lambda(\theta)$ is weakly decreasing. Fix θ_N, then there exists $\bar{\theta}_{N-1} < \theta_N$ such that if $0 < \theta_1 < \ldots < \theta_{N-1} \leq \bar{\theta}_{N-1}$, then at the solution to the government's problem we have

$$u_1(\theta', \beta') - \frac{1}{\theta'} v(\theta', \beta') + \beta'\delta u_2(\theta', \beta') > u_1(\theta_N, \beta) - \frac{1}{\theta'} v(\theta_N, \beta) + \beta'\delta u_2(\theta_N, \beta)$$

and $v(\theta_N, \beta) > v(\theta', \beta')$ for all $\theta' \in \{\theta_1, \ldots, \theta_{N-1}\}$ and for all $\beta, \beta' \in \{\beta_1, \ldots, \beta_M\}$.

Proof. We can use again the same strategy of the proof of the previous Lemma. If $\bar{\theta}_{N-1} = 0$, then the result holds as $v(\theta_N, \beta) > 0$ at the solution to the planner's

[1] In fact consider a perturbation to all agents $\tilde{v}(\theta, \beta) = v(\theta, \beta) + \varepsilon$ and $u_1(\theta, \beta) = u_1(\theta, \beta) + \nu$ for $\varepsilon > 0$ and $\nu > 0$. Since the original allocation is incentive compatible and the perturbation is uniform across all agents, the perturbation is incentive compatible. The change in resources used is given by $dE = \sum \pi(\theta, \beta) [C'(u_1(\theta, \beta))\nu - Y'(v(\theta, \beta))\varepsilon]$. Since $C'(u_1(\theta, \beta)) < \infty$ for all (θ, β), then for any $\varepsilon > 0$ we have $dE = -\infty$ as $Y'(v(\theta', \beta')) = +\infty$. The welfare change is given by $dW = \sum \pi(\theta, \beta) \lambda(\theta)(\nu - \varepsilon) = \nu - \varepsilon$. Therefore, for $\nu > \varepsilon$ and for ν and ε small enough, the perturbation is feasible and increases welfare, a contradiction.

problem. By continuity of the solution to the planner's problem there exists $\bar{\theta}_{N-1}$ that satisfy the conditions of the theorem. □

Lemma 17. *Assume $\lambda(\theta)$ is weakly decreasing. Fix θ_N and $\{\beta_2, \ldots, \beta_M\}$, then there exists $\bar{\theta}_{N-1} < \theta_N$ such that if $0 < \theta_1 < \ldots < \theta_{N-1} \leq \bar{\theta}_{N-1}$, then at the solution to the government's problem we have either $u_1(\theta_N, \beta_1) > u_1(\theta, \beta)$ or $u_2(\theta_N, \beta_1) > u_2(\theta, \beta)$ for all $\theta \in \{\theta_1, \ldots, \theta_{N-1}\}$ and all $\beta \in B$.*

Proof. From Corollary A.1 there is $0 < \bar{\theta}'_{N-1} < \theta_N$ such that we have $v(\theta_N, \beta_1) > v(\theta, \beta)$ for $0 < \theta_1 < \ldots < \theta_{N-1} \leq \bar{\theta}'_{N-1}$ and all $\beta \in B$. Notice that for $\bar{\theta}''_{N-1} = 0$, we have either $u_1(\theta_N, \beta) > u_1(\theta, \beta)$ or $u_2(\theta_N, \beta) > u_1(\theta, \beta)$ for $\theta = 0$. By the Maximum theorem, the solution to the planner's problem is continuous. Therefore, since the solution to the planner's problem is continuous, there exists $\bar{\theta}'_{N-1} \geq \bar{\theta}_{N-1} > 0$ such that the inequality continues to hold at the solution to the planner's problem for $0 < \theta_1 < \ldots < \theta_{N-1} \leq \bar{\theta}_{N-1}$. □

Proof of Theorem 1

Let's first show part 1 of Theorem 1. From Lemma 16 we know that there exists $\bar{\theta}_1 > 0$ such that $\bar{\theta}_1 < \theta_2$ and so that

$$u_1(\theta_1, \beta) - \frac{1}{\theta_1} v(\theta_1, \beta) + \beta\delta u_2(\theta_1, \beta) > u_1(\theta', \beta') - \frac{1}{\theta_1} v(\theta', \beta') + \beta\delta u_2(\theta', \beta')$$

for all $\beta, \beta' \in \{\beta_1, \ldots, \beta_M\}$ and $\theta' \in \{\theta_2, \ldots, \theta_N\}$. Assume by way of contradiction that types $(\theta_1, \beta) \neq (\theta_1, \beta')$ face a different allocation. Then consider a perturbation

$$\tilde{u}_t(\theta_1, \beta) = \tilde{u}_t(\theta_1) = \sum_{\beta} \left(\frac{\pi(\theta_1, \beta)}{\sum_{\beta'} \pi(\theta_1, \beta')} \right) u_t(\theta_1, \beta)$$

$$\tilde{v}(\theta_1, \beta) = \tilde{v}(\theta_1) = \sum_{\beta} \left(\frac{\pi(\theta_1, \beta)}{\sum_{\beta'} \pi(\theta_1, \beta')} \right) v(\theta_1, \beta)$$

and keep all other allocations the same. Since incentive constraints are linear, they are convex and therefore this perturbation is incentive compatible. But note that since the allocation for (θ_1, β) was initially different from the allocation for (θ_1, β'), and since C is convex and Y is concave, now we have extra resources available in the economy. This is a contradiction as the planner can now improve its objective by distributing those resources uniformly across agents. Therefore, we conclude agents with type in $\{(\theta_1, \beta) \,|\beta \in \{\beta_1, \ldots, \beta_M\}\}$ are bunched.

Now we turn to part 2 of Theorem 1. Assume by way of contradiction that all agents with types in $\{(\theta_N, \beta) \,|\beta \in \{\beta_1, \ldots, \beta_M\}\}$ receive the same allocation denoted by $(u_1(\theta_N), u_2(\theta_N), v(\theta_N))$. This implies that

$$\frac{R\delta C'(u_1(\theta_N))}{C'(u_2(\theta_N))} = \kappa$$

for some fixed $\kappa > 0$. Now consider the perturbation

$$\tilde{u}_1(\theta_N) = u_1(\theta_N) + \delta\varepsilon$$
$$\tilde{u}_2(\theta_N) = u_2(\theta_N) - \varepsilon$$

that keeps the planner's objective constant. Agents with a disagreement level β of discounting will face a payoff change of

$$dU(\theta_N, \beta) = (1 - \beta)\delta\varepsilon$$

so that for $\varepsilon > 0$ this perturbation is incentive compatible. The marginal change in resources used by such a perturbation is given by

$$dE = \left(\sum_{\beta} \pi(\theta_N, \beta)\right)(\kappa - 1)\frac{1}{R}C'(u_2(\theta_N))\varepsilon$$

If $\kappa < 1$, then we reach a contradiction as the perturbation is incentive compatible and generates extra resources for the government. Thus $\kappa \geq 1$.

Now if $\kappa > 1$, consider a perturbation only to the allocation offered to an agent with type (θ_N, β_M) where $\beta_M = 1$:

$$\tilde{u}_1(\theta_N, \beta_M) = u_1(\theta_N) - \delta\varepsilon$$
$$\tilde{u}_2(\theta_N, \beta_M) = u_2(\theta_N) + \varepsilon$$

This perturbation keeps the planner objective constant and is incentive compatible as agents with type $\beta < 1$ find it worsens the allocation for $\varepsilon > 0$. The change in resources used by the planner is given by

$$dE = \frac{1}{R} C'(u_2(\theta_N))\, \pi(\theta_N, \beta_M)\, (1 - \kappa)\, \varepsilon$$

Hence for $\varepsilon > 0$ we have $dE < 0$ as $\kappa > 1$, a contradiction as the government generates extra resource while keeping the objective constant. Therefore we conclude that $\kappa = 1$.

From Lemma 17 we have that either $u_1(\theta_N) > u_1(\theta, \beta)$ or $u_2(\theta_N) > u_2(\theta, \beta)$ for all $\theta \in \{\theta_1, \ldots, \theta_{N-1}\}$ and all $\beta \in \{\beta_1, \ldots, \beta_M\}$. We prove the Theorem in the first case, the second case is analogous. Consider the following perturbation

$$\tilde{u}_1(\theta_N, \beta_1) = u_1(\theta_N) + \beta_1 \delta\varepsilon + \nu$$
$$\tilde{u}_2(\theta_N, \beta_1) = u_2(\theta_N) - \varepsilon$$

for $\varepsilon > 0$ and $\nu > 0$. For $(\theta, \beta) \neq (\theta_N, \beta_1)$ we set

$$\tilde{u}_1(\theta, \beta) = \tilde{u}_1(\theta, \beta) + \nu$$

and keep all other allocations constant. It is easy to check this is incentive compatible. The change in resources used by the government is given by

$$dE = \pi\left(\theta_N, \beta_1\right)\left(\beta_1 - 1\right)\frac{1}{R}C'\left(u_2\left(\theta_N\right)\right)\varepsilon + \nu\sum_{\theta,\beta}\pi\left(\theta,\beta\right)C'\left(u_1\left(\theta,\beta\right)\right)$$

If we set $dE = 0$ we obtain

$$\nu = \pi\left(\theta_N, \beta_1\right)\left(1 - \beta_1\right)\delta\frac{C'\left(u_1\left(\theta_N\right)\right)}{\sum_{\theta,\beta}\pi\left(\theta,\beta\right)C'\left(u_1\left(\theta,\beta\right)\right)}\varepsilon$$

The government objective change is given by

$$dW = \pi\left(\theta_N, \beta_1\right)\left(1 - \beta_1\right)\delta\left[\frac{C'\left(u_1\left(\theta_N\right)\right)}{\sum_{\theta,\beta}\pi\left(\theta,\beta\right)C'\left(u_1\left(\theta,\beta\right)\right)} - \lambda\left(\theta_N\right)\right]\varepsilon$$

Since $\lambda\left(\theta_N\right) \leq 1$ and $u_1\left(\theta_N\right) > u_1\left(\theta,\beta\right)$, we then have $dW > 0$, a contradiction.

Proof of Theorem 2

First let's show the results for low earnings ability. By Lemma 16, given $\{\theta_2, \ldots, \theta_N\}$, there exists $\bar{\theta}_1 > 0$ such that for $\theta_1 < \bar{\theta}_1$ incentive constraints of type (θ_1, β) are strictly slack with respect to agents' (θ', β') allocations for $\theta' > \theta_1$. From Theorem 1, agents with type (θ_1, β) receive the same allocation, independent of their β-type. Then we have

$$\frac{R\delta C'\left(u_1\left(\theta_1\right)\right)}{C'\left(u_2\left(\theta_1\right)\right)} = \kappa$$

for some $\kappa > 0$. We want to show that $\kappa < 1$.[2] Assume by way of contradiction that $\kappa > 1$. Then consider the following perturbation in θ_1's allocation

$$\tilde{u_1}(\theta_1) = u_1(\theta_1) - \delta\varepsilon$$
$$\tilde{u_2}(\theta_1) = u_2(\theta_1) + \varepsilon$$

This keeps the government objective constant and for $\varepsilon > 0$ sufficiently small is incentive compatible because incentive constraints of types θ_1 are slack and agents with $\beta \leq 1$ find it (weakly) worsens the allocation. The marginal change in resources used with this perturbation is

$$\begin{aligned} dE &= \left(\sum_{\beta} \pi(\theta_1, \beta)\right) \left[\frac{1}{R} C'(u_2(\theta_1)) - \delta C'(u_1(\theta_1))\right] \varepsilon \\ &= \left(\sum_{\beta} \pi(\theta_1, \beta)\right) (1 - \kappa) \frac{1}{R} C'(u_2(\theta_1)) \varepsilon \end{aligned}$$

As $\kappa > 1$ we have for $\varepsilon > 0$ that $dE < 0$, so that the perturbation keeps the government objective constant, it is incentive compatible and it generates extra resources to the government. This is a contradiction as those resources can then be used to improve the government objective. Therefore we proved part 1.(b) of the theorem, part 1.(a) is then a consequence of $\beta < 1$.

Now let's consider the results for high earnings ability agents. First let's show that agent's (θ_N, β_M) where $\beta_M = 1$ face weakly negative intertemporal wedge. Assume by way of contradiction that

$$\frac{R\delta C'(u_1(\theta_N, \beta_M))}{C'(u_2(\theta_N, \beta_M))} > 1$$

[2]Remember that $C'(\bar{u}) = \frac{1}{u'(C(\bar{u}))}$.

Then consider the following perturbation to the allocation of (θ_N, β_M)

$$\tilde{u}_1(\theta_N, \beta_M) = u_1(\theta_N, \beta_M) - \delta\varepsilon$$
$$\tilde{u}_2(\theta_N, \beta_M) = u_2(\theta_N, \beta_M) + \varepsilon$$

and leave all other allocations constant. This perturbation is incentive compatible as agents with $\beta_M = 1$ are indifferent between the original allocation and this one, and agents with $\beta < 1$ like the original allocation for type (θ_N, β_M) better. But this perturbation generates extra resources at the margin as we have

$$dE = \pi(\theta_N, \beta_M)\left(1 - \frac{R\delta C'(u_1(\theta_N, \beta_M))}{C'(u_2(\theta_N, \beta_M))}\right)\varepsilon$$

and since for $\varepsilon > 0$ we have $dE < 0$. This is a contradiction as the extra resources can then be used to improve the planner's objective. Therefore we just proved 2.(a). Notice that the proof of 2.(b) is within the proof of Theorem 1.

A.2 Extension: life-cycle model with hyperbolic preference shocks

In this section, we present a life-cycle model with stochastic earnings ability and self-control shocks and characterize the efficient dynamic provision of commitment in the model. We show that a trade-off between providing insurance and providing commitment arises when agents face high income shocks but does not arise for agents with low income shocks. As a result commitment is provided for those that face low income shocks, but not so much for those that face high income shocks. This is the exact counterpart of the results obtained in our two-period economy with redistribution.

Our assumption of hyperbolic preferences shocks, instead of stable hyperbolic preferences, allows for a considerable level of tractability as well as changes in behavior over the life-cycle. We effectively sidestep the intricacies studied in detail by [37] where out of equilibrium allocations play an important role in the optimal design of commitment devices. In addition, this setup allows for changes in the cross sectional distribution of present bias through the life cycle. Hence it allows, for example, for more self-control when individuals are close to the retirement age. Finally, this setup allows for an agent to have both time consistent and time inconsistent behavior during its life-cycle.[3]

The economy is composed of a measure one of agents that have a life-cycle of $T \geq 3$ periods that is composed of working life and retirement.[4] At each period $t = 1, \ldots, T_w$ each agent face an earnings ability shock $\theta_t \in \Theta = \{\theta_1, \ldots, \theta_N\}$ with a transition probability distribution denoted by $\rho_{t+1}(\theta_{t+1}|\theta_t)$ which we allow to vary over the life-cycle and has full support at all periods over Θ for all $\theta_t \in \Theta$. We assume as well that ρ_{t+1} has a stochastic ordering so that higher levels of θ_t imply a distribution that first order stochastically dominates a distribution for lower levels of θ_t. With an abuse of notation we denote the probability distribution over the initial level of earnings ability θ_1 by $\rho_1(\theta_1)$ and assume that it has full support. The period payoff during working life over consumption and obtained earnings is given by

$$U_W(c_t, y_t; \theta_t) = u(c_t) - \frac{1}{\theta_t} v(y_t)$$

where we assume $u' > 0$, $u'' < 0$, $v' > 0$, $v'(0) = 0$, $v'' > 0$ and $v(0) = 0$. At each period $t = T_w + 1, \ldots, T$ the agent is retired so that it only consumes and it's period

[3] In the traditional example of present bias, the postponement of going to the gym, our setup allows for agents that would go to the gym once in a while without a commitment device in place. This is not allowed by stable hyperbolic preferences as in that setup agents would decide to never go to the gym without help of a commitment device.

[4] Here we implicitly assume that retirement lasts for at least one period.

payoff is given by

$$U_R(c_t) = u(c_t)$$

Below, for ease of notation we write $U_t(c_t, y_t; \theta_t)$ for the period payoff but one should keep in mind that we refer to the two possibilities above.[5] Finally, without loss of generality we order earnings ability shocks by $\theta_1 < \ldots < \theta_N$ so that θ_1 is the earnings ability type of an agent that finds it extremely hard to obtain labor earnings whereas θ_N is the earnings ability type of an agent that finds it relatively easy to obtain labor earnings.

Agents in this economy face self-control shocks during their life-cycle. At each period $t = 1, \ldots, T-1$ each agent faces a hyperbolic self-control shock $\beta_t \in B = \{\beta_1, \ldots, \beta_M\}$ which is assumed to be independently distributed both over time and from earnings ability shocks. Without loss of generality we order $\beta_1 < \beta_2 < \ldots < \beta_M$. We allow the probability distribution of self-control shocks at period t,$\gamma_t(\beta_t)$, to vary over the life-cycle but assume it has full support at all periods. We denote the joint type by $h_t = (\beta_t, \theta_t)$ and its distribution at time t by π_t. We follow the usual notation in the literature using a superscript t for the history of types realized until period t, so that $h^t = (h_1, \ldots, h_t) \in H^t$ and with an abuse of notation also denote by π_t the probability distribution over H^t.

Types are unobservable and we rely on the revelation principle to characterize implementable allocations. In Appendix A.3 we show that it is sufficient to consider mechanisms in which at each period the agent report its current period type, and not the whole history of types up to that point.This result is an implication of the assumption that hyperbolic preference shocks are independent over time. An allocation can therefore be written as a sequence of pair of functions $(c_t, y_t) : H^t \to \mathbb{R}^2_+$.

[5]One might as well think that at retirement $\theta_t = 0$ and therefore $y_t = 0$ and use only the first period payoff for reference.

The planner evaluates welfare of an allocation (c, y) according to the period 0 preference of agents in the economy

$$W(c,y) = \sum_{t=1}^{T} \delta^{t-1} \sum_{h^t} \pi_t\left(h^t\right) U_t\left(c_t\left(h^t\right), y_t\left(h^t\right), h_t\right)$$

We therefore interpret this as the problem of an agent at period 0 obtaining the optimal level of insurance over earnings ability shocks and a commitment device to deal with self-control shocks over its life-cycle. The efficient allocation could therefore be implemented both by the government or by competitive private insurance companies as long as both are able to enforce the contract.

Finally there is a perfect credit market that the planner has access to at a gross rate of return R per period. We say a contract is feasible for the planner if

$$\sum_{t=1}^{T} \frac{1}{R^{t-1}} \sum_{h^t} \pi\left(h^t\right)\left[c_t\left(h^t\right) - y_t\left(h^t\right)\right] \leq 0$$

that is to say that assuming a law of large numbers hold, if all agents in the economy take this contract then the insurance provider is able to fulfill the contract without outside resources. An allocation (c, y) is said to be implementable if it is feasible and incentive compatible. The planner's problem then is

$$\max_{c,y} W(c,y)$$
$$s.t.\ (c,y) \text{ is implementable}$$

We now characterize properties of the planner's problem solution, i.e., characteristics of efficient insurance provision and efficient commitment provision. Full-insurance of self-control shocks in a period t implies that agents with lack of self-control ($\beta_t < 1$) and agents with self control ($\beta_t = 1$) are assigned the same allocation conditional on the whole history of earnings ability shocks they reported. Therefore, it's natural

to interpret insurance of self-control shocks as the provision of commitment by the planner. Next we show that full commitment is provided under some conditions in our economy.

Theorem 18. *Given θ_N and $\{\beta_2, \ldots, \beta_M\}$, then*

1. *There is $\bar{\theta}_1 > 0$ such that if $\theta_1 \leq \bar{\theta}_1$ then at the solution to the planner's problem we have that for any fixed $t = 1, \ldots, T-1$ and a fixed history h^{t-1} agents with types in $\{(h^{t-1}, (\theta_1, \beta)) : \beta \in B\}$ are all assigned the same level of consumption and earnings in period t and are assigned the same continuation allocation for all future periods:*

$$\begin{aligned}
c_t\left(h^{t-1}, (\theta_1, \beta)\right) &= c_t\left(h^{t-1}, (\theta_1, \beta')\right) \\
y_t\left(h^{t-1}, (\theta_1, \beta)\right) &= y_t\left(h^{t-1}, (\theta_1, \beta')\right) \\
c_{t+s}\left(h^{t-1}, (\theta_1, \beta), (h_{t+1}, \ldots, h_{t+s})\right) &= c_{t+s}\left(h^{t-1}, (\theta_1, \beta'), (h_{t+1}, \ldots, h_{t+s})\right) \\
y_{t+s}\left(h^{t-1}, (\theta_1, \beta), (h_{t+1}, \ldots, h_{t+s})\right) &= y_{t+s}\left(h^{t-1}, (\theta_1, \beta'), (h_{t+1}, \ldots, h_{t+s})\right)
\end{aligned}$$

   for all $\beta, \beta' \in B$ and for all $(h_{t+1}, \ldots, h_{t+s}) \in H_{t+1} \times \cdots \times H_{t+s}$;

2. *there is $\bar{\theta}_{N-1} < \theta_N$ and $\bar{\beta}_1 > 0$ such that if $\theta_1 < \ldots < \theta_{N-1} \leq \bar{\theta}_{N-1}$ and $\beta_1 \leq \bar{\beta}_1$ then at the solution to the planner's problem we have that at all periods $t = 1, \ldots, T-1$ and after all histories h^{t-1} agents with types $\{(h^{t-1}, (\theta_N, \beta)) : \beta \in B\}$ are not all assigned the same current allocation and continuation allocations.*

The planner in our economy desires to provide both insurance against earnings ability shocks and commitment against self-control shocks. This result informs us that it is efficient to provide perfect commitment for low earnings agents in our economy but not for high earnings agents. To understand why we need to think about the interaction between the self-control problem and the insurance problem. The lack of

self-control that a share of agents in our economy faces imply a demand for flexibility from a period t perspective. These agents with lack of self-control are willing to pay to obtain an allocation that caters to their biased preferences at period t. Without an insurance problem, the planner would not be willing to sell them flexibility and would provide commitment to all agents.[6] However, there is an insurance problem against earnings ability shocks in our environment to which the planner is not able to provide full insurance. Therefore, it is possible to charge high earnings agents for flexibility and use the proceeds to improve insurance against labor earnings shocks. This transfer of resources in exchange for flexibility is made possible because high earnings individuals are in a relatively better position than lower earnings individuals to pay for flexibility. Therefore, the planner is able to offer this flexibility to high earners without losing the ability to provide commitment for the lowest earnings agents in the economy.

Our second set of results characterize the effects of the available choices on ameliorating (or not) the time inconsistency problem faced by agents. Indeed, so far the results show fundamental differences in the choice sets offered to agents but there was no discussion of how those choices would compare to the case without self-control shocks. In the case of hyperbolic agents, one natural measure for comparison of those choices is the wedge to a efficient time-consistent intertemporal consumption decision. Without self-control shocks ($\beta = 1$), efficient insurance implies that intertemporal choices satisfy the inverse Euler equation[7]

$$\sum_{\theta_{t+1} \in \Theta_{t+1}} \rho_{t+1}\left(\theta_{t+1} | \theta_t\right) \frac{u'\left(c_t\left(\theta^t\right)\right)}{\delta R u'\left(c_{t+1}\left(\theta^t, \theta_{t+1}\right)\right)} = 1$$

As long as this intertemporal condition is satisfied the detrimental effects of time-inconsistency have been completely dealt with. We can define the time inconsistency

[6] This is the case if $\theta_t = \theta_0$ for all agents in the economy at all histories.

[7] For applications of this result for optimal taxation see [38], [89] and [40].

wedge in our economy for an agent with history h^t as

$$\tau\left(h^t\right) = \sum_{h_{t+1} \in H_{t+1}} \pi_{t+1}\left(h_{t+1}|h^t\right) \frac{u'\left(c_t\left(h^t\right)\right)}{\delta R u'\left(c_{t+1}\left(h^t, h_{t+1}\right)\right)} - 1$$

On one hand, if the time inconsistency wedge is positive it means that consumption into the future is relatively high as compared to current consumption. On the other hand, a negative time inconsistency wedge implies that agents are consuming too much in the current period as compared to what they consume in the future. If agents face self-control shocks, it's natural to assume that without the availability of commitment devices agents would present a negative time consistency wedge and as a result would have relatively little consumption into the future. The next result extends Theorem 2 to this environment.

Theorem 19. *Given θ_N and $\{\beta_2, \ldots, \beta_M\}$, then*

1. *there is $\bar{\theta}_1 > 0$ such that if $\theta_1 \leq \bar{\theta}_1$ then at the solution to the planner's problem we have that at any fixed $t = 1, \ldots, T-1$ and after any fixed history h^{t-1} agents with types in $\{(h^{t-1}, (\theta_1, \beta)) : \beta \in B\}$ have a weakly positive time inconsistency wedge:*

$$\tau\left(h^{t-1}, (\theta_1, \beta)\right) = \sum_{h_{t+1} \in H_{t+1}} \pi_{t+1}\left(h_{t+1}|\theta_1\right) MRS_{t,t+1}\left(h_{t+1}|\left(\theta_1, \beta\right)\right) - 1 \geq 0$$

   where

$$MRS_{t,t+1}\left(h_{t+1}|\left(\theta_1, \beta\right)\right) = \frac{u'\left(c_t\left(h^{t-1}, (\theta_1, \beta)\right)\right)}{\delta R u'\left(c_{t+1}\left(h^{t-1}, (\theta_1, \beta), (\theta_{t+1}, \beta_{t+1})\right)\right)}$$

2. *there is $\bar{\theta}_{N-1} < \theta_N$ and $\bar{\beta}_1 > 0$ such that if $\theta_1 < \ldots < \theta_{N-1} \leq \bar{\theta}_{N-1}$ and $\beta_1 < \bar{\beta}_1$, then at the solution to the planner's problem we have for any fixed periods $t = 1, \ldots, T-1$ and after any fixed history h^{t-1} agents with types*

(a) $(h^{t-1}, (\theta_N, \beta_M))$ *have a weakly positive time-inconsistency wedge*

(b) $(h^{t-1}, (\theta_N, \beta_1))$ *have a negative time inconsistency wedge:*

$$\tau\left(h^{t-1}, (\theta_N, \beta_1)\right) < 0$$

A.3 Proofs of results for dynamic model with self-control shock

Types are unobservable and we rely on the revelation principle to characterize implementable allocations. Hence we define an allocation as a pair of functions $(c_t, y_t) : H^1 \times \cdots \times H^{t-1} \times H^t \to \mathbb{R}^2_+$ for each period t that assigns a consumption level and an earnings level for any reported history $h^t \in H^t$ at period t and any past reported history $\hat{r}^{t-1} = (h^1, \ldots, h^{t-1}) \in H^1 \times \cdots \times H^{t-1}$. A strategy for an agent is a sequence of reporting strategies $\sigma_t : H^1 \times \cdots \times H^{t-1} \times H^t \to H^t$. The overall payoff after history h^t, previous reports $\hat{r}^{t-1} = (r^1, \ldots, r^{t-1}) \in H^1 \times \cdots \times H^{t-1}$ and following a strategy $(\sigma_s)_{s=t}^T$ from period t on is given by

$$\begin{aligned} V_t\left(\hat{r}^{t-1}, h^t, (\sigma_s)_{s=t}^T\right) =& U_t\left(c_t\left(\hat{r}^t, \sigma_t\left(\hat{r}^t, h^t\right)\right), y_t\left(\hat{r}^{t-1}, \sigma_t\left(\hat{r}^{t-1}, h^t\right)\right), \theta_t\right) \\ &+ \beta_t \sum_{s=t+1}^{T} \delta^{s-t} \sum_{h^s \succ h^t} \pi_s\left(h^s | \theta_t\right) U_s\left(\hat{c}_s\left(\hat{r}^{s-1}, h^s\right), \hat{y}_s\left(\hat{r}^{s-1}, h^s\right), \theta_s\right) \end{aligned} \tag{A.1}$$

where with an abuse of notation $\hat{c}_s\left(\hat{r}^{s-1}, h^s\right) = c_s\left(\sigma_s\left(\hat{r}^{s-1}, h^s\right)\right)$ and $\hat{y}_s\left(\hat{r}^{s-1}, h^s\right) = y_s\left(\sigma_s\left(\hat{r}^{s-1}, h^s\right)\right)$. So the preference is hyperbolic with a present-bias of β_t at period t.[8]

[8] We use the symbol $h^s \succ h^t$ to denote continuation histories for $s > t$ that are consistent with h^t.

Agents are sophisticated as they take into account their time inconsistencies into the future. An allocation is said to be incentive compatible if truth-telling is a sub-game perfect equilibrium of the game played between the selves at all periods and after all histories of reports and realized types. Hence incentive compatibility requires that after any history of reports $\hat{r}^{t-1} \in H^1 \times \cdots \times H^{t-1}$ and an actually realized type h^{t-1}

$$\sigma_t^{Truth} \in \arg\max_{\sigma_t} V_t\left(\hat{r}^{t-1}, h^t, \left(\sigma_t, \left(\sigma_s^{Truth}\right)_{s=t}^{T}\right)\right)$$

that is to say: taking into account that future selves will consider it optimal to report the truth, reporting the truth at period t after history h^t is optimal for any previous reports $\hat{r}^t$. The revelation principle guarantees us that the outcome of any mechanism can be obtained using the allocations defined above.[9]

Our assumptions of full support over types, the Markovian nature of the stochastic process over types and the planner's objective allow us to further simplify incentive constraints. From the Markovian nature of the problem we have that, conditional on $\hat{r}^{t-1}$, the preferences after any history $\tilde{h}^t \in H^t$ with $h_t = \tilde{h}_t$ have the same ordering as the preferences after history h^t. As we will show below, the planner's objective function is strictly concave so we must have that at an optimal allocation the allocations from period t on of those types of agents coincide. Hence we can write

$$c_{t+s}\left(\hat{r}^{t-1}, h^t, \ldots, h^s\right) = c_{t+s}\left(\hat{r}^{t-1}, h_t, \ldots, h^s\right)$$
$$y_{t+s}\left(\hat{r}^{t-1}, h^t, \ldots, h^s\right) = y_{t+s}\left(\hat{r}^{t-1}, h_t, \ldots, h^s\right)$$

Using this argument recursively for all periods $s > t$ we obtain

$$c_{t+s}\left(\hat{r}^{t-1}, h^t, \ldots, h^s\right) = c_{t+s}\left(\hat{r}_1, \ldots, \hat{r}_{t-1}, h_t, \ldots, h_s\right)$$
$$y_{t+s}\left(\hat{r}^{t-1}, h^t, \ldots, h^s\right) = y_{t+s}\left(\hat{r}_1, \ldots, \hat{r}_{t-1}, h_t, \ldots, h_s\right)$$

[9]Indeed this is a Bayesian game with positive probabilities at all nodes of the game.

where we used that $\hat{r}_1, \ldots, \hat{r}_t$ are optimal reports for an agent with that history of types. Therefore it is without loss of generality that the mechanism requires only reporting of the current period type and not of the full history of types.[10]

The next result is the extension of Lemma 16 for the dynamic case.

Lemma 20. *Given $\{\theta_2, \ldots, \theta_N\}$, there is $\bar{\theta}_1 > 0$ such that if $\theta_1 < \bar{\theta}_1$ then at the solution to the planner's problem we have*

$$\underbrace{V_t\left(h^{t-1}, (\beta, \theta_1)\right)}_{\text{truthful report}} \geq \underbrace{V_t\left((\beta', \theta_1) \,|h^{t-1}, (\beta, \theta_1)\right)}_{\text{deviation in } \beta} > \underbrace{V_t\left((\beta', \theta') \,|h^{t-1}, (\beta_t, \theta_1)\right)}_{\text{deviation in } \theta}$$

and $y_t\left(h^{t-1}, (\beta', \theta')\right) > y_t\left(h^{t-1}, (\beta, \theta_1)\right)$ for all $\theta' > \theta_1$, for all $\beta' \neq \beta$, for all h^{t-1} and for all t.

Proof. In fact, if $\theta_1 = 0$, then $y_t\left(h^{t-1}, (\beta, \theta_1)\right) = 0$ for all $\beta \in B$ and all h^{t-1}. For any $\theta' > 0$, then $y_t\left(h^{t-1}, (\beta', \theta')\right) > 0$, therefore

$$V_t\left((\beta', \theta') \,|h^{t-1}, (\beta, \theta_1)\right) = -\infty$$

and this proves the strict inequality. The first inequality is a requirement from incentive compatibility. From continuity of the solution to the planner's problem, it follows that for fixed $\{\theta_2, \theta_3, \ldots, \theta_N\}$, there is $\bar{\theta}_1$ such that for all $\theta_1 \leq \bar{\theta}_1$ we have at the solution to the planner's problem

$$V_t\left(h^{t-1}, (\beta, \theta_1)\right) \geq V_t\left((\beta', \theta_1) \,|h^{t-1}, (\beta, \theta_1)\right) > V_t\left((\beta', \theta') \,|h^{t-1}, (\beta_t, \theta_1)\right)$$

and we proved the result. □

[10] This characterization implies that only equilibrium path allocations are important for incentive compatibility (see [35] and [60]). This argument can break down in problems with perfect correlated types. One example is in [15]. The full support in β_t types is particularly important in the current analysis for this alternative characterization to be valid. If there is no full support in β_t it is possible to design a mechanism in which off equilibrium path allocations relax incentive constraint on the equilibrium path and therefore this simplified characterization does not hold.

Fix θ_N, then there exists $\bar{\theta}_{N-1} < \theta_N$ such that if $0 < \theta_1 < \ldots < \theta_{N-1} \leq \bar{\theta}_{N-1}$, then at the solution to the government's problem we have

$$V_t\left(h^{t-1}, (\beta', \theta')\right) > V_t\left((\beta, \theta_N) | h^{t-1}, (\beta', \theta')\right)$$

and $v_t\left(h^{t-1}, (\beta, \theta_N)\right) > v_t\left(h^{t-1}, (\beta', \theta')\right)$ for all h^{t-1} , for all $\theta' \in \{\theta_1, \ldots, \theta_{N-1}\}$ and for all $\beta, \beta' \in \{\beta_1, \ldots, \beta_M\}$.

Proof. We can use again the same strategy of the proof of the previous Lemma. If $\bar{\theta}_{N-1} = 0$, then the result holds as $v_t\left(h^{t-1}, (\beta, \theta_N)\right) > 0$ at the solution to the planner's problem. By continuity of the solution to the planner's problem there exists $\bar{\theta}_{N-1}$ that satisfy the conditions of the theorem. □

Lemma 21. *Fix θ_N and $\{\beta_2, \ldots, \beta_M\}$, then there exists $\bar{\theta}_{N-1} < \theta_N$ and $\bar{\beta}_1$ such that if $0 < \theta_1 < \ldots < \theta_{N-1} \leq \bar{\theta}_{N-1}$ and $\beta_1 < \bar{\beta}_1$, then at the solution to the government's problem we have $u_t\left(h^{t-1}, (\beta_1, \theta_N)\right) > u_t\left(h^{t-1}, (\beta, \theta)\right)$ for all h^{t-1}, for all $\theta \in \{\theta_1, \ldots, \theta_{N-1}\}$, for all $\beta \in \{\beta_1, \ldots, \beta_N\}$ and for all $t = 1, \ldots, T-1$.*

Proof. From Corollary A.3 we have that $v_t\left(h^{t-1}, (\beta_1, \theta_N)\right) > v_t\left(h^{t-1}, (\beta, \theta)\right)$. Note that if $\beta_1 \approx 0$, incentive compatibility requires $u_t\left(h^{t-1}, (\beta_1, \theta_N)\right) > u_t\left(h^{t-1}, (\beta, \theta)\right)$. By the Maximum theorem, the solution to the planner's problem is continuous. Therefore there exists $\bar{\beta}_1 > 0$ such that this inequality continues to be strict for all $\beta_1 < \bar{\beta}_1$. Since $T < \infty$ we can pick a uniform level of $\bar{\beta}_1 > 0$. □

(Theorem 18 part 1) Given $\{\theta_2, \ldots, \theta_N\}$, there is $\bar{\theta}_1 > 0$ such that if $\theta_1 \leq \bar{\theta}_1$ then at the solution to the planner's problem we have that at all periods $t = 1, \ldots, T-1$ and after all histories h^{t-1} agents with types in $\{(\theta_1, \beta) : \beta \in B_t\}$ are all assigned the same level of consumption and earnings in period t and are assigned the same continuation allocation for all future periods.

Proof. Consider the problem in terms of utility levels from consumption and disutility levels from working.[11] Assume by way of contradiction that for a fixed $t < T$ and fixed history $h^{t-1} \in H^{t-1}$ and for $\beta, \beta' \in B_t$ we have

$$u_t\left(h^{t-1}, (\beta, \theta_1)\right) > u_t\left(h^{t-1}, (\beta', \theta_1)\right)$$

at the solution to the planner's problem. Consider now an allocation that is a convex combination between between $(h^{t-1}, (\beta^*, \theta_1))$ allocations for all $\beta^* \in B$ and offer it to all types with low ability θ_1 after history h^{t-1}:

$$\tilde{u}_t\left(h^{t-1}, (\beta^*, \theta_1), h_{t+1}^{t+s}\right) = \sum_{b \in B} \frac{\pi_t\left((b, \theta_1) | h^{t-1}\right)}{\sum_{b' \in B} \pi_t\left((b', \theta_1) | h^{t-1}\right)} u_t\left(h^{t-1}, (b, \theta_1), h_{t+1}^{t+s}\right)$$
$$\tilde{v}_t\left(h^{t-1}, (\beta^*, \theta_1), h_{t+1}^{t+s}\right) = \sum_{b \in B} \frac{\pi_t\left((b, \theta_1) | h^{t-1}\right)}{\sum_{b' \in B} \pi_t\left((b', \theta_1) | h^{t-1}\right)} v_t\left(h^{t-1}, (b, \theta_1), h_{t+1}^{t+s}\right)$$

for all $\beta^* \in B$, where $h_{t+1}^{t+s} = (h_{t+1}, \ldots, h_{t+s})$ We have from Lemma (20) that there exists $\bar{\theta}_1$ such that for $\theta_1 < \bar{\theta}_1$

$$V_t\left(h^{t-1}, (\beta, \theta_1)\right) \geq V_t\left((b, \theta_1) | h^{t-1}, (\beta, \theta_1)\right) > V_t\left((\beta'', \theta') | h^{t-1}, (\beta, \theta_1)\right)$$

for all $\theta' > \theta_1$, for all β''. Therefore, incentive compatibility constraints at nodes $(h^{t-1}, (b, \theta_1))$ are satisfied for all $b \in B$. From linearity of the objective function, it follows as well that incentive compatibility of types $(h^{t-1}, (b, \theta))$ are satisfied for all $\theta > \theta_1$ and all $b \in B$. Therefore this perturbation is incentive compatible at period t. Further, notice that the perturbation is such that from the planner's point of view, continuation utility at h^{t-1} is unchanged. Therefore welfare is unchanged and from

[11] That is to say, consider the standard transformation

$$u\left(c_t\left(h^t\right)\right) = u_t\left(h^t\right)$$
$$v\left(y_t\left(h^t\right)\right) = v_t\left(h^t\right)$$

and let $C(u)$ and $Y(v)$ denote the inverses of u and v, respectively.

hyperbolic preferences, all incentive constraints for period $s \leq t-1$ are satisfied. For histories $h^s \succ h^{t-1}$ for $s > t$, note once again that the convex combination does not affect incentives because the objective is linear. However, we then reach a contradiction since $C(u) = u^{-1}(u)$ is a convex function, then the total cost of the new allocation is strictly lower as $u_t(h^{t-1}, (\beta, \theta_1)) > u_t(h^{t-1}, (\beta', \theta_1))$ and π_t has full support (so these types have positive mass). □

(Theorem 18 part 2) Given θ_N and $\{\beta_2, \ldots, \beta_M\}$, there is $\bar{\theta}_{N-1} < \theta_N$ and $\bar{\beta}_1 > 0$ such that if $\theta_1 < \ldots < \theta_{N-1} \leq \bar{\theta}_{N-1}$ and $\beta_1 < \bar{\beta}_1$ then at the solution to the planner's problem we have that at all periods $t = 1, \ldots, T-1$ and after all histories h^{t-1} agents with types (θ_N, β) are not bunched all at the same allocation.

Proof. Assume by way of contradiction that for some h^{t-1} we have that $(h^{t-1}, (\beta_t, \theta_N))$ for all $\beta_t \in B_t$ face the same allocation. Then

$$\mathbb{E}_t \left[\frac{C'\left(u_{t+1}\left(h^{t-1}, (\beta_t, \theta_N), (\beta_{t+1}, \theta_{t+1})\right)\right)}{\delta R_t C'\left(u_t\left(h^{t-1}, (\beta_t, \theta_N)\right)\right)} | \theta_N \right] = \kappa$$

for some constant $\kappa > 0$. I'm going to show this cannot be optimal. Recalling that $\beta_M = 1$, consider the following perturbation for the allocation of type (β_M, θ_N):

$$\tilde{u}_t\left(h^{t-1}, (\beta_M, \theta_N)\right) = u_t\left(h^{t-1}, (\beta_M, \theta_N)\right) - \varepsilon$$
$$\tilde{u}_{t+1}\left(h^{t-1}, (\beta_m, \theta_n), (\beta_{t+1}, \theta_{t+1})\right) = u_{t+1}\left(h^{t-1}, (\beta_M, \theta_N), (\beta_{t+1}, \theta_{t+1})\right) + \frac{1}{\delta}\varepsilon$$

for $\varepsilon > 0$. The welfare of type $(h^{t-1}, (\beta_M, \theta_N))$ is kept constant by such a change. Further, types $(h^{t-1}, (\beta_j, \theta_N))$ for $\beta_j < 1$ dislike the perturbation, so it is incentive compatible. The marginal change in the amount of resources used by type $(h^{t-1}, (\beta_M, \theta_N))$

is

$$\begin{aligned} dE &= -C'\left(u_t\left(h^{t-1},(\beta_M,\theta_N)\right)\right)\varepsilon \\ &\quad + \frac{1}{R_t}\mathbb{E}_t\left[C'\left(u_{t+1}\left(h^{t-1},(\beta_M,\theta_N),(\beta_{t+1},\theta_{t+1})\right)\right)|\theta_N\right]\frac{1}{\delta}\varepsilon \\ &= (\kappa-1)\,C'\left(u_t\left(h^{t-1},(\beta_M,\theta_N)\right)\right)\varepsilon \end{aligned}$$

At the solution to the planner's problem it must be the case $dE \geq 0$. Therefore we conclude $\kappa \geq 1$. Now assume by way of contradiction that $\kappa > 1$. Since $\beta_t \leq 1$ for all $\beta_t \in B_t$, then consider the following perturbation to all types $(h^{t-1},(\beta_t,\theta_N))$:

$$\begin{aligned} \tilde{u}_t\left(h^{t-1},(\beta_t,\theta_N)\right) &= u_t\left(h^{t-1},(\beta_t,\theta_N)\right)+\varepsilon \\ \tilde{u}_{t+1}\left(h^{t-1},(\beta_t,\theta_N),(\beta_{t+1},\theta_{t+1})\right) &= u_{t+1}\left(h^{t-1},(\beta_t,\theta_N),(\beta_{t+1},\theta_{t+1})\right)-\frac{1}{\delta}\varepsilon \end{aligned}$$

Type $(h^{t-1},(\beta_M,\theta_N))$ is indifferent between the original allocation and the perturbed one, types $(h^{t-1},(\beta_t,\theta_N))$ with $\beta_t < 1$ strictly prefer the perturbed allocation for $\varepsilon > 0$. Since no incentive constraint from types $\{\theta_1,\theta_2,\ldots,\theta_{n-1}\}$ are binding upwards, then the perturbation is incentive compatible for $\varepsilon > 0$ small enough. The change in resources used by each type of agent is

$$dE = (1-\kappa)\,C'\left(u_t\left(h^{t-1},(\beta_t,\theta_N)\right)\right)\varepsilon$$

and we reach a contradiction since $\kappa > 1$. Therefore we conclude $\kappa = 1$ and if agents are bunched at θ_N, then the inverse Euler equation holds.

Now we need to show this implies a contradiction. By Lemma 17, since agent's with the high earnings shock are bunched, we have $u_t\left(h^{t-1},(\beta_t,\theta_N)\right) >$

$u_t\left(h^{t-1},(\beta_t,\theta_j)\right)$ for all $j<N$. Then consider the following perturbation

$$\begin{aligned}\tilde{u}_t\left(h^{t-1},(\beta_1,\theta_N)\right) &= u_t\left(h^{t-1},(\beta_1,\theta_N)\right)+\varepsilon-\nu\\ \tilde{u}_{t+1}\left(h^{t-1},(\beta_1,\theta_N),(\beta_{t+1},\theta_{t+1})\right) &= u_{t+1}\left(h^{t-1},(\beta_1,\theta_N),(\beta_{t+1},\theta_{t+1})\right)-\frac{1}{\delta\beta_2}\varepsilon\end{aligned}$$

From the point of view of β_1 the payoff change is

$$dU(\beta_1)=\left(1-\frac{\beta_1}{\beta_2}\right)\varepsilon-\nu$$

Since $\beta_1<\beta_2$, then we can choose $\varepsilon>0$ and $\nu>0$ such that $\left(1-\frac{\beta_1}{\beta_2}\right)\varepsilon=\nu$. In this case agent $(h^{t-1},(\beta_1,\theta_N))$ is made indifferent with the perturbation and the original allocation. Further, agents with type $\beta>\beta_1$ dislike the perturbation. Therefore, since other incentive constraints with respect to type $(h^{t-1},(\beta_1,\theta_N))$ are slack, the perturbation is incentive compatible. However, in terms of resources we have

$$\begin{aligned}dE &= C'\left(u_t\left(h^{t-1},(\beta_1,\theta_N)\right)\right)(\varepsilon-\nu)\\ &\quad-\frac{1}{R_t}\mathbb{E}_t\left[C'\left(u_{t+1}\left(h^{t-1},(\beta_M,\theta_N),(\beta_{t+1},\theta_{t+1})\right)\right)|\theta_N\right]\frac{1}{\delta\beta_2}\varepsilon\\ &= C'\left(u_t\left(h^{t-1},(\beta_1,\theta_N)\right)\right)\left[\left(1-\frac{\kappa}{\beta_2}\right)\varepsilon-\nu\right]\\ &= C'\left(u_t\left(h^{t-1},(\beta_1,\theta_N)\right)\right)\left(\frac{\beta_1}{\beta_2}-\frac{\kappa}{\beta_2}\right)\varepsilon\end{aligned}$$

Since $\beta_2\leq 1$, then for $\varepsilon>0$ and $\nu>0$ we have $dE<0$ so that the planner saves resources. Those resources can be uniformly distributed across all types in the economy. Since $u_t\left(h^{t-1},(\beta_t,\theta_N)\right)>u_t\left(h^{t-1},(\beta_t,\theta_j)\right)$ for all $j<N$. There exists $\varepsilon>0$ small enough such that welfare improves with this redistribution and we reach a contradiction. □

Proof of Theorem 19

We first show result 1 in Theorem 19. Fix a period t and a history h^{t-1}. From Theorem 18 we have that there exists $\bar{\theta}_1 > 0$ such that all agents with a history in $\{(h^{t-1}, (\beta, \theta_1)) : \beta \in B\}$ are bunched at the same continuation allocation. In particular they all face the same inverse Euler equation distortion which we denote by

$$\kappa = \sum_{(\beta_{t+1}, \theta_{t+1}) \in B \times \Theta} \gamma_{t+1}(\beta_{t+1}) \rho_{t+1}(\theta_{t+1}|\theta_1) \left[\frac{C'(u_{t+1}(h^{t-1}, (\beta, \theta_1), (\beta_{t+1}, \theta_{t+1})))}{\delta R_t C'(u_t(h^{t-1}, (\beta, \theta_1)))} | \theta_N \right]$$

for all $\beta \in B$ and for a constant $\kappa > 0$. We want to show that $\kappa \geq 1$. Assume by way of contradiction that $\kappa > 1$. Then consider the following perturbation

$$\tilde{u}_t\left(h^{t-1}, (\beta, \theta_1)\right) = u_t\left(h^{t-1}, (\beta, \theta_1)\right) - \delta\varepsilon$$
$$\tilde{u}_{t+1}\left(h^{t-1}, (\beta, \theta_1), (\beta_{t+1}, \theta_{t+1})\right) = u_{t+1}\left(h^{t-1}, (\beta, \theta_1), (\beta_{t+1}, \theta_{t+1})\right) + \varepsilon$$

for all $(\beta_{t+1}, \theta_{t+1}) \in B \times \Theta$. This perturbation keeps welfare constant and therefore does not affect incentive compatibility at period $s < t$ (hyperbolic preferences). From Lemma 20 there exists $\bar{\theta}_1$ such that for $\theta_1 < \bar{\theta}_1$ we have that agents with histories in $\{(h^{t-1}, (\beta, \theta_1)) : \beta \in B\}$ incentives are strictly slack with respect to any other agent in the economy not in this group. Since for $\varepsilon > 0$ agents with $\beta \leq 1$ find this perturbation to weakly decrease the payoff of this allocation, we conclude the perturbation is incentive compatible. The marginal change in usage of resources is given by

$$dE = (\kappa - 1)\delta C'\left(u_t\left(h^{t-1}, (\beta, \theta_1)\right)\right)\varepsilon$$

For $\varepsilon > 0$ and $\kappa < 1$ we then get $dE < 0$, a contradiction as the planner can economize resources while keeping welfare constant. Hence we proved the first part of the theorem.

For part 2(a) of the Theorem, let

$$\kappa_H = \sum_{(\beta_{t+1},\theta_{t+1})\in B\times\Theta} \gamma_{t+1}\left(\beta_{t+1}\right)\rho_{t+1}\left(\theta_{t+1}|\theta_N\right)\frac{C'\left(u_{t+1}\left(h^{t-1},\left(\beta_M,\theta_N\right),\left(\beta_{t+1},\theta_{t+1}\right)\right)\right)}{\delta R_t C'\left(u_t\left(h^{t-1},\left(\beta_M,\theta_N\right)\right)\right)}$$

and assume by way of contradiction that $\kappa_H < 1$. Consider the perturbation

$$\tilde{u}_t\left(h^{t-1},\left(\beta_M,\theta_N\right)\right) = u_t\left(h^{t-1},\left(\beta_M,\theta_N\right)\right) - \delta\varepsilon$$
$$\tilde{u}_{t+1}\left(h^{t-1},\left(\beta_M,\theta_N\right),\left(\beta_{t+1},\theta_{t+1}\right)\right) = u_{t+1}\left(h^{t-1},\left(\beta_M,\theta_N\right),\left(\beta_{t+1},\theta_{t+1}\right)\right) + \varepsilon$$

for all $(\beta_{t+1},\theta_{t+1}) \in B\times\Theta$. Since $\beta_M = 1$, this is clearly incentive compatible for $\varepsilon > 0$. The marginal change in the usage of resource is given by

$$dE = \left(\kappa_H - 1\right)\delta C'\left(u_t\left(h^{t-1},\left(\beta_M,\theta_N\right)\right)\right)\varepsilon$$

therefore if $\kappa_H < 1$ we have that $dE < 0$ for $\varepsilon > 0$. From full support on the distribution of types, the planner can generate strictly positive resources with this incentive compatible perturbation, while keeping welfare constant. This is a contradiction as the extra resources can be used to strictly improve welfare. Thus we proved part 2(a) of the Theorem.

For part 2(b), notice that from Lemma 21 and from Theorem 18 we have that there exists $\bar{\theta}_{N-1} < \theta_N$ and $\bar{\beta}_1 > 0$ such that for $\theta_1 < \ldots < \theta_{N-1} \leq \bar{\theta}_{N-1}$ and $0 < \beta_1 < \bar{\beta}_1$ we have that agents with histories $\{(h^{t-1},(\beta,\theta)) : \beta \in B\, and\, \theta < \theta_N\}$ strictly prefer their own allocation to the allocation of any agent with a history $\{(h^{t-1},(\beta,\theta_N)) : \beta \in B\}$. Further we have $u_t\left(\left(h^{t-1},\left(\beta_1,\theta_N\right)\right)\right) > u_t\left(\left(h^{t-1},\left(\beta,\theta\right)\right)\right)$ for all $\theta < \theta_N$ and all $\beta \in B$. Denote

$$\kappa_H^1 = \sum_{(\beta_{t+1},\theta_{t+1})\in B\times\Theta} \gamma_{t+1}\left(\beta_{t+1}\right)\rho_{t+1}\left(\theta_{t+1}|\theta_N\right)\frac{C'\left(u_{t+1}\left(h^{t-1},\left(\beta_1,\theta_N\right),\left(\beta_{t+1},\theta_{t+1}\right)\right)\right)}{\delta R_t C'\left(u_t\left(h^{t-1},\left(\beta_1,\theta_N\right)\right)\right)}$$

and assume by way of contradiction that $\kappa_H^1 \geq 1$. Consider the following perturbation

$$\begin{aligned}
\tilde{u}_t\left(h^{t-1},(\beta_1,\theta_N)\right) &= u_t\left(h^{t-1},(\beta_1,\theta_N)\right)+\beta_1\delta\varepsilon+\nu \\
\tilde{u}_{t+1}\left(h^{t-1},(\beta_1,\theta_N),(\beta_{t+1},\theta_{t+1})\right) &= u_{t+1}\left(h^{t-1},(\beta_1,\theta_N),(\beta_{t+1},\theta_{t+1})\right)-\varepsilon \\
\tilde{u}_t\left(h^{t-1},(\beta,\theta)\right) &= \tilde{u}_t\left(h^{t-1},(\beta,\theta)\right)+\nu
\end{aligned}$$

for all $(\beta_{t+1},\theta_{t+1}) \in B\times\Theta$ and all $(\beta,\theta)\neq(\beta_1,\theta_N)$. For $\varepsilon>0$ and $\nu>0$, this perturbation is incentive compatible since agents with $\beta\geq\beta_1$ find it (weakly) unattractive. In order to keep welfare unchanged with this perturbation we need

$$dW = \pi\left(\beta_1,\theta_N|h^{t-1}\right)(\beta_1-1)\,\delta\varepsilon+\nu=0$$

so that $\nu=\pi\left(\beta_1,\theta_N|h^{t-1}\right)(1-\beta_1)\,\delta\varepsilon$. The marginal change in resource usage of this perturbation is given by

$$\begin{aligned}
dE =& \pi\left(\beta_1,\theta_N|h^{t-1}\right)C'\left(u_t\left(h^{t-1},(\beta_1,\theta_N)\right)\right)\delta\left(\beta_1-\kappa_H^1\right)\varepsilon \\
&+\nu\sum_{(\beta,\theta)}\pi\left(\beta,\theta|h^{t-1}\right)C'\left(u_t\left(h^{t-1},(\beta,\theta)\right)\right) \\
=& A\left(h^{t-1}\right)\left[\left(\frac{\beta_1-\kappa_H^1}{1-\beta_1}\right)+\sum_{(\beta,\theta)}\pi\left(\beta,\theta|h^{t-1}\right)\frac{C'\left(u_t\left(h^{t-1},(\beta,\theta)\right)\right)}{C'\left(u_t\left(h^{t-1},(\beta_1,\theta_N)\right)\right)}\right]\varepsilon
\end{aligned}$$

where $A\left(h^{t-1}\right) = \pi\left(\beta_1,\theta_N|h^{t-1}\right)\delta C'\left(u_t\left(h^{t-1},(\beta_1,\theta_N)\right)\right)(1-\beta_1) > 0$ Since we have $u_t\left(h^{t-1},(\beta_1,\theta_N)\right) > u_t\left(h^{t-1},(\beta,\theta)\right)$ for $\theta<\theta_N$ and $u_t\left(h^{t-1},(\beta_1,\theta_N)\right)\geq u_t\left(h^{t-1},(\beta,\theta_N)\right)$ for all β, then we have

$$\sum_{(\beta,\theta)}\pi\left(\beta,\theta|h^{t-1}\right)\frac{C'\left(u_t\left(h^{t-1},(\beta,\theta)\right)\right)}{C'\left(u_t\left(h^{t-1},(\beta_1,\theta_N)\right)\right)}<1$$

and since $\beta_1 < 1 \leq \kappa_H^1$ we conclude that $dE < 0$ for $\varepsilon > 0$, a contradiction. Hence we proved 2(b).

A.4 Proofs for Section 1.5

In this Appendix we provide a proof for Theorem 4. For the first part of the Theorem the proof is analogous to the proof of Theorem 1 and relies only on the convexity of the set of incentive constraints when we rewrite the problem in terms of utility levels and the wedge level. For the second part of the Theorem, note that for $\theta_1 < \ldots < \theta_{N-1} \leq \bar{\theta}_N$ we have $v(\theta_N, \alpha) > v(\theta_1)$. Therefore by incentive compatibility it must be the case that

$$u(\theta_N, \alpha) + \alpha\varepsilon(\theta_N, \alpha) > u(\theta_1) + \alpha\varepsilon(\theta_1)$$

In particular, for $\alpha = 0$ we have $u(\theta_N, \alpha) > u(\theta, \alpha')$ for all (θ, α'). Assume by way of contradiction that $u(\theta_N, \alpha) = u(\theta_N)$, $v(\theta_N, \alpha) = v(\theta_N)$ and $a(\theta_N, \alpha) = a(\theta_N)$ at the solution to the planner's problem. Then we have $u(\theta_N) > u(\theta_1)$. Let $\alpha_{max} = \max A \geq 0$. If $b'(a(\theta_N)) > p$, then consider the following change to the allocation for :

$$\begin{aligned}
\tilde{a}(\theta_N, \alpha_{max}) &= \alpha(\theta_N) + \nu \\
\tilde{u}(\theta_N, \alpha_{max}) &= u(\theta_N) - \alpha_{max}\nu + \eta \\
\tilde{u}(\theta, \alpha) &= \tilde{u}(\theta, \alpha) + \eta
\end{aligned}$$

for all $(\theta, \alpha) \neq (\theta_N, \alpha_{max})$. For $\nu > 0$ this perturbation is incentive compatible since for $\alpha < \alpha_{max}$ the change in the deviation payoff into the allocation of (θ_N, α_{max}) is $(\alpha - \alpha_{max})\nu < 0$ and the original allocation is incentive compatible. Now the

marginal change in resources used is given by

$$dE = \pi\left(\theta_N, \alpha_{max}\right)\left[-\alpha_{max}\nu C'\left(u\left(\theta_N\right)\right) - \left(\frac{b'\left(\varepsilon^{-1}\left(\varepsilon\left(\theta_N\right)\right)\right)}{\varepsilon'\left(\varepsilon^{-1}\left(\varepsilon\left(\theta_N\right)\right)\right)} - \frac{p}{\varepsilon'\left(\varepsilon^{-1}\left(\varepsilon\left(\theta_N\right)\right)\right)}\right)\nu\right]$$
$$+ \eta \sum_{\theta,\alpha} \pi\left(\theta, \alpha\right) C'\left(u\left(\theta, \alpha\right)\right)$$

If $\alpha_{max} = 0$, then we already get a contradiction since for $\nu > 0$ we can set $\eta > 0$ small enough such that welfare increases. For $\alpha_{max} > 0$, the marginal change in the government's objective is given by

$$d\mathcal{W} = -\pi\left(\theta_N, \alpha_{max}\right)\lambda\left(\theta_N\right)\alpha_{max}\nu + \eta$$

If we set $\eta = \pi\left(\theta_N, \alpha_{max}\right)\lambda\left(\theta_N\right)\alpha_{max}\nu$ so that $d\mathcal{W} = 0$, then the marginal change in resources used is given by

$$\frac{dE}{B} = \nu\left\{\lambda\left(\theta_N\right)\sum_{\theta,\alpha}\pi\left(\theta, \alpha\right)\frac{C'\left(u\left(\theta, \alpha\right)\right)}{C'\left(u\left(\theta_N\right)\right)} - 1\right\}$$
$$- \frac{\nu}{\alpha_{max}C'\left(u\left(\theta_N\right)\right)}\left(\frac{b'\left(\varepsilon^{-1}\left(\varepsilon\left(\theta_N\right)\right)\right)}{\varepsilon'\left(\varepsilon^{-1}\left(\varepsilon\left(\theta_N\right)\right)\right)} - \frac{p}{\varepsilon'\left(\varepsilon^{-1}\left(\varepsilon\left(\theta_N\right)\right)\right)}\right)$$

where $B = \pi\left(\theta_N, \alpha_{max}\right)\alpha_{max}C'\left(u\left(\theta_N\right)\right) > 0$Since $\lambda\left(\theta_N\right)$ and since $C\left(\cdot\right)$ is convex and $u\left(\theta_N\right) \geq u\left(\theta, \alpha\right)$, then $dE < 0$ for $\nu > 0$, a contradiction. Therefore we must have $b'\left(a\left(\theta_N\right)\right) \leq p$. Again we can make an analogous perturbation for $\alpha_{min} = \min A \leq 0$. Hence we conclude $b'\left(a\left(\theta_N\right)\right) = p$. If $\alpha_{max} > 0$, note that the exact same perturbation above also implies $dE < 0$ since $u\left(\theta_N\right) > u\left(\theta, \alpha\right)$ for $\theta < \theta_N$, a contradiction. The case for $\alpha_{min} < 0$ is analogous. Finally we exhausted all possible cases and reach a contradiction to our initial assumption that all agents in $\{(\theta_N, \alpha) : \alpha \in A\}$ receive the same allocation.

A.5 Life-cycle calibration with hyperbolic preferences

In this appendix we build a incomplete markets life-cycle model with hyperbolic discounting. This model allows us to calibrate for β's and δ separately. We compare the heterogeneity generated by this calibration and the benchmark calibration and find it to be very similar so that in terms of savings distortions the two calibrations yield very similar results.

Households start their lives at age 25 and have a working life of up to age 65. At age 65 households retire and live up to age 99 at most. At each period there is a probability s_t of survival until the next period (from 2010 U.S. Life Tables). At each period, households receive endowment shocks e_t that follows

$$\ln e_{i,t+1} = \phi_{t+1} + \rho \ln e_{i,t} + \nu_{i,t+1}$$

where ϕ_{t+1} is the life-cycle component and $\nu_{i,t+1} \sim N(0, \sigma^2)$. We estimate ϕ_{t+1}, ρ and σ^2 from the PSID from 1999-2009. We find $\rho = 0.829$ and $\sigma = 0.42$.[12]

Each period workers can save $a_t \geq 0$ resources into the next period at a gross rate of interest R. Finally workers pay taxes on labor and asset income according to

$$T(y_t, Ra_{t-1}) = \left(t_0 + y_t - t_1 y_t^{1-t_2}\right) + t_a Ra_{t-1}$$

[12]In particular, in order to estimate ϕ_{t+1} as the set of age dummies in a regression of reported labor income on fixed household effects and a set of controls for the business cycle (polinomial on state level unemployment rates) and for demographics (family size, marital status, number of children). Then as a measure of e_t we use fixed effects plus residuals from this regression. Then we estimate an AR(1) with this measure to obtain ρ and σ estimates.

We assume linearity on asset income taxes to make the dynamic model more tractable. Using the Cross National equivalent PSID files from 1999-2009 and using non-linear least squares, we find $t_0 = 2488$, $t_1 = 1.67$, $t_2 = 0.07$ and $t_a = 0.28$.[13]

After age 65, individuals receive no labor income, only asset income and social security benefits $b\left(y_{65}\right)$. We follow [43] and [67] in assuming that retirement benefits depend only on labor income only in the last period. For $b(\cdot)$ we use the current schedule on social security old-age benefits. Finally for tax purposes only 55% of social security benefits are included into taxable income so we take that into account.

Agent's have hyperbolic preferences (with a constant β present-bias over time) and therefore choices during working life are solutions to

$$U_t\left(x_t, s_t\right) = \max u\left(c_t\right) + \beta\delta\mathbb{E}_t\left[V_{t+1}\left(a_{t+1}, y_{t+1}\right)|y_t\right]$$

$$s.t. c_t + a_{t+1} = x_t$$

$$a_{t+1} \geq 0$$

where $u\left(c\right) = \frac{c^{1-\frac{1}{\gamma}}}{1-\frac{1}{\gamma}}$ and

$$V_t\left(a_t, s_t\right) = u\left(c_t\left(a_t, y_t\left(s_t\right), s_t\right)\right) + \delta\mathbb{E}_t\left[V_{t+1}\left(a_{t+1}\left(a_t, y_t\left(s_t\right), s_t\right), s_{t+1}\right)\right]$$

$$x_{t+1} = R_{t+1}a_{t+1} + y_{t+1}\left(s_{t+1}\right) - T\left(y_{t+1}\left(s_{t+1}\right), R_{t+1}a_{t+1}\right)$$

[13] All variables in 2010 dollars, scaled by family size as

$$size = 1 + 0.7 * other\ adults + 0.5 * children$$

used in Attanazio and Pistaferri (2014). We consider overall household income for households a head within 25-65 years old, that worked more than 260 hours in the year and that made at least \$4000 in labor income. We exclude households with negative asset income and household in the highest percentile of asset income. Finally, we exclude the lowest and highest percentile on TAXSIM estimated taxes paid.

We then obtain the following Euler equation:

$$u'\left(c_t^i\left(x_t^i, s_t^i\right)\right) = (1-\tau_a) R\delta \mathbb{E}_t\left[u'\left(c_{t+1}^i\left(x_{t+1}^i, s_{t+1}^i\right)\right)|s_t^i\right] - (1-\tau_a) R\delta (1-\beta) \mathbb{E}_t\left[MPC_{t+1}^i\left(x_{t+1}, s_{t+1}\right) u'\left(c_{t+1}^i\left(x_{t+1}^i, s_{t+1}^i\right)\right)|s_t^i\right]$$

where

$$MPC_{t+1}^i\left(x_{t+1}, s_{t+1}\right) = \frac{\partial c_{t+1}^i\left(x_{t+1}^i, s_{t+1}^i\right)}{\partial x_{t+1}^i}$$

Finally we assume $\frac{\beta}{2} \sim Beta\left(a, b\right)$ in the population so that we allow both for present bias and future bias.

We match exactly the same percentiles of the distribution of private wealth at retirement to lifetime earnings ratio. In Table A.1 we see the model matches exactly the three statistics from the data. Finally in the table we report statistics on the distribution of β we find as well as on the distribution of lifetime average time inconsistency wedge:

$$\tau_\beta = 1 - (1-\beta) MPC_{t+1}^i\left(x_{t+1}, s_{t+1}\right)$$

This second measure gives us a better sense of how distorted savings decisions are each year on average. Notice that those distortions are relatively small in a yearly basis, but when compounded over the life-cycle a distortion of $\tau_\beta = 0.96$ implies individuals care very little about retirement when young in life as compared with individuals with $\tau_\beta = 1.0$.

Table A.1: Life-Cycle Calibration Results

Statistic	Data	Value
δ		0.9856
$avg\ \beta$		0.80
$std\,\beta$		0.62
avg		0.96
$std\ b$		0.058
$corr\,(b, LT\ Labor\ Earnings)$		-0.01
W/Y percentile 25%	0.043	0.043
W/Y percentile 50%	0.0906	0.0906
W/Y percentile 75%	0.1768	0.1768

Appendix B

Appendix to Chapter 2

B.1 Proofs

B.1.1 Proof of Theorem 7

We can re-write the problem in terms of utility levels and eliminate for consumption. Then in the relaxed problem, the planner chooses functions U and y to solve

$$\max_{U,y} \int_{\underline{p}}^{\bar{p}} \left[\int_{\underline{\theta}}^{\bar{\theta}} \alpha(\theta) U(\theta,p) f(\theta|p)\, d\theta \right] h(p)\, dp$$

$$s.t.\ U(\theta,p) = U(\underline{\theta},p) + \int_{\underline{\theta}}^{\theta} \frac{y\left(\tilde{\theta},p\right)}{\tilde{\theta}^2} v'\left(\frac{y\left(\tilde{\theta},p\right)}{\tilde{\theta}}\right) d\tilde{\theta}\ , \forall (\theta,p) \in \left[\underline{\theta},\bar{\theta}\right] \times [\underline{p},\bar{p}]$$

$$\int_{\underline{\theta}}^{\bar{\theta}} U(\theta,p') f(\theta|p')\, d\theta = V(\underline{p}) + \int_{\underline{p}}^{p'} \left(\int_{\underline{\theta}}^{\bar{\theta}} U(\theta,p) \frac{\partial f(\theta|p)}{\partial p} d\theta \right) dp\ , \forall p' \in [\underline{p},\bar{p}]$$

$$\int_{\underline{p}}^{\bar{p}} \left[\int_{\underline{\theta}}^{\bar{\theta}} \left[y(\theta,p) - U(\theta,p) - v\left(\frac{y(\theta,p)}{\theta}\right) \right] f(\theta|p)\, d\theta \right] h(p)\, dp = R$$

The Lagrangian to this problem can be written as Let $\mu(\theta,p)$, $\gamma(p)$ and λ be the Lagrange multipliers on each of the respective constraints. Then we can write the

Lagrangian as

$$\begin{aligned}
\mathcal{L} = & \int_{\underline{\mathrm{p}}}^{\bar{p}} \left[\int_{\underline{\theta}}^{\bar{\theta}} \alpha(\theta) U(\theta, p) f(\theta|p) \, d\theta \right] h(p) \, dp \\
& + \int_{\underline{\mathrm{p}}}^{\bar{p}} \left[\int_{\underline{\theta}}^{\bar{\theta}} \mu(\theta, p) \left[U(\theta, p) - U(\underline{\theta}, p) \right] f(\theta|p) \, d\theta \right] h(p) \, dp \\
& - \int_{\underline{\mathrm{p}}}^{\bar{p}} \int_{\underline{\theta}}^{\bar{\theta}} \mu(\theta, p) \int_{\underline{\theta}}^{\theta} \frac{y\left(\tilde{\theta}, p\right)}{\tilde{\theta}^2} v' \left(\frac{y\left(\tilde{\theta}, p\right)}{\tilde{\theta}} \right) d\tilde{\theta} f(\theta|p) \, d\theta h(p) \, dp \\
& + \int_{\underline{\mathrm{p}}}^{\bar{p}} \gamma(p) \int_{\underline{\theta}}^{\bar{\theta}} U(\theta, p) f(\theta|p) \, d\theta - V(\underline{\mathrm{p}}) h(p) \, dp \\
& - \int_{\underline{\mathrm{p}}}^{\bar{p}} \gamma(p) \int_{\underline{\mathrm{p}}}^{p} \left(\int_{\underline{\theta}}^{\bar{\theta}} U(\theta, p') \frac{\partial f(\theta|p')}{\partial p} d\theta \right) dp' h(p) \, dp \\
& + \lambda \left[R + \int_{\underline{\mathrm{p}}}^{\bar{p}} \left[\int_{\underline{\theta}}^{\bar{\theta}} \left[y(\theta, p) - U(\theta, p) - v\left(\frac{y(\theta, p)}{\theta} \right) \right] f(\theta|p) \, d\theta \right] h(p) \, dp \right]
\end{aligned}$$

Define $\Gamma(p) = \int_{p}^{\bar{p}} \gamma(\tilde{p}) h(\tilde{p}) \, d\tilde{p}$, then notice that by a change of variables we get

$$\begin{aligned}
& \int_{\underline{\mathrm{p}}}^{\bar{p}} \gamma(p) \int_{\underline{\mathrm{p}}}^{p} \int_{\underline{\theta}}^{\bar{\theta}} U(\theta, p') \frac{\partial f(\theta|p')}{\partial p} d\theta dp' h(p) \, dp \\
& \qquad = \int_{\underline{\mathrm{p}}}^{\bar{p}} \int_{\underline{\theta}}^{\bar{\theta}} \Gamma(p) U(\theta, p) \frac{\partial f(\theta|p)}{\partial p} d\theta dp
\end{aligned}$$

Define $M(\theta, p) = \int_{\theta}^{\bar{\theta}} \mu\left(\tilde{\theta}, p\right) f\left(\tilde{\theta}|p\right) d\tilde{\theta}$, the using a similar change of variables we have

$$\begin{aligned}
& \int_{\underline{\theta}}^{\bar{\theta}} \mu(\theta, p) \int_{\underline{\theta}}^{\theta} \frac{y\left(\tilde{\theta}, p\right)}{\tilde{\theta}^2} v' \left(\frac{y\left(\tilde{\theta}, p\right)}{\tilde{\theta}} \right) d\tilde{\theta} f(\theta|p) \, d\theta \\
& \qquad = \int_{\underline{\theta}}^{\bar{\theta}} M(\theta, p) \frac{y(\theta, p)}{\tilde{\theta}^2} v' \left(\frac{y(\theta, p)}{\theta} \right) d\theta
\end{aligned}$$

Further we have

$$\int_{\underline{\mathrm{p}}}^{\bar{p}} \gamma(p) V(\underline{\mathrm{p}}) \, dp h(p) \, dp = \Gamma(\underline{\mathrm{p}}) \int_{\underline{\theta}}^{\bar{\theta}} U(\theta, \underline{\mathrm{p}}) f(\theta|\underline{\mathrm{p}}) \, d\theta$$

Therefore the Lagrangian can be re-written as Therefore the Lagrangian can be rewritten as

$$
\begin{aligned}
\mathcal{L} = & \int_{\underline{\mathrm{p}}}^{\bar{p}} \left[\int_{\underline{\theta}}^{\bar{\theta}} \alpha(\theta) U(\theta, p) f(\theta|p) h(p) d\theta \right] dp \\
& + \int_{\underline{\mathrm{p}}}^{\bar{p}} \left[\int_{\underline{\theta}}^{\bar{\theta}} \mu(\theta, p) U(\theta, p) f(\theta|p) h(p) d\theta \right] dp - M(\underline{\theta}, p) U(\underline{\theta}, p) \\
& - \int_{\underline{\mathrm{p}}}^{\bar{p}} \left[\int_{\underline{\theta}}^{\bar{\theta}} M(\theta, p) \frac{y(\theta, p)}{\theta^2} v' \left(\frac{y(\theta, p)}{\theta} \right) h(p) d\theta \right] dp \\
& + \int_{\underline{\mathrm{p}}}^{\bar{p}} \left[\int_{\underline{\theta}}^{\bar{\theta}} \gamma(p) U(\theta, p) f(\theta|p) h(p) d\theta \right] dp \\
& - \Gamma(\underline{\mathrm{p}}) \int_{\underline{\theta}}^{\bar{\theta}} U(\theta, \underline{\mathrm{p}}) f(\theta|\underline{\mathrm{p}}) d\theta \\
& - \int_{\underline{\mathrm{p}}}^{\bar{p}} \left(\int_{\underline{\theta}}^{\bar{\theta}} \Gamma(p) U(\theta, p) \frac{\partial f(\theta|p)}{\partial p} d\theta \right) dp \\
& + \int_{\underline{\mathrm{p}}}^{\bar{p}} \left[\int_{\underline{\theta}}^{\bar{\theta}} \lambda \left[y(\theta, p) - U(\theta, p) - v \left(\frac{y(\theta, p)}{\theta} \right) \right] f(\theta|p) h(p) d\theta \right] dp \\
& + \lambda R
\end{aligned}
$$

The first order conditions to this problem give us

$$[U(\theta,p),\ p>\underline{p},\ \theta>\underline{\theta}]:\alpha(\theta)+\mu(\theta,p)+\gamma(p)-\frac{\Gamma(p)}{f(\theta|p)\,h(p)}\frac{\partial f(\theta|p)}{\partial p}=\lambda$$

$$[U(\underline{\theta},p)]:M(\underline{\theta},p)=0$$

$$[y(\theta,p)]:\lambda\left(1-\frac{1}{\theta}v'\left(\frac{y(\theta,p)}{\theta}\right)\right)f(\theta|p)$$
$$=M(\theta,p)\left[\frac{1}{\theta^2}v'\left(\frac{y(\theta,p)}{\theta}\right)+\frac{y(\theta,p)}{\theta^3}v''\left(\frac{y(\theta,p)}{\theta}\right)\right]$$

$$[U(\theta,\underline{p}),\ \theta>\underline{\theta}]:\ -\Gamma(\underline{p})\,f(\theta|\underline{p})=0$$

$$[IC_2]:U(\theta,p)=U(\underline{\theta},p)+\int_{\underline{\theta}}^{\theta}\frac{y\left(\tilde{\theta},p\right)}{\tilde{\theta}^2}v'\left(\frac{y\left(\tilde{\theta},p\right)}{\tilde{\theta}}\right)d\tilde{\theta}$$

$$[Feasibility]:R+\int_{\underline{p}}^{\bar{p}}\left[\int_{\underline{\theta}}^{\bar{\theta}}\left[y(\theta,p)-U(\theta,p)\right]f(\theta|p)\,d\theta\right]h(p)\,dp$$
$$-\int_{\underline{p}}^{\bar{p}}\left[\int_{\underline{\theta}}^{\bar{\theta}}v\left(\frac{y(\theta,p)}{\theta}\right)f(\theta|p)\,d\theta\right]h(p)\,dp=0$$

$$[IC_1]:\ \int_{\underline{\theta}}^{\bar{\theta}}U(\theta,p)\,f(\theta|p)\,d\theta$$
$$=\int_{\underline{\theta}}^{\bar{\theta}}U(\theta,\underline{p})\,f(\theta|\underline{p})\,d\theta$$
$$+\int_{\underline{p}}^{p}\left(\int_{\underline{\theta}}^{\bar{\theta}}U(\theta,p')\,\frac{\partial f(\theta|p')}{\partial p}d\theta\right)dp'$$

So that

$$\mu(\theta,p)=\lambda-\alpha(\theta)-\gamma(p)+\frac{\Gamma(p)}{f(\theta|p)\,h(p)}\frac{\partial f(\theta|p)}{\partial p}$$

$$M(\theta,p)=\int_{\theta}^{\bar{\theta}}\mu(\theta',p)\,f(\theta'|p)\,d\theta'$$
$$=\int_{\theta}^{\bar{\theta}}\left[\lambda-\alpha(\theta')-\gamma(p)\right]f(\theta'|p)\,d\theta'$$
$$+\int_{\theta}^{\bar{\theta}}\frac{\Gamma(p)}{f(\theta'|p)\,h(p)}\frac{\partial f(\theta'|p)}{\partial p}f(\theta'|p)\,d\theta'$$

Since $M(\underline{\theta}, p) = 0$ and $\int_{\underline{\theta}}^{\bar{\theta}} \frac{\partial f(\theta'|p)}{\partial p} d\theta' = 0$ we conclude

$$\gamma(p) = \int_{\underline{\theta}}^{\bar{\theta}} [\lambda - \alpha(\theta)] f(\theta|p) \, d\theta$$

From the definition of Γ we then get

$$\begin{aligned}\Gamma(p') &= \int_{p'}^{\bar{p}} \int_{\underline{\theta}}^{\bar{\theta}} [\lambda - \alpha(\theta)] f(\theta|p) \, d\theta h(p) \, dp \\ &= \lambda [1 - H(p')] - \int_{p'}^{\bar{p}} \int_{\underline{\theta}}^{\bar{\theta}} \alpha(\theta) f(\theta|p) \, d\theta h(p) \, dp\end{aligned}$$

and

$$\Gamma(\underline{\mathrm{p}}) = \lambda - \int_{\underline{\mathrm{p}}}^{\bar{p}} \int_{\underline{\theta}}^{\bar{\theta}} \alpha(\theta) f(\theta|p') \, d\theta h(p') \, dp'$$

hence $\lambda = 1$. Therefore

$$\begin{aligned}M(\theta', p) = &(1 - F(\theta'|p)) \int_{\underline{\theta}}^{\bar{\theta}} \alpha(\theta) f(\theta|p) \, d\theta \\ &- (1 - F(\theta'|p)) \int_{\theta'}^{\bar{\theta}} \alpha(\theta) \frac{f(\theta|p)}{1 - F(\theta'|p)} d\theta \\ &+ \left(\frac{1 - H(p)}{h(p)}\right) \int_{\theta'}^{\bar{\theta}} \frac{\partial f(\theta|p)}{\partial p} d\theta \\ &- \left(\frac{1 - H(p)}{h(p)}\right) \int_{p}^{\bar{p}} \int_{\underline{\theta}}^{\bar{\theta}} \alpha(\theta) f(\theta|p') \, d\theta \left(\frac{h(p')}{1 - H(p)}\right) dp' \int_{\theta'}^{\bar{\theta}} \frac{\partial f(\theta|p)}{\partial p} d\theta\end{aligned}$$

Define the labor wedge by

$$\frac{1}{\theta} v' \left(\frac{y(\theta, p)}{\theta}\right) = 1 - \tau^{\ell}(\theta, p)$$

Then from the first order conditions we have

$$M(\theta, p) \left[\frac{1 - \tau^{\ell}(\theta, p)}{\theta} + \frac{1 - \tau^{\ell}(\theta, p)}{\theta} \frac{1}{\varepsilon(\theta, p)}\right] = \tau^{\ell}(\theta, p) f(\theta|p)$$

Where $\varepsilon(\theta, p)$ is the labor supply elasticity. Thus

$$\frac{\tau^{\ell}(\theta, p)}{1-\tau^{\ell}(\theta, p)} = \frac{M(\theta, p)}{\theta f(\theta|p)}\left(1+\frac{1}{\varepsilon(\theta, p)}\right)$$

and we obtain the formula in Theorem 7.

B.1.2 Proof of Theorem 8

In this case the planner's problem is given by

$$\max_{U,y} \sum_{p} h(p)\left[\int_{\underline{\theta}}^{\bar{\theta}} \alpha(\theta) U(\theta, p) f(\theta|p)\, d\theta\right]$$

$$s.t.\ U(\theta, p) = U(\underline{\theta}, p) + \int_{\underline{\theta}}^{\theta} \frac{y\left(\tilde{\theta}, p\right)}{\tilde{\theta}^2} v'\left(\frac{y\left(\tilde{\theta}, p\right)}{\tilde{\theta}}\right) d\tilde{\theta}\ , \forall(\theta, p) \in \left[\underline{\theta}, \bar{\theta}\right] \times [\underline{\mathrm{p}}, \bar{p}]$$

$$\int_{\underline{\theta}}^{\bar{\theta}} \left[U(\theta, p_H) - U(\theta, p_L)\right] f(\theta|p_H)\, d\theta \geq 0$$

$$\int_{\underline{\theta}}^{\bar{\theta}} \left[U(\theta, p_L) - U(\theta, p_H)\right] f(\theta|p_L)\, d\theta \geq 0$$

$$\sum_{p} h(p) \int_{\underline{\theta}}^{\bar{\theta}} \left[y(\theta, p) - U(\theta, p) - v\left(\frac{y(\theta, p)}{\theta}\right)\right] f(\theta|p)\, d\theta = R$$

Analogously to the proof of Theorem 7, we can write the Lagrangian as

$$\begin{aligned}
\mathcal{L} = & \sum_{p} h(p) \int_{\underline{\theta}}^{\bar{\theta}} \alpha(\theta) U(\theta, p) f(\theta|p) \, d\theta \\
& + \sum_{p} h(p) \int_{\underline{\theta}}^{\bar{\theta}} \mu(\theta, p) U(\theta, p) f(\theta|p) \, d\theta - M(\underline{\theta}, p) U(\underline{\theta}, p) \\
& - \sum_{p} h(p) \int_{\underline{\theta}}^{\bar{\theta}} M(\theta, p) \frac{y(\theta, p)}{\theta^2} v'\left(\frac{y(\theta, p)}{\theta}\right) d\theta \\
& + \sum_{p} h(p) \gamma(p) \int_{\underline{\theta}}^{\bar{\theta}} \left[U(\theta, p) - U(\theta, p')\right] f(\theta|p) \, d\theta \\
& + \lambda \left\{ R + \sum_{p} h(p) \int_{\underline{\theta}}^{\bar{\theta}} \left(y(\theta, p) - U(\theta, p) - v\left(\frac{y(\theta, p)}{\theta}\right)\right) f(\theta|p) \, d\theta \right\}
\end{aligned}$$

The first order conditions then are given by

$$\begin{aligned}
\left[U(\theta, p), \theta > \underline{\theta}\right] : & \alpha(\theta) + \mu(\theta, p) + \gamma(p) - \gamma(p') \frac{h(p')}{h(p)} \frac{f(\theta|p')}{f(\theta|p)} = \lambda \\
\left[U(\underline{\theta}, p)\right] : & M(\underline{\theta}, p) = 0 \\
\left[y(\theta, p)\right] : & -M(\theta, p) \left[\frac{1}{\theta^2} v'\left(\frac{y(\theta, p)}{\theta}\right) + \frac{y(\theta, p)}{\theta^3} v''\left(\frac{y(\theta, p)}{\theta}\right)\right] \\
& + \lambda \left[1 - \frac{1}{\theta} v'\left(\frac{y(\theta, p)}{\theta}\right)\right] f(\theta|p) = 0
\end{aligned}$$

Integrating the first equation with respect to θ and using the cdf $f(\theta|p)$, we obtain

$$\int_{\underline{\theta}}^{\bar{\theta}} \left[\alpha(\theta) + \mu(\theta, p) + \gamma(p) - \gamma(p') \frac{h(p')}{h(p)} \frac{f(\theta|p')}{f(\theta|p)}\right] f(\theta|p) \, d\theta = \lambda$$

Since $M(\underline{\theta}, p) = 0$, we obtain

$$\int_{\underline{\theta}}^{\bar{\theta}} \alpha(\theta) f(\theta|p) \, d\theta + \gamma(p) - \gamma(p') \frac{h(p')}{h(p)} = \lambda$$

So

$$\lambda = \sum_{p} h(p) \int_{\underline{\theta}}^{\bar{\theta}} \alpha(\theta) f(\theta|p) \, d\theta = 1$$

Since $f(\theta|p_H)$ FOSD $f(\theta|p_L)$ and $\alpha'(\cdot) < 0$. Then

$$\int_{\underline{\theta}}^{\bar{\theta}} \alpha(\theta) f(\theta|p_L) \, d\theta > \int_{\underline{\theta}}^{\bar{\theta}} \alpha(\theta) f(\theta|p_H) \, d\theta$$

Since $\lambda = 1$, we must have

$$\gamma(p_H) - \gamma(p_L) \frac{h(p_L)}{h(p_H)} > \gamma(p_L) - \gamma(p_H) \frac{h(p_H)}{h(p_L)}$$

Hence

$$\gamma(p_H) \left[1 + \frac{h(p_H)}{h(p_L)}\right] > \gamma(p_L) \left[1 + \frac{h(p_L)}{h(p_H)}\right]$$

Hence $\gamma(p_H) > 0$.

Note further that

$$\frac{\tau(\theta, p)}{1 - \tau(\theta, p)} = \left(1 + \frac{1}{\varepsilon}\right) \frac{M(\theta, p)}{\theta f(\theta|p)}$$

and

$$\begin{aligned} M(\theta, p) = 1 - F(\theta|p) - \int_{\theta}^{\bar{\theta}} \alpha\left(\tilde{\theta}\right) f\left(\tilde{\theta}|p\right) d\tilde{\theta} \\ - (1 - F(\theta|p)) \left[\gamma(p) - \gamma(p') \left(\frac{1 - F(\theta|p')}{1 - F(\theta|p)}\right) \left(\frac{h(p')}{h(p)}\right)\right] \end{aligned}$$

Thus

$$\begin{aligned} \frac{\tau(\theta, p)}{1 - \tau(\theta, p)} = \left(1 + \frac{1}{\varepsilon}\right) \frac{1 - F(\theta|p)}{\theta f(\theta|p)} \left\{1 - \int_{\theta}^{\bar{\theta}} \alpha\left(\tilde{\theta}\right) \frac{f\left(\tilde{\theta}|p\right)}{1 - F(\theta|p)} d\tilde{\theta}\right\} \\ - \left(1 + \frac{1}{\varepsilon}\right) \frac{1 - F(\theta|p)}{\theta f(\theta|p)} \left[\gamma(p) - \gamma(p') \left(\frac{1 - F(\theta|p')}{1 - F(\theta|p)}\right) \left(\frac{h(p')}{h(p)}\right)\right] \end{aligned}$$

Hence

$$\begin{aligned}
\frac{\tau(\theta, p_H)}{1-\tau(\theta, p_H)} &= \left(1+\frac{1}{\varepsilon}\right)\frac{1-F(\theta|p_H)}{\theta f(\theta|p_H)}\left\{1-\int_{\theta}^{\bar{\theta}}\alpha\left(\tilde{\theta}\right)\frac{f\left(\tilde{\theta}|p_H\right)}{1-F(\theta|p_H)}d\tilde{\theta}\right\} \\
&\quad -\left(1+\frac{1}{\varepsilon}\right)\frac{1-F(\theta|p_H)}{\theta f(\theta|p_H)}\gamma(p_H) \\
&\quad +\left(1+\frac{1}{\varepsilon}\right)\frac{1-F(\theta|p_H)}{\theta f(\theta|p_H)}\gamma(p_L)\left(\frac{1-F(\theta|p_L)}{1-F(\theta|p_H)}\right)\left(\frac{h(p_L)}{h(p_H)}\right) \\
\frac{\tau(\theta, p_L)}{1-\tau(\theta, p_L)} &= \left(1+\frac{1}{\varepsilon}\right)\frac{1-F(\theta|p_L)}{\theta f(\theta|p_L)}\left\{1-\int_{\theta}^{\bar{\theta}}\alpha\left(\tilde{\theta}\right)\frac{f\left(\tilde{\theta}|p_L\right)}{1-F(\theta|p_L)}d\tilde{\theta}\right\} \\
&\quad -\left(1+\frac{1}{\varepsilon}\right)\frac{1-F(\theta|p_L)}{\theta f(\theta|p_L)}\gamma(p_L) \\
&\quad +\left(1+\frac{1}{\varepsilon}\right)\frac{1-F(\theta|p_L)}{\theta f(\theta|p_L)}\gamma(p_H)\left(\frac{1-F(\theta|p_H)}{1-F(\theta|p_L)}\right)\left(\frac{h(p_H)}{h(p_L)}\right)
\end{aligned}$$

Now assume that

$$\lim_{\theta\to\bar{\theta}}\left(\frac{1-F(\theta|p_L)}{1-F(\theta|p_H)}\right)=1$$

and also

$$\lim_{\theta\to\bar{\theta}}\left(\frac{1-F(\theta|p_H)}{\theta f(\theta|p_H)}\right)=\lim_{\theta\to\bar{\theta}}\left(\frac{1-F(\theta|p_L)}{\theta f(\theta|p_L)}\right)=\frac{1}{\rho}$$

that both distributions have a similar upper tail. Then finally we obtain

$$\begin{aligned}
\lim_{\theta\to\bar{\theta}}\left(\frac{\tau(\theta, p_H)}{1-\tau(\theta, p_H)}\right) &= \left(1+\frac{1}{\varepsilon}\right)\frac{1}{\rho}\int_{\underline{\theta}}^{\bar{\theta}}\alpha(\theta)\, f(\theta|p_H)\, d\theta \\
\lim_{\theta\to\bar{\theta}}\left(\frac{\tau(\theta, p_L)}{1-\tau(\theta, p_L)}\right) &= \left(1+\frac{1}{\varepsilon}\right)\frac{1}{\rho}\int_{\underline{\theta}}^{\bar{\theta}}\alpha(\theta)\, f(\theta|p_L)\, d\theta
\end{aligned}$$

B.1.3 Proof of Theorem 9

Proof. In this case we have

$$1-F(\theta|p_H)\propto\theta^{-\rho_H}$$

and

$$1 - F\left(\theta | p_H\right) \propto \theta^{-\rho_L}$$

at the upper tails. Then

$$\begin{aligned}
\lim_{\theta \to \infty} \frac{1 - F\left(\theta | p_L\right)}{1 - F\left(\theta | p_H\right)} &\propto \lim_{\theta \to \infty} \theta^{\rho_H - \rho_L} = 0 \\
\lim_{\theta \to \infty} \frac{1 - F\left(\theta | p_H\right)}{1 - F\left(\theta | p_L\right)} &= +\infty
\end{aligned}$$

Since $\gamma\left(p_H\right) > 0$, we obtain

$$\begin{aligned}
\lim_{\theta \to +\infty} \frac{\tau\left(\theta, p_H\right)}{1 - \tau\left(\theta, p_H\right)} &= \left(1 + \frac{1}{\varepsilon^*\left(p_H\right)}\right) \frac{1}{\rho_H} \left[1 - \gamma\left(p_H\right)\right] \\
\lim_{\theta \to +\infty} \frac{\tau\left(\theta, p_L\right)}{1 - \tau\left(\theta, p_L\right)} &= +\infty
\end{aligned}$$

We conclude that

$$\lim_{\theta \to +\infty} \tau\left(\theta, p_L\right) = 1 > \lim_{\theta \to +\infty} \tau\left(\theta, p_H\right)$$

□

Appendix C

Appendix to Chapter 3

C.1 Empirical Analysis with Demographic Controls

In the analysis of Section 3.2, we used raw data on labor income to gather information on inequality in labor income across occupations and educational levels. In this Section of the Appendix, we show that results are basically unchanged when we add a first stage in which we control for some observable characteristics. In particular, our estimates of the Pareto tails all remain significantly different across the two categories of occupations and the two different educational levels.

In Table C.1 we see that all Pareto tails parameters estimated are slightly higher than in the raw data. However, coefficients are still statistically different across occupation categories and educational levels, pointing to important differences at the upper tail of the labor income distribution across different occupations and educational levels. In Table C.2, we report the results on the first stage regression of log labor income of the head of household on standard demographic characteristics.

C.2 Proofs

The planner's objective is given by

Table C.1: Descriptive statistics of conditional labor income distribution (with demographic controls

Category	Pareto Tail	s.e.
Benchmark	2.26	0.07
Occupation		
Managerial/Professional	2.01	0.07
Others	3.25	0.19
Education		
College	2.02	0.07
No Finished College	3.33	0.15

Source: Author's computations using 2013 Survey of Consumer Finances. Standard errors (s.e.) computed using replicate weights provided in the original data and method proposed by [84] to adjust for imputation. The estimate for the Pareto tail is computed using the standard method proposed by [52] with a threshold of $100,000 for the upper tail. All statistics computed using labor income of head of household. Sample is restricted to head of households between 25 and 60 years old, as well as labor income of at least $1,000.

Table C.2: Liner Regression of Log-Wages on Demographics

log(Wage)	Coefficient	s.e.
(Intercept)	6.01	1.07
age	0.31	0.08
age^2	-0.01	0.00
age^3	0.00	0.00
African-American	-0.38	0.04
Hispanic	-0.54	0.03
Other race (not white)	0.07	0.07
Sex	-0.67	0.03

Source: Author's computations using 2013 Survey of Consumer Finances. Standard errors (s.e.) computed using replicate weights provided in the original data and method proposed by [84] to adjust for imputation. All statistics computed using labor income of head of household. Sample is restricted to head of households between 25 and 60 years old, as well as labor income of at least $1,000.

$$\begin{aligned}\mathcal{W}(\tau,T,s) = & \int_0^\infty G\left(U^a\left(\theta_a;\tau,T\right)\right) F\left(\theta_b^*\left(\theta_a;\tau,T,s\right)|\theta_a\right) f\left(\theta_a\right) d\theta_a \\ & + \int_0^\infty \int_{\theta_b^*(\theta_a;\tau,T,s)}^\infty G\left(U^b\left(\theta_b;\tau,T,s\right)\right) f\left(\theta_a,\theta_b\right) d\theta_b d\theta_a\end{aligned}$$

Therefore

$$\begin{aligned}\frac{\partial \mathcal{W}}{\partial s} = & \int_0^\infty G\left(U^a\left(\theta_a;\tau,T\right)\right) f\left(\theta_a,\theta_b^*\left(\theta_a;\tau,T,s\right)\right) \frac{\partial \theta_b^*\left(\theta_a;\tau,T,s\right)}{\partial s} d\theta_a \\ & + e\int_0^\infty \int_{\theta_b^*(\theta_a;\tau,T,s)}^\infty G'\left(U^b\left(\theta_b;\tau,T,s\right)\right) u_c^b\left(\theta_b\right) f\left(\theta_a,\theta_b\right) d\theta_b d\theta_a \\ & - \int_0^\infty G\left(U^b\left(\theta_b^*\left(\theta_a;\tau,s\right);\tau,T,s\right)\right) f\left(\theta_a,\theta_b^*\left(\theta_a;\tau,T,s\right)\right) \frac{\partial \theta_b^*\left(\theta_a;\tau,T,s\right)}{\partial s} d\theta_a\end{aligned}$$

where

$$u_c^b\left(\theta_b\right) = u_c\left(T + \left(1-\tau\right) y^b\left(\theta_b;\tau,T,s\right) - \left(1-s\right) e, \frac{y^b\left(\theta_b;\tau,T,s\right)}{\theta_b}\right)$$

is the marginal utility from consumption for an agent in occupation b.

By definition of $\theta_b^*\left(\theta_a;\tau,T,s\right)$ we have

$$U^a\left(\theta_a;\tau,T\right) = U^b\left(\theta_b^*\left(\theta_a;\tau,s\right);\tau,T,s\right)$$

Therefore, the marginal welfare change from the subsidy is simplified to

$$\frac{\partial \mathcal{W}\left(\tau,T,s\right)}{\partial s} = e\int_0^\infty \int_{\theta_b^*(\theta_a;\tau,T,s)}^\infty G'\left(U^b\left(\theta_b;\tau,T,s\right)\right) u_c^b\left(\theta_b\right) f\left(\theta_a,\theta_b\right) d\theta_b d\theta_a$$

On the other hand, the budget constraint is given byBut this is desirable by a redistributive

$$\tau Y\left(\tau,T,s\right) - T - s\cdot e\cdot E\left(\tau,T,s\right) = 0$$

where

$$\begin{aligned}Y\left(\tau,T,s\right) = & \int_{\theta_b \geq \theta_b^*(\theta_a;\tau,T,s)} y^b\left(\theta_b;\tau,T,s\right) dF\left(\theta_a,\theta_b\right) \\ & + \int_{\theta_b < \theta_b^*(\theta_a;\tau,T,s)} y^a\left(\theta_a;\tau,T\right) dF\left(\theta_a,\theta_b\right)\end{aligned}$$

and $E(\tau,T,s)=\int_{\theta_a}\int_{\theta_b\geq\theta_b^*(\theta_a;\tau,T,s)}dF(\theta_a,\theta_b)$.

The first order condition to the planner's problem on s can then be written as

$$\frac{\partial\mathcal{W}(\tau,T,s)}{\partial s}=\lambda\left[\left(E(\tau,T,s)+s\frac{\partial E(\tau,T,s)}{\partial s}\right)e-\tau\frac{\partial Y^*(\tau,T,s)}{\partial s}\right] \tag{C.1}$$

The derivative of $\mathcal{W}$ with respect to the lump-sum transfer is

$$\begin{aligned}\frac{\partial\mathcal{W}(\tau,T,s)}{\partial T}=&\int_0^\infty G'\left(U^a(\theta_a;\tau,T)\right)u_c^a(\theta_a)F\left(\theta_b^*(\theta_a;\tau,T,s)\,|\theta_a\right)f(\theta_a)\,d\theta_a\\&+\int_0^\infty\int_{\theta_b^*(\theta_a;\tau,T,s)}^\infty G'\left(U^b(\theta_b;\tau,T,s)\right)u_c^b(\theta_b)f(\theta_a,\theta_b)\,d\theta_b d\theta_a\end{aligned}$$

Let's denote this quantity by $\Omega(\tau,T,s)$, the marginal welfare from a uniform increase in consumption. Then FOC with respect to T is given by

$$\Omega(\tau,T,s)=\lambda\left(1-\tau\frac{\partial Y(\tau,T,s)}{\partial T}+s\cdot e\cdot\frac{\partial E(\tau,T,s)}{\partial T}\right)$$

C.2.1 Proof of Theorem (13)

Therefore from (C.1) we obtain

$$\begin{aligned}e\Omega_e(\tau,T,s)=&\frac{\Omega(\tau,T,s)}{1-\tau\frac{\partial Y^*(\tau,T,s)}{\partial T}+s\cdot e\cdot\frac{\partial E(\tau,T,s)}{\partial T}}E(\tau,T,s)\,e\\&+\frac{\Omega(\tau,T,s)}{1-\tau\frac{\partial Y^*(\tau,T,s)}{\partial T}+s\cdot e\cdot\frac{\partial E(\tau,T,s)}{\partial T}}s\frac{\partial E(\tau,T,s)}{\partial s}e\\&-\frac{\Omega(\tau,T,s)}{1-\tau\frac{\partial Y^*(\tau,T,s)}{\partial T}+s\cdot e\cdot\frac{\partial E(\tau,T,s)}{\partial T}}\tau\frac{\partial Y(\tau,T,s)}{\partial s}\end{aligned}$$

Therefore at the optimum

$$\begin{aligned}\frac{s}{1-s} = & \frac{1}{\xi^E_{1-s}(\tau,T,s)} - \frac{1}{\xi^E_{1-s}(\tau,T,s)} \frac{\frac{\Omega_e(\tau,T,s)}{E(\tau,T,s)}}{\Omega(\tau,T,s)} \left(1 - \tau Y(\tau,T,s)\, \eta^Y_T(\tau,T,s)\right) \\ & - \frac{1}{\xi^E_{1-s}(\tau,T,s)} \frac{\frac{\Omega_e(\tau,T,s)}{E(\tau,T,s)}}{\Omega(\tau,T,s)} s \cdot e \cdot E(\tau,T,s)\, \eta^E_T(\tau,T,s) \\ & + \left(\frac{\tau}{1-s}\right)\left(\frac{1}{R(\tau,T,s)}\right) \frac{\xi^Y_{1-s}(\tau,T,s)}{\xi^E_{1-s}(\tau,T,s)}\end{aligned}$$

where

$$\begin{aligned}\xi^E_s(\tau,T,s) &= \frac{\partial E(\tau,T,s)}{\partial(1-s)} \frac{1-s}{E(\tau,T,s)} \\ \xi^Y_s(\tau,T,s) &= \frac{1-s}{Y^*(\tau,s)} \frac{\partial Y(\tau,T,s)}{\partial(1-s)} \\ R(\tau,T,s) &= \frac{eE(\tau,T,s)}{Y(\tau,T,s)} \\ \eta^Y_T(\tau,T,s) &= \frac{1}{Y(\tau,T,s)} \frac{\partial Y(\tau,T,s)}{\partial T} \\ \eta^E_T(\tau,T,s) &= \frac{1}{E(\tau,T,s)} \frac{\partial E(\tau,T,s)}{\partial T}\end{aligned}$$

and $R(\tau,s)$ accounts for educational expenses as a share of total output.

C.2.2 Optimal Income Tax

The welfare effect of a change in τ is given by

$$\frac{\partial \mathcal{W}(\tau,T,s)}{\partial \tau} = -\Omega^Y_\tau$$

Therefore the first order condition to the government's problem with respect to τ is given by

$$\Omega^Y_\tau = \lambda \left[Y(\tau,T,s) + \tau \frac{\partial Y(\tau,T,s)}{\partial \tau} - se \frac{\partial E(\tau,T,s)}{\partial \tau} \right]$$

Hence

$$\frac{\tau}{1-\tau} = \frac{1}{\xi^Y_{1-\tau}}\left[1 + \frac{s}{1-\tau}R\xi^E_{1-\tau} - \frac{\Omega^Y_\tau}{\lambda Y(\tau, T, s)}\right]$$

Since $\lambda = \Omega/\omega$ we get the result.

C.2.3 Optimal Educational Subsidies with Non-Linear Labor Income Taxes

The only term that changes as compared to the previous analysis is the effect on government labor income tax revenue. Revenue from labor income taxes is given by

$$\begin{aligned}\mathcal{I}(s) &= \int_0^\infty \int_0^\infty t\left(y\left(\theta_a, \theta_b; s\right)\right) f\left(\theta_a, \theta_b\right) d\theta_b d\theta_a \\ &= \int_0^\infty \int_{\theta_b^*(\theta_a;s)}^\infty t\left(y^b\left(\theta_b; s\right)\right) f\left(\theta_a, \theta_b\right) d\theta_b d\theta_a \\ &+ \int_0^\infty t\left(y^a\left(\theta_a\right)\right)\left(\int_0^{\theta_b^*(\theta_a;s)} f\left(\theta_a, \theta_b\right) d\theta_b\right) d\theta_a\end{aligned}$$

Therefore

$$\begin{aligned}\frac{\partial \mathcal{I}(s)}{\partial s} &= \int_0^\infty \left[t\left(y^a\left(\theta_a\right)\right) - t\left(y^b\left(\theta_b^*\left(\theta_a; s\right); s\right)\right)\right] f\left(\theta_a, \theta_b^*\left(\theta_a; s\right)\right) \frac{\partial \theta_b^*\left(\theta_a; s\right)}{\partial s} d\theta_a \\ &+ \int_0^\infty \int_{\theta_b^*(\theta_a;s)}^\infty t'\left(y^b\left(\theta_b; s\right)\right) \frac{\partial y^b\left(\theta_b; s\right)}{\partial s} f\left(\theta_a, \theta_b\right) d\theta_b d\theta_a\end{aligned}$$

The first term captures the effect on government revenue from the change in educational subsidies through the extensive margin channel. As more workers switch into occupation b, their labor income jumps to a new level, what results in a different amount of taxes paid. The second component captures changes in the intensive margin on labor supply due to the change in educational subsidies. Those arise from income effects on labor supply of the change in subsidies for workers that would obtain a high level education degree even before the policy change.

Proceeding in an analogous fashion as before, we have the first order condition

$$e \cdot \Omega_e \cdot E(s) = \lambda \cdot \left[e \cdot E(s) + s \cdot e \cdot \frac{\partial E(s)}{\partial s} - \frac{\partial \mathcal{I}}{\partial s} \right]$$

Rewriting this equation in terms of elasticities we obtain

$$\frac{s}{1-s} = \frac{1}{\xi^E_{1-s}} \left(1 - \frac{\Omega_e}{\lambda} \right) + \left(\frac{\mathcal{I}}{Y(s)} \right) \frac{\xi^{\mathcal{I}}_{1-s}}{(1-s) R \xi^E_{1-s}}$$

Bibliography

[1] Daron Acemoglu and David Autor. Skills, tasks and technologies: Implications for employment and earnings. *Handbook of labor economics*, 4:1043–1171, 2011.

[2] Mark Aguiar and Manuel Amador. Growth in the shadow of expropriation*. *Quarterly Journal of Economics*, 126(2), 2011.

[3] Sule Alan and Martin Browning. Estimating intertemporal allocation parameters using synthetic residual estimation. *The Review of Economic Studies*, 77(4):1231–1261, 2010.

[4] Sule Alan, Martin Browning, and Mette Ejrnaes. Income and consumption: a micro semi-structural analysis with pervasive heterogeneity1. *Available at SSRN 2566477*, 2014.

[5] Manuel Amador, Ivan Werning, and George-Marios Angeletos. Commitment vs. flexibility. *Econometrica*, 74(2):365–396, 2006.

[6] Dan Anderberg. Optimal policy and the risk properties of human capital reconsidered. *Journal of Public Economics*, 93(9):1017–1026, 2009.

[7] Mark Armstrong. Multiproduct nonlinear pricing. *Econometrica: Journal of the Econometric Society*, pages 51–75, 1996.

[8] Mark Armstrong and Jean-Charles Rochet. Multi-dimensional screening:: A user's guide. *European Economic Review*, 43(4):959–979, 1999.

[9] Anthony B Atkinson, Thomas Piketty, and Emmanuel Saez. Top incomes in the long run of history. *Journal of Economic Literature*, 49(1):3–71, 2011.

[10] Anthony B Atkinson and Joseph E Stiglitz. The structure of indirect taxation and economic efficiency. *Journal of Public Economics*, 1(1):97–119, 1972.

[11] Ned Augenblick, Muriel Niederle, and Charles Sprenger. Working over time: Dynamic inconsistency in real effort tasks*. *The Quarterly Journal of Economics*, page qjv020, 2015.

[12] David H Autor, Lawrence F Katz, and Melissa S Kearney. Trends in us wage inequality: Revising the revisionists. *The Review of economics and statistics*, 90(2):300–323, 2008.

[13] David H Autor, Frank Levy, and Richard J Murnane. The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, pages 1279–1333, 2003.

[14] Alejandro Badel and Mark Huggett. Taxing top earners: a human capital perspective. *Federal Reserve Bank of St. Louis Working Paper Series*, (2014-017), 2014.

[15] Marco Battaglini and Rohit Lamba. Optimal dynamic contracting. *Economic Theory Center Working Paper*, (46-2012), 2015.

[16] Roland Benabou. Unequal societies: Income distribution and the social contract. *American Economic Review*, pages 96–129, 2000.

[17] Roland Benabou. Tax and education policy in a heterogeneous-agent economy: What levels of redistribution maximize growth and efficiency? *Econometrica*, 70(2):481–517, 2002.

[18] John Beshears, James J Choi, Christopher Harris, David Laibson, Brigitte C Madrian, and Jung Sakong. Self control and commitment: Can decreasing the liquidity of a savings account increase deposits? Technical report, National Bureau of Economic Research, 2015.

[19] Radim Bohacek and Marek Kapicka. Optimal human capital policies. *Journal of Monetary Economics*, 55(1):1–16, 2008.

[20] A Lans Bovenberg and Bas Jacobs. Redistribution and education subsidies are siamese twins. *Journal of Public Economics*, 89(11), 2005.

[21] Raj Chetty, Adam Guren, Day Manoli, and Andrea Weber. Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins. *The American Economic Review*, 101(3):471–475, 2011.

[22] Stephen Coate. Altruism, the samaritan's dilemma, and government transfer policy. *The American Economic Review*, pages 46–57, 1995.

[23] Jacques Cremer and Richard P McLean. Full extraction of the surplus in bayesian and dominant strategy auctions. *Econometrica: Journal of the Econometric Society*, pages 1247–1257, 1988.

[24] Carlos E Da Costa and Lucas J Maestri. The risk properties of human capital and the design of government policies. *European Economic Review*, 51(3):695–713, 2007.

[25] Peter Diamond and Johannes Spinnewijn. Capital income taxes with heterogeneous discount rates. *American Economic Journal: Economic Policy*, 3(4):52–76, 2011.

[26] Peter A Diamond. A framework for social security analysis. *Journal of Public Economics*, 8(3):275–298, 1977.

[27] Peter A Diamond. Optimal income taxation: an example with a u-shaped pattern of optimal marginal tax rates. *American Economic Review*, pages 83–95, 1998.

[28] Eric M Engen, Eric M Engen, and Eric M Engen. Lifetime earnings, social security benefits, and the adequacy of retirement wealth accumulation. *Soc. Sec. Bull.*, 66:38, 2005.

[29] Financial Engines. Missing out:how much employer 401(k) matching contributions do employees leave on the table? Technical report, Financial Engines, 2015.

[30] Emmanuel Farhi and Xavier Gabaix. Optimal taxation with behavioral agents. Technical report, National Bureau of Economic Research, 2015.

[31] Emmanuel Farhi and Iván Werning. Progressive estate taxation. *The Quarterly Journal of Economics*, 125(2):635–673, 2010.

[32] Emmanuel Farhi and Iván Werning. Insurance and taxation over the life cycle. *The Review of Economic Studies*, 80(2):596–635, 2013.

[33] Martin Feldstein. The optimal level of social security benefits. *The Quarterly Journal of Economics*, 100(2):303–320, 1985.

[34] Martin S Feldstein. The effects of taxation on risk taking. *The Journal of Political Economy*, pages 755–764, 1969.

[35] Ana Fernandes and Christopher Phelan. A recursive formulation for repeated agency with history dependence. *Journal of Economic Theory*, 91(2):223–247, 2000.

[36] Sergio Firpo, Nicole M Fortin, and Thomas Lemieux. Occupational tasks and changes in the wage structure. 2011.

[37] Simone Galperti. Commitment, flexibility, and optimal screening of time inconsistency. *Econometrica*, 83(4):1425–1465, 2015.

[38] Mikhail Golosov, Narayana Kocherlakota, and Aleh Tsyvinski. Optimal indirect and capital taxation. *Review of Economic studies*, pages 569–587, 2003.

[39] Mikhail Golosov, Maxim Troshkin, and Aleh Tsyvinski. Redistribution and social insurance. *American Economic Review*, Forthcoming.

[40] Mikhail Golosov, Maxim Troshkin, and Aleh Tsyvinski. Redistribution and social insurance. *American Economic Review*, Forthcoming.

[41] Mikhail Golosov, Maxim Troshkin, Aleh Tsyvinski, and Matthew Weinzierl. Preference heterogeneity and optimal capital income taxation. *Journal of Public Economics*, 97:160–175, 2013.

[42] Maarten Goos and Alan Manning. Lousy and lovely jobs: The rising polarization of work in britain. *The review of economics and statistics*, 89(1):118–133, 2007.

[43] Pierre-Olivier Gourinchas and Jonathan A Parker. Consumption over the life cycle. *Econometrica*, 70(1):47–89, 2002.

[44] Borys Grochulski and Tomasz Piskorski. Risky human capital and deferred capital income taxation. *Journal of Economic Theory*, 145(3):908–943, 2010.

[45] Jonathan Gruber and Botond Kőszegi. Tax incidence when individuals are time-inconsistent: the case of cigarette excise taxes. *Journal of Public Economics*, 88(9):1959–1987, 2004.

[46] Fatih Guvenen. Learning your earning: Are labor income shocks really very persistent? *The American economic review*, pages 687–712, 2007.

[47] Fatih Guvenen. An empirical investigation of labor income processes. *Review of Economic Dynamics*, 12(1):58–79, 2009.

[48] Fatih Guvenen, Fatih Karahan, Serdar Ozkan, and Jae Song. What do data on millions of us workers reveal about life-cycle earnings risk? Technical report, National Bureau of Economic Research, 2015.

[49] Fatih Guvenen and Anthony A Smith. Inferring labor income risk and partial insurance from economic choices. *Econometrica*, 82(6):2085–2129, 2014.

[50] Jonathan Heathcote, Kjetil Storesletten, and Giovanni L Violante. Optimal tax progressivity: An analytical framework. Technical report, National Bureau of Economic Research, 2014.

[51] Jonathan Heathcote and Hitoshi Tsujiyama. Optimal income taxation: Mirrlees meets ramsey. 2015.

[52] Bruce M Hill et al. A simple general approach to inference about the tail of a distribution. *The annals of statistics*, 3(5):1163–1174, 1975.

[53] Harold M Hochman and James D Rodgers. Pareto optimal redistribution. *The American Economic Review*, pages 542–557, 1969.

[54] Mark Huggett, Gustavo Ventura, and Amir Yaron. Sources of lifetime inequality. *The American Economic Review*, 101(7):2923–2954, 2011.

[55] Bas Jacobs and A Bovenberg. Optimal taxation of human capital and the earnings function. *Journal of Public Economic Theory*, 13(6):957–971, 2011.

[56] Bas Jacobs and A Lans Bovenberg. Human capital and optimal positive taxation of capital income. *International Tax and Public Finance*, 17(5):451–478, 2010.

[57] Damon Jones and Aprajit Mahajan. Time-inconsistency and saving: Experimental evidence from low-income tax filers. Working Paper 21272, National Bureau of Economic Research, June 2015.

[58] Kenneth Judd and Che-Lin Su. Optimal income taxation with multidimensional taxpayer types. In *Computing in Economics and Finance*, volume 471, 2006.

[59] Gueorgui Kambourov and Iourii Manovskii. Occupational mobility and wage inequality. *The Review of Economic Studies*, 76(2):731–759, 2009.

[60] Marek Kapička. Efficient allocations in dynamic private information economies with persistent shocks: A first-order approach. *The Review of Economic Studies*, page rds045, 2013.

[61] Marek Kapicka. Optimal mirrleesean taxation with unobservable human capital formation. *University of California, Santa Barbara WP*, 2013.

[62] Marek Kapicka and Julian Neira. Optimal taxation with risky human capital. *CERGE-EI Working Paper Series*, (553), 2015.

[63] Henrik Jacobsen Kleven, Claus Thustrup Kreiner, and Emmanuel Saez. The optimal income taxation of couples. *Econometrica*, 77(2):537–560, 2009.

[64] Laurence J Kotlikoff, Avia Spivak, and Lawrence H Summers. The adequacy of savings. *American Economic Review*, 72(5):1056–69, 1982.

[65] Dirk Krueger and Alexander Ludwig. Optimal progressive labor income taxation and education subsidies when education decisions and intergenerational transfers are endogenous. *The American Economic Review*, 103(3):496–501, 2013.

[66] David Laibson. Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2):443–478, 1997.

[67] David Laibson, Andrea Repetto, and Jeremy Tobacman. Estimating discount functions with consumption choices over the lifecycle. Working Paper 13314, National Bureau of Economic Research, August 2007.

[68] Lee A Lillard and Yoram Weiss. Components of variation in panel earnings data: American scientists 1960-70. *Econometrica: Journal of the Econometric Society*, pages 437–454, 1979.

[69] Ben Lockwood and Dmitry Taubinsky. Regressive sin taxes. Technical report, Working Paper, 2015.

[70] Thomas E MaCurdy. The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of econometrics*, 18(1):83–114, 1982.

[71] Darío Maldonado. Education policies and optimal taxation. *International Tax and Public Finance*, 15(2):131–143, 2008.

[72] N Gregory Mankiw and Matthew Weinzierl. The optimal taxation of height: A case study of utilitarian income redistribution. *American Economic Journal. Economic Policy*, 2(1):155, 2010.

[73] R Preston McAfee and John McMillan. Multidimensional incentive compatibility and mechanism design. *Journal of Economic Theory*, 46(2):335–354, 1988.

[74] Paul Milgrom and Ilya Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, pages 583–601, 2002.

[75] James A Mirrlees. An exploration in the theory of optimum income taxation. *The review of economic studies*, pages 175–208, 1971.

[76] José Luis Montiel Olea and Tomasz Strzalecki. Axiomatization and measurement of quasi-hyperbolic discounting. *The Quarterly Journal of Economics*, 129(3):1449–1499, 2014.

[77] Sendhil Mullainathan, Joshua Schwartzstein, and William J Congdon. A reduced-form approach to behavioral public finance. *Annu. Rev. Econ*, 4:17–1, 2012.

[78] Ted O'Donoghue and Matthew Rabin. Studying optimal paternalism, illustrated by a model of sin taxes. *American Economic Review*, pages 186–191, 2003.

[79] Ted O'Donoghue and Matthew Rabin. Optimal sin taxes. *Journal of Public Economics*, 90(10):1825–1849, 2006.

[80] Thomas Piketty and Emmanuel Saez. Income inequality in the united states, 1913–1998. *The Quarterly Journal of Economics*, 118(1):1–41, 2003.

[81] Jean-Charles Rochet and Philippe Choné. Ironing, sweeping, and multidimensional screening. *Econometrica*, pages 783–826, 1998.

[82] Casey Rothschild and Florian Scheuer. Redistributive Taxation in the Roy Model. *The Quarterly Journal of Economics*, 128(2):623–668, 2013.

[83] Andrew Donald Roy. Some thoughts on the distribution of earnings. *Oxford economic papers*, 3(2):135–146, 1951.

[84] Donald B Rubin. The calculation of posterior distributions by data augmentation: Comment: A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The sir algorithm. *Journal of the American Statistical Association*, 82(398):543–546, 1987.

[85] Dominik Sachs, Sebastian Findeisen, et al. Education and optimal dynamic taxation. In *2012 Meeting Papers*, number 365. Society for Economic Dynamics, 2012.

[86] Emmanuel Saez. Using elasticities to derive optimal income tax rates. *Review of economic studies*, 68(1):205, 2001.

[87] Emmanuel Saez. Using elasticities to derive optimal income tax rates. *Review of Economic Studies*, 68(1):205–29, 2001.

[88] Emmanuel Saez. The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes. *Journal of Public Economics*, 83(2):217 – 230, 2002.

[89] Stefanie Stantcheva. Optimal taxation and human capital policies over the life cycle. Technical report, National Bureau of Economic Research, 2015.

[90] Tomomi Tanaka, Colin F Camerer, and Quang Nguyen. Risk and time preferences: linking experimental and household survey data from vietnam. *The American Economic Review*, 100(1):557–571, 2010.

[91] Pei Cheng Yu. Optimal taxation with time-inconsistent agents. Technical report, University of Minessota - Working Paper, 2015.