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

Christian Moser
Princeton University

Pedro Olea de Souza e Silva
Princeton University

JOB MARKET PAPER

December 7, 2015

CLICK HERE FOR LAST VERSION
http://scholar.princeton.edu/pedro-olea

Abstract

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, independent 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.

*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, the University of São Paulo Economics Department, and the 2014 Conference of German Economists Abroad at Kiel IfW.

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 (Diamond (1977), Kotlikoff et al. (1982) and Feldstein (1985)). 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.[1] 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 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.[2] 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

[1] See Tanaka et al. (2010), Montiel Olea and Strzalecki (2014), Augenblick et al. (2015), Jones and Mahajan (2015) and Beshears et al. (2015) for recent findings both on the presence of behavioral biases and its considerable heterogeneity. Piketty and Saez (2003) and Atkinson et al. (2011) document the increase in labor earnings and wealth inequality in the United States.

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

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 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.[3] 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.[4]

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 (Judd and Su (2006)). 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 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 Heathcote et al. (2014) 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.[5] 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

[3] In Appendix E, 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.

[4] See Alan and Browning (2010) and Alan et al. (2014) for heterogeneity in discount factors. Montiel Olea and Strzalecki (2014), Augenblick et al. (2015) and Beshears et al. (2015) using experimental evidence find also considerable heterogeneity in present-bias.

[5] See Heathcote et al. (2014) and Heathcote and Tsujiyama (2015).

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 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 5 we extend the approach in Mullainathan et al. (2012) 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 Aguiar and Amador (2011). 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 Laibson (1997), starting with O'Donoghue and Rabin (2003), Gruber and Kőszegi (2004) and O'Donoghue and Rabin (2006), 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, Farhi and Gabaix (2015) 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. Lockwood and Taubinsky (2015) 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, Farhi and Werning (2010) consider a model with a single level of present-bias.[6] 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. Amador et al. (2006) 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, Galperti (2015) 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, Yu (2015) 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 Cremer and McLean (1988). 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 C, 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. Atkinson and Stiglitz (1972) show that without

[6] 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.

behavioral biases and uncertainty on earnings ability, the government should not distort savings decisions for redistribution purposes. Saez (2002) and more recently Diamond and Spinnewijn (2011) and Golosov et al. (2013) 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. Armstrong and Rochet (1999) emphasize that the lack of a general toolbox for the analysis of contracting problems with multi-dimensional heterogeneity hindered their usage in applications.[7] 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 2 develops a benchmark model for the joint analysis of paternalistic and redistributive policies. Then Section 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 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 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 6 concludes the paper.

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

[7] See McAfee and McMillan (1988), Armstrong (1996) and Rochet and Choné (1998) for important contributions to the mechanism design literature with multi-dimensional types. However, there are important exceptions in public finance. Kleven et al. (2009) analyzes the taxation of couples and find that a higher earnings by a secondary earner generally reduce optimal taxes on the primary earner. Rothschild and Scheuer (2013) characterizes optimal income taxes when there is heterogeneity in earnings skills across different occupations.

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.[8] 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 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*

[8] Amador et al. (2006) 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 B.

compatible if

$$U\left(c_1\left(\theta,\beta\right),c_2\left(\theta,\beta\right),y\left(\theta,\beta\right);\theta,\beta\right)\geq U\left(c_1\left(\theta',\beta'\right),c_2\left(\theta',\beta'\right),y\left(\theta',\beta'\right);\theta,\beta\right) \tag{1}$$

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:

$$\sum_{(\theta,\beta)\in\Theta\times B}\pi\left(\theta,\beta\right)\left[y\left(\theta,\beta\right)-c_1\left(\theta,\beta\right)-\frac{c_2\left(\theta,\beta\right)}{R}\right]\geq 0 \tag{2}$$

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

The government evaluates individual payoffs according to

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

Therefore the government objective can be written as

$$W\left(\left\{c_1\left(\theta,\beta\right),c_2\left(\theta,\beta\right),y\left(\theta,\beta\right)\right\}_{(\theta,\beta)\in\Theta\times B}\right)=\sum_{(\theta,\beta)\in\Theta\times B}\pi\left(\theta,\beta\right)\lambda\left(\theta\right)V\left(\theta,\beta\right) \tag{3}$$

where $\lambda\left(\theta\right)\geq 0$ 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.[9] Without loss of generality, Pareto weights are normalized so that $\sum_{(\theta,\beta)\in\Theta\times B}\pi\left(\theta,\beta\right)\lambda\left(\theta\right)=1$. The government problem is then to choose an allocation that maximizes (3) subject to both incentive constraints in (1) and the resource constraint in (2).

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.,

[9] Using the language of Laibson (1997), 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.

$\lambda(\theta_L) = \lambda(\theta_H) = 1$.

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 ability types into the allocations of low ability types:

$$u(c_1(\theta_H, \beta)) - \frac{1}{\theta_H} v(y_1(\theta_H, \beta)) + \beta\delta u(c_2(\theta_H, \beta)) \geq u(c_1(\theta_L, \beta')) + \beta\delta u(c_2(\theta_L, \beta')) \quad (4)$$

for $\beta, \beta' \in \{\beta_L, 1\}$. Since incentive constraint (4) is linear in utility levels $u(c_1(\theta_L, \beta'))$ and $u(c_2(\theta_L, \beta'))$, then a convex combination of those utility levels satisfy incentive compatibility

$$u(c_1(\theta_H, \beta)) - \frac{1}{\theta_H} v(y_1(\theta_H, \beta)) + \beta\delta u(c_2(\theta_H, \beta)) \geq \bar{u}_1(\theta_L) + \beta\delta \bar{u}_2(\theta_L) \quad (5)$$

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

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

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, $c_1(\theta_H) > c_2(\theta_H)$. In Figure 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.[10] 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.

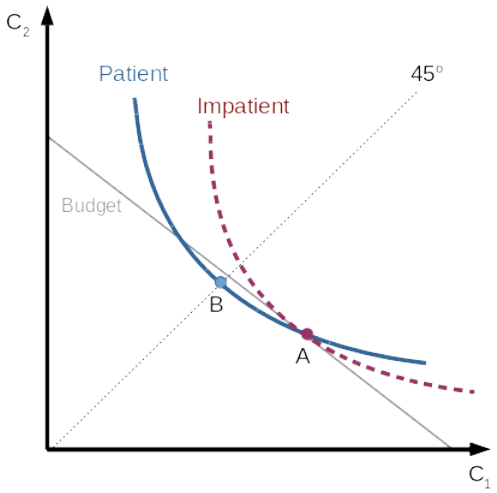


Figure 1. Incentives with bunching.

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 2 by point D. In this case, 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 ei-

[10]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.

ther $c_1(\theta_H) > c_1(\theta_L)$ 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.

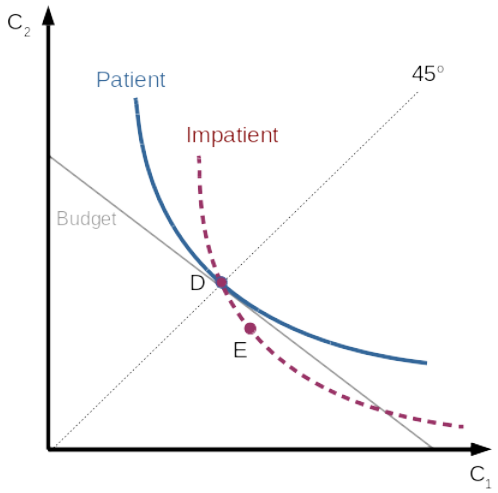


Figure 2. Separation at high ability

Distortions at low ability – We already know that low ability agents are bunched. Now, assume that point A in Figure 1 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 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, 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.

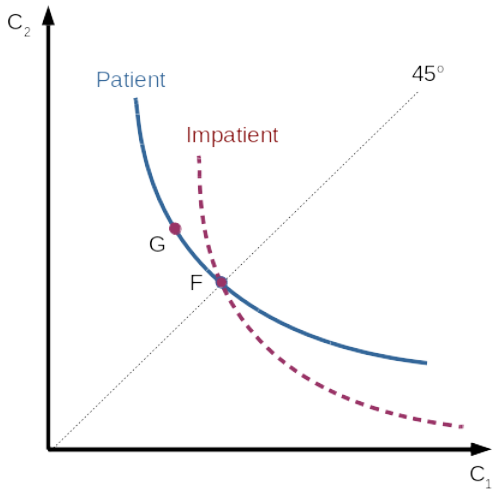


Figure 3. Over consumption at retirement of low ability relaxes incentives of impatient agents with high ability.

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.

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

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

At the solution to the planner's problem we must have $c_2\left(\theta_H,\beta_H\right)>c_2\left(\theta_L\right)$.[11] Therefore we conclude

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

But since high ability agents are separated, we also must have $c_2\left(\theta_H,\beta_H\right)>c_2\left(\theta_H,\beta_L\right)$ 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 $c_2\left(\theta_H,\beta_L\right)>c_2\left(\theta_L\right)$ so that

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

where we used our last result that the incentive constraint of $\left(\theta_H,\beta_L\right)$ 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 $\lambda\left(\theta_H\right)\approx 0$, then $c_2\left(\theta_H,\beta_L\right)\approx 0<c_2\left(\theta_L\right)$, and the incentive constraint of type $\left(\theta_H,\beta_H\right)$ with respect to low ability types will be binding. With many types, the characterization of which incentive constraints bind becomes an intractable problem. However, it is still possible to provide key implications for optimal policy even without this characterization.

[11] Assume by way of contradiction that $c_2\left(\theta_L\right)\geq c_2\left(\theta_H,\beta_H\right)$. Since there is separation of high ability agents, then $c_2\left(\theta_L\right)\geq c_2\left(\theta_H,\beta_H\right)>c_2\left(\theta_H,\beta_L\right)$. 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 $c_2\left(\theta_H,\beta_H\right)>c_2\left(\theta_L\right)$.

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 Mirrlees (1971), 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 Armstrong (1996); Rochet and Choné (1998) and Rochet and Choné (1998). 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.[12] 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 $\beta \in B$;

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.

[12] In section 4 we find through numerical simulations that those properties are not particular to the endpoints of the distribution of types, but that they are in fact a force present at all earnings levels.

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 B, 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:*

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

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

2. *for types with higher earnings ability:*

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

   2.2 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.

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 Atkinson and Stiglitz (1972). 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 Farhi and Werning (2010) 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 Atkinson and Stiglitz (1972) 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 Farhi and Werning (2010), 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 $c_1(\theta_H)$ in exchange for a reduction of $c_2(\theta_H)$ 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 Atkinson and Stiglitz (1972) and in Farhi and Werning (2010) are present in our model. As we can inspect in Theorem 2, the intertemporal distortions display a monotonic shape as in Farhi and Werning (2010). 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 Farhi and Werning (2010) while not distorting some agents without disagreement as in Atkinson and Stiglitz (1972), but it will also bunch them across disagreement types at low earnings ability.

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.[13]

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 the earnings level of individuals in the decentralization. In the U.S. 401(k) accounts are characterized by a

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

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

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

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.

Corollary 1. *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 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.

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.

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 Engen et al. (2005) on the ratio of wealth at retirement to life-time earnings of households.[14] This statistic is informative about how people save controlling for their life-time labor earnings, the essential

[14]The wealth measure used in Engen et al. (2005) 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 types of retirement accounts and estimates of defined contribution pensions benefits (excluding social security).

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 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.

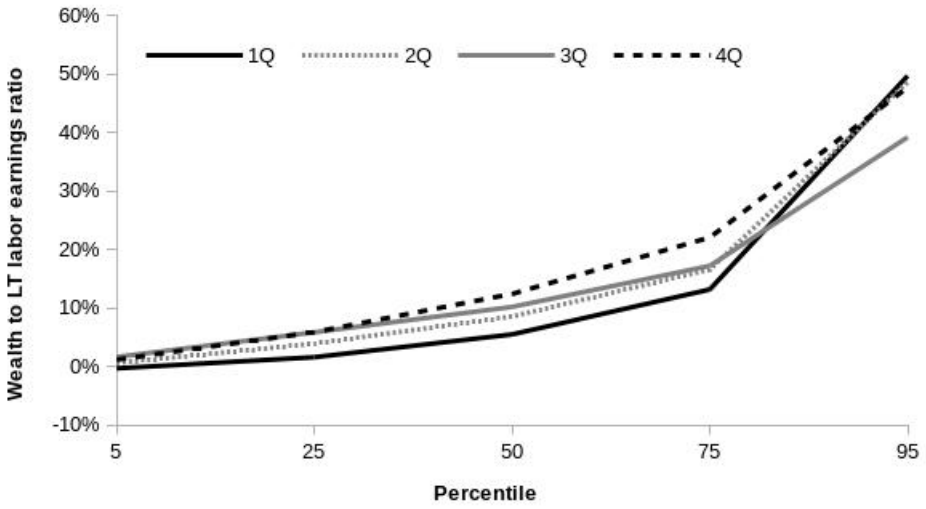


Figure 4. Distribution of the ratio of non social security wealth at retirement to life-time earnings. Each line represents the distribution within a different lifetime earnings quantile. Data from Engen et al. (2005) using 1992 Health and Retirement Survey sample of households. Non social security household 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.

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

Using the history of current and past reported earnings and estimates from Khitatrakun, Kitamura, and Scholz (2000), Engen et al. (2005) obtain a measure of lifetime household earnings.

δ in the two-period model calibration.[15] 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.[16] 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 Heathcote et al. (2014) and Heathcote and Tsujiyama (2015) in using the actual observed labor income distribution. Heathcote et al. (2014) 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.[17]

For the marginal distribution of ψ we assume a Beta(a, b) distribution with shape parameters $a > 0$ and $b > 0$.[18] 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.[19]

In modeling the tax-transfer system, we follow Benabou (2000), Benabou (2002), Heathcote et al. (2014)and Heathcote and Tsujiyama (2015) by using a parsimonious functional form calibrated to the level and progressivity of the net tax schedule that was introduced by Feldstein (1969). In this formulation, we write net transfers T to individuals as function of taxable income Y :

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

where taxable income include both labor earnings and taxable asset income. Estimating

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

[16] In Appendix E 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.

[17] 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.

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

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

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

the functional parameters off PSID data for 2002–2006, Heathcote et al. (2014) 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 through employers.[20] 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 Gourinchas and Parker (2002).

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 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.

Statistic\Quantile	Data Q1	Model Q1	Data Q2	Model Q2	Data Q3	Model Q3	Data Q4	Model Q4
W/Y percentile 25%	0.0165	0.0163	0.0398	0.0367	0.059	0.0451	0.0593	0.653
W/Y percentile 50%	0.0554	0.0899	0.086	0.1021	0.1024	0.1083	0.1248	0.1254
W/Y percentile 75%	0.1322	0.1431	0.1664	0.1712	0.1726	0.1765	0.2211	0.1839

Table 1. Calibration fit by life time earnings quantile. 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 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 Alan et al. (2014). 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 Alan et al. (2014). 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.

[20] This is the most common matching rate in the U.S. (Engines (2015))

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

Table 2. Calibration results for 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. Alan et al. (2014) 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 Tanaka et al. (2010) found no evidence of a correlation between present-bias and income levels in their experimental setting, Mullainathan et al. (2012) argues that different types of behavioral biases are more prevalent at the poor (e.g., addiction to cigarettes).

In Appendix E 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 Montiel Olea and Strzalecki (2014) 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. Laibson et al. (2007) find an estimate of 0.7 by calibrating a life-cycle model with hyperbolic preferences. More recently, Augenblick et al. (2015) finds that on real effort tasks present bias is 0.88 in a laboratory experiment. Jones and Mahajan (2015) find a present bias of 0.34 in a field experiment providing commitment devices incentivizing savings of income tax returns in the United

States.

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 Heathcote and Tsujiyama (2015) 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.[21]

4.2.1 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. Judd and Su (2006) 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

[21] See Heathcote et al. (2014) and Heathcote and Tsujiyama (2015).

subsets of incentive constraints.[22] 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.[23] 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.

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.[24] 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.[25]

4.2.2 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.[26] In doing so, we follow the approach

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

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

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

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

[26] 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.

in Heathcote et al. (2014); Heathcote and Tsujiyama (2015) 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.[27] 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.[28] 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.[29]

In Figure 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 5b we plot the proportion of retirement consumption at the optimal allocation for a utilitarian planner with the same level of discounting.

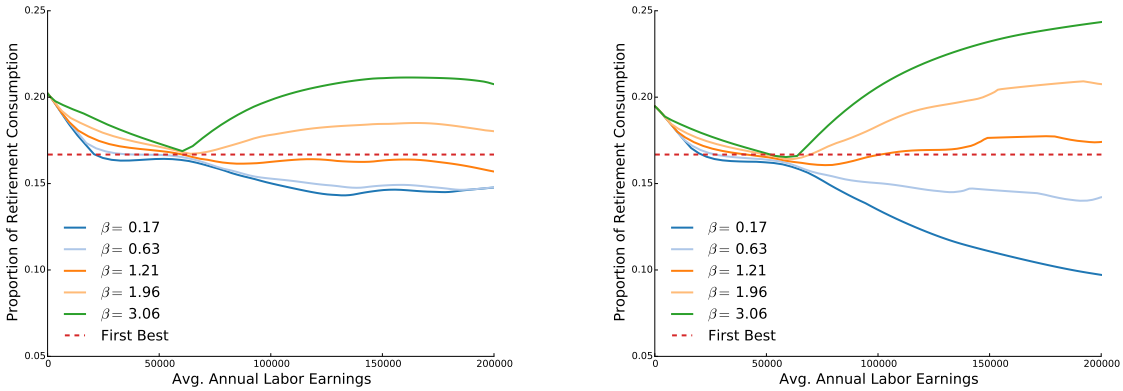
The level of dispersion at low earnings level is very similar to the benchmark normative model, however there is considerably more dispersion at high earnings. A utilitarian planner is considerably more redistributive than the planner in the benchmark normative model, therefore it will offer considerably 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 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 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 individuals have very little access to these accounts. This points in the direction that optimal policy can be well approximated with simple policy tools.

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

[28] Montiel Olea and Strzalecki (2014) indeed find a considerable level of forward bias present in their experiment.

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



(a) Benchmark normative model

(b) Utilitarian planner (same discounting as benchmark)

Figure 5. 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.

In Figure 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 8 we compare the current system of old-age benefits by earnings level with the normative model implication under the 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 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

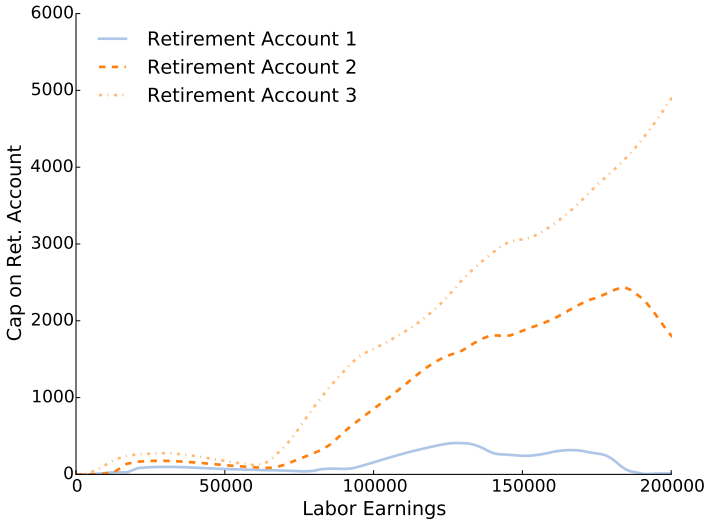


Figure 6. Caps on retirement savings accounts on normative model that better approximate actual policy. On the horizontal axis we have average labor earnings of the individual during its working life and in the vertical axis we have the cap on the contribution to each of the 3 different retirement savings accounts.

the difficulty in reconciling current social security benefits with current redistribution levels in the lenses of our normative model.

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

Table 3. 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.

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: 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).

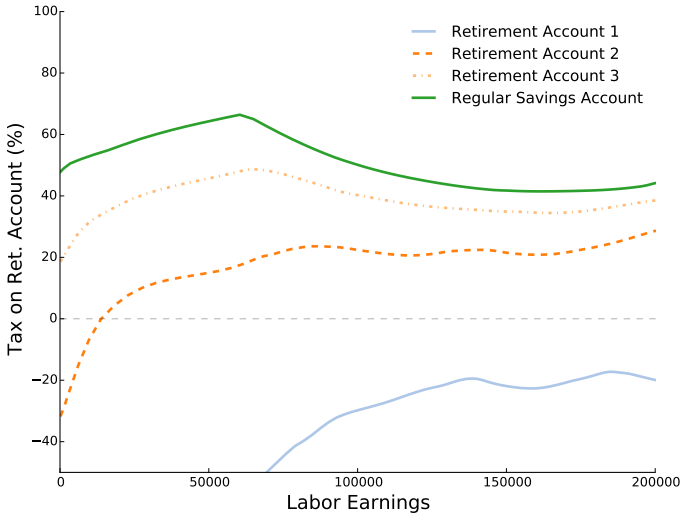


Figure 7. Tax rates on savings. On the horizontal axis we have average labor earnings of the individual during its working life and in the vertical axis we have the tax rate on savings. Each line represents a different retirement savings account or the regular savings account.

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.

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.

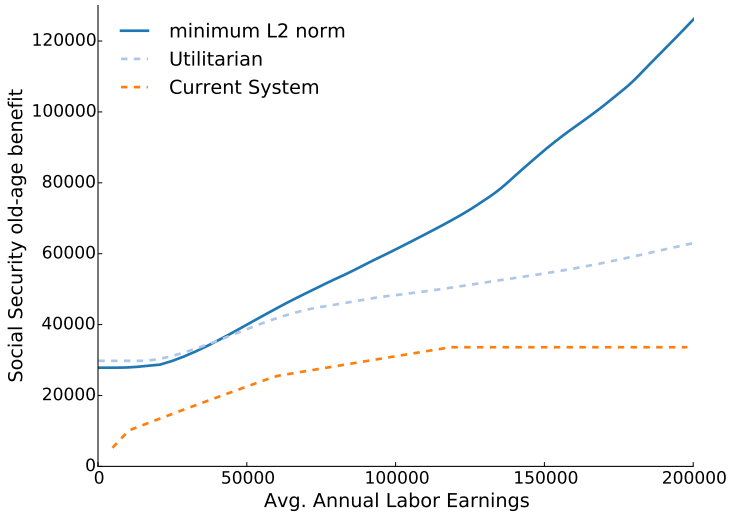


Figure 8. Old-age benefits by labor earnings level. Optimal policy for benchmark normative model and utilitarian planner (same discounting level as benchmark model). On the horizontal axis we have average labor earnings of the individual during its working life and in the vertical axis we have the level 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).

Mullainathan et al. (2012) 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).[30]Conversely, in a classical model the wedge is caused by an externality in which 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

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

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

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

In this model the first best level of the action a does not vary with the agent's type and is independent of redistribution.[31] 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

[31] 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.

redistribution policy so that agent's budget is still given by

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

It is then straightforward to show Lemma 1.

Lemma 1. *(Quantity restriction without redistribution) A quantity restriction of* $a\left(\theta, \alpha\right) = a^*$ *such that*

$$b'\left(a^*\right) = p$$

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

Proof. In fact, let Pareto weights be given by $\lambda\left(\theta\right) = \frac{1}{u'(c^{LF*}(\theta))}$ where $c^{LF*}\left(\theta\right)$ 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 3. *Fix* θ_N *and assume that* $0 \in A$*. If* $\lambda\left(\theta\right)$ *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 3 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.

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.[32] Some individuals do care considerably about pollution and buy more

[32] Allcott et al. (2014) also studies energy policies that deal with externalities.

energy efficient cars, whereas some other individuals are not particularly 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.[33] 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 1 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.

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

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

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

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.[34]

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 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

[34] 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.

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 Saez (2001)). Further, Heathcote and Tsujiyama (2015) 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, Hochman and Rodgers (1969) and Coate (1995) 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.

References

Mark Aguiar and Manuel Amador. Growth in the shadow of expropriation*. *Quarterly Journal of Economics*, 126(2), 2011.

Sule Alan and Martin Browning. Estimating intertemporal allocation parameters using synthetic residual estimation. *The Review of Economic Studies*, 77(4):1231–1261, 2010.

Sule Alan, Martin Browning, and Mette Ejrnaes. Income and consumption: a micro semi-structural analysis with pervasive heterogeneity1. *Available at SSRN 2566477*, 2014.

Hunt Allcott, Sendhil Mullainathan, and Dmitry Taubinsky. Energy policy with externalities and internalities. *Journal of Public Economics*, 112:72 – 88, 2014.

Manuel Amador, Ivan Werning, and George-Marios Angeletos. Commitment vs. flexibility. *Econometrica*, 74(2):365–396, 2006.

Mark Armstrong. Multiproduct nonlinear pricing. *Econometrica: Journal of the Econometric Society*, pages 51–75, 1996.

Mark Armstrong and Jean-Charles Rochet. Multi-dimensional screening:: A user's guide. *European Economic Review*, 43(4):959–979, 1999.

Anthony B Atkinson and Joseph E Stiglitz. The structure of indirect taxation and economic efficiency. *Journal of Public Economics*, 1(1):97–119, 1972.

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.

Ned Augenblick, Muriel Niederle, and Charles Sprenger. Working over time: Dynamic inconsistency in real effort tasks*. *The Quarterly Journal of Economics*, page qjv020, 2015.

Marco Battaglini and Rohit Lamba. Optimal dynamic contracting. *Economic Theory Center Working Paper*, (46-2012), 2015.

Roland Benabou. Unequal societies: Income distribution and the social contract. *American Economic Review*, pages 96–129, 2000.

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.

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.

Stephen Coate. Altruism, the samaritan's dilemma, and government transfer policy. *The American Economic Review*, pages 46–57, 1995.

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.

Peter Diamond and Johannes Spinnewijn. Capital income taxes with heterogeneous discount rates. *American Economic Journal: Economic Policy*, 3(4):52–76, 2011.

Peter A Diamond. A framework for social security analysis. *Journal of Public Economics*, 8 (3):275–298, 1977.

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.

Financial Engines. Missing out:how much employer 401(k) matching contributions do employees leave on the table? Technical report, Financial Engines, 2015.

Emmanuel Farhi and Xavier Gabaix. Optimal taxation with behavioral agents. Technical report, National Bureau of Economic Research, 2015.

Emmanuel Farhi and Iván Werning. Progressive estate taxation. *The Quarterly Journal of Economics*, 125(2):635–673, 2010.

Martin Feldstein. The optimal level of social security benefits. *The Quarterly Journal of Economics*, 100(2):303–320, 1985.

Martin S Feldstein. The effects of taxation on risk taking. *The Journal of Political Economy*, pages 755–764, 1969.

Ana Fernandes and Christopher Phelan. A recursive formulation for repeated agency with history dependence. *Journal of Economic Theory*, 91(2):223–247, 2000.

Simone Galperti. Commitment, flexibility, and optimal screening of time inconsistency. *Econometrica*, 83(4):1425–1465, 2015. ISSN 1468-0262. doi: 10.3982/ECTA11851. URL http://dx.doi.org/10.3982/ECTA11851.

Mikhail Golosov, Narayana Kocherlakota, and Aleh Tsyvinski. Optimal indirect and capital taxation. *Review of Economic studies*, pages 569–587, 2003.

Mikhail Golosov, Maxim Troshkin, Aleh Tsyvinski, and Matthew Weinzierl. Preference heterogeneity and optimal capital income taxation. *Journal of Public Economics*, 97: 160–175, 2013.

Mikhail Golosov, Maxim Troshkin, and Aleh Tsyvinski. Redistribution and social insurance. *American Economic Review*, Forthcoming.

Pierre-Olivier Gourinchas and Jonathan A Parker. Consumption over the life cycle. *Econometrica*, 70(1):47–89, 2002.

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.

Jonathan Heathcote and Hitoshi Tsujiyama. Optimal income taxation: Mirrlees meets ramsey. 2015.

Jonathan Heathcote, Kjetil Storesletten, and Giovanni L Violante. Optimal tax progressivity: An analytical framework. Technical report, National Bureau of Economic Research, 2014.

Harold M Hochman and James D Rodgers. Pareto optimal redistribution. *The American Economic Review*, pages 542–557, 1969.

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.

Kenneth Judd and Che-Lin Su. Optimal income taxation with multidimensional taxpayer types. In *Computing in Economics and Finance*, volume 471, 2006.

Marek Kapička. Efficient allocations in dynamic private information economies with persistent shocks: A first-order approach. *The Review of Economic Studies*, page rds045, 2013.

Henrik Jacobsen Kleven, Claus Thustrup Kreiner, and Emmanuel Saez. The optimal income taxation of couples. *Econometrica*, 77(2):537–560, 2009.

Laurence J Kotlikoff, Avia Spivak, and Lawrence H Summers. The adequacy of savings. *American Economic Review*, 72(5):1056–69, 1982.

David Laibson. Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2):443–478, 1997.

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.

Ben Lockwood and Dmitry Taubinsky. Regressive sin taxes. Technical report, Working Paper, 2015.

R Preston McAfee and John McMillan. Multidimensional incentive compatibility and mechanism design. *Journal of Economic Theory*, 46(2):335–354, 1988.

James A Mirrlees. An exploration in the theory of optimum income taxation. *The review of economic studies*, pages 175–208, 1971.

José Luis Montiel Olea and Tomasz Strzalecki. Axiomatization and measurement of quasi-hyperbolic discounting. *The Quarterly Journal of Economics*, 129(3):1449–1499, 2014.

Sendhil Mullainathan, Joshua Schwartzstein, and William J Congdon. A reduced-form approach to behavioral public finance. *Annu. Rev. Econ*, 4:17–1, 2012.

Ted O'Donoghue and Matthew Rabin. Studying optimal paternalism, illustrated by a model of sin taxes. *American Economic Review*, pages 186–191, 2003.

Ted O'Donoghue and Matthew Rabin. Optimal sin taxes. *Journal of Public Economics*, 90 (10):1825–1849, 2006.

Thomas Piketty and Emmanuel Saez. Income inequality in the united states, 1913–1998. *The Quarterly Journal of Economics*, 118(1):1–41, 2003.

Jean-Charles Rochet and Philippe Choné. Ironing, sweeping, and multidimensional screening. *Econometrica*, pages 783–826, 1998.

Casey Rothschild and Florian Scheuer. Redistributive Taxation in the Roy Model. *The Quarterly Journal of Economics*, 128(2):623–668, 2013. URL http://ideas.repec.org/a/oup/qjecon/v128y2013i2p623-668.html.

Emmanuel Saez. Using elasticities to derive optimal income tax rates. *Review of economic studies*, 68(1):205, 2001.

Emmanuel Saez. The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes. *Journal of Public Economics*, 83(2):217 – 230, 2002. ISSN 0047-2727.

Stefanie Stantcheva. Optimal taxation and human capital policies over the life cycle. Technical report, National Bureau of Economic Research, 2015.

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.

Pei Cheng Yu. Optimal taxation with time-inconsistent agents. Technical report, University of Minessota - Working Paper, 2015.

A Proofs for Section 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 2 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 2 builds upon the proof of Lemma 2 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 2. *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.[35] 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 $\theta' \in \{\theta_2, \ldots, \theta_N\}$. □

Corollary 2. *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 $\bar{\theta}_{N-1} = 0$,

[35] 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 $dE = \sum \pi(\theta, \beta)\left[C'(u_1(\theta, \beta))\nu - Y'(v(\theta, \beta))\varepsilon\right]$. 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.

then the result holds as $v(\theta_N, \beta) > 0$ 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 3. *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 2 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 2 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 3 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

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

If we set $dE = 0$ we obtain

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

The government objective change is given by

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

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

Proof of Theorem 2

First let's show the results for low earnings ability. By Lemma 2, 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

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

for some $\kappa > 0$. We want to show that $\kappa < 1$.[36] 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

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

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

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

extra resources at the margin as we have

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

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.

B 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 Galperti (2015) 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.[37]

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.[38] 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 $\rho_{t+1}\left(\theta_{t+1}|\theta_t\right)$ 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

[37] 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.

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

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 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.[39] 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 C 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 $(c_t, y_t) : H^t \to \mathbb{R}^2_+$.

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

[39] 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.

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

$$\begin{aligned} &\max_{c,y} W(c,y) \\ &\quad s.t.\ (c,y) \text{ is implementable} \end{aligned}$$

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 4. *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.[40] 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

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

insurance implies that intertemporal choices satisfy the inverse Euler equation[41]

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

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 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 5. *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:*

$$\tau\left(h^{t-1},(\theta_1,\beta)\right) = \sum_{h_{t+1}\in H_{t+1}} \pi_{t+1}\left(h_{t+1}|\theta_1\right) \frac{u'\left(c_t\left(h^{t-1},(\theta_1,\beta)\right)\right)}{\delta R u'\left(c_{t+1}\left(h^{t-1},(\theta_1,\beta),(\theta_{t+1},\beta_{t+1})\right)\right)} - 1 \geq 0$$

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*

   2.1 $(h^{t-1},(\theta_N,\beta_M))$ have a weakly positive time-inconsistency wedge

   2.2 $(h^{t-1},(\theta_N,\beta_1))$ have a negative time inconsistency wedge:

$$\tau\left(h^{t-1},(\theta_N,\beta_1)\right) = \sum_{h_{t+1}\in H_{t+1}} \pi_{t+1}\left(h_{t+1}|\theta_N\right) \frac{u'\left(c_t\left(h^{t-1},(\theta_N,\beta_1)\right)\right)}{\delta R u'\left(c_{t+1}\left(h^{t-1},(\theta_N,\beta_1),(\theta_{t+1},\beta_{t+1})\right)\right)} - 1 < 0$$

[41] For applications of this result for optimal taxation see Golosov et al. (2003), Stantcheva (2015) and Golosov et al. (Forthcoming).

C 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

$$\begin{aligned} V_t\left(\hat{r}^{t-1}, h^t, (\sigma_s)_{s=t}^T\right) =& U_t\left(c_t\left(\hat{r}^t, \sigma_t\left(\hat{r}^t, h^t\right)\right), y_t\left(\hat{r}^{t-1}, \sigma_t\left(\hat{r}^{t-1}, h^t\right)\right), \theta_t\right) \\ &+ \beta_t \sum_{s=t+1}^{T} \delta^{s-t} \sum_{h^s \succ h^t} \pi_s\left(h^s | \theta_t\right) U_s\left(c_s\left(\sigma_s\left(\hat{r}^{s-1}, h^s\right)\right), y_s\left(\sigma_s\left(\hat{r}^{s-1}, h^s\right)\right), \theta_s\right) \end{aligned} \tag{6}$$

so the preference is hyperbolic with a present-bias of β_t at period t.[42]

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.[43]

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

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

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

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

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

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.[44]

The next result is the extension of Lemma 2 for the dynamic case.

Lemma 4. *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

[44] This characterization implies that only equilibrium path allocations are important for incentive compatibility (see Fernandes and Phelan (2000) and Kapička (2013)). This argument can break down in problems with perfect correlated types. One example is in Battaglini and Lamba (2015). 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.

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. □

Corollary 3. *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 5. *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 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. ***(Theorem 4 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.[45] Assume by way of contradiction that for a fixed $t<T$ and fixed history

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

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

$h^{t-1} \in H^{t-1}$ 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}:

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

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

for all $\beta^* \in B$. We have from Lemma (4) 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 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. ***(Theorem 4 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

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

all $\beta_t \in B_t$ face the same allocation. Then

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

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):

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

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

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

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 $(h^{t-1}, (\beta_t, \theta_N))$:

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

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

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

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 3, 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

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

From the point of view of β_1 the payoff change is

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

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

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

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 5

We first show result 1 in Theorem 5. Fix a period t and a history h^{t-1}. From Theorem 4 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

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

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

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

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 4 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

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

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

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

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

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

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 5 and from Theorem 4 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 $u_t\left((h^{t-1}, (\beta_1, \theta_N))\right) > u_t\left((h^{t-1}, (\beta, \theta))\right)$ for all $\theta < \theta_N$ and all $\beta \in B$. Denote

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

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

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

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

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

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

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

Since we have $u_t\left(h^{t-1}, \left(\beta_1, \theta_N\right)\right) > u_t\left(h^{t-1}, \left(\beta, \theta\right)\right)$ for $\theta < \theta_N$ and $u_t\left(h^{t-1}, \left(\beta_1, \theta_N\right)\right) \geq u_t\left(h^{t-1}, \left(\beta, \theta_N\right)\right)$ for all β, then we have

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

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

D Proofs for Section 5

In this Appendix we provide a proof for Theorem 3. 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

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

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

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

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

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

Since $\lambda(\theta_N)$ and since $C(\cdot)$ is convex and $u(\theta_N) \geq u(\theta,\alpha)$, then $dE < 0$ for $\nu > 0$, a contradiction. Therefore we must have $b'(a(\theta_N)) \leq p$. Again we can make an analogous perturbation for $\alpha_{min} = \min A \leq 0$. Hence we conclude $b'(a(\theta_N)) = p$. If $\alpha_{max} > 0$, note that the exact same perturbation above also implies $dE < 0$ since $u(\theta_N) > u(\theta,\alpha)$ 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.

E 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 $\sigma = 0.42$.[46]

Each period workers can save $a_t \geq 0$ resources into the next period at a gross rate of

[46] 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.

interest R. Finally workers pay taxes on labor and asset income according to

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

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 $t_a = 0.28$.[47]

After age 65, individuals receive no labor income, only asset income and social security benefits $b\left(y_{65}\right)$. We follow Gourinchas and Parker (2002) and Laibson et al. (2007) 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

$$\begin{aligned} U_t\left(x_t, s_t\right) &= \max u\left(c_t\right) + \beta\delta\mathbb{E}_t\left[V_{t+1}\left(a_{t+1}, y_{t+1}\right)|y_t\right] \\ s.t. & c_t + a_{t+1} = x_t \\ & a_{t+1} \geq 0 \end{aligned}$$

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

$$\begin{aligned} V_t\left(a_t, s_t\right) &= u\left(c_t\left(a_t, y_t\left(s_t\right), s_t\right)\right) + \delta\mathbb{E}_t\left[V_{t+1}\left(a_{t+1}\left(a_t, y_t\left(s_t\right), s_t\right), s_{t+1}\right)\right] \\ x_{t+1} &= R_{t+1}a_{t+1} + y_{t+1}\left(s_{t+1}\right) - T\left(y_{t+1}\left(s_{t+1}\right), R_{t+1}a_{t+1}\right) \end{aligned}$$

We then obtain the following Euler equation:

$$u'\left(c_t^i\left(x_t^i, s_t^i\right)\right) = \left(1-\tau_a\right)R\delta\mathbb{E}_t\left[\left\{1-\left(1-\beta\right)MPC_{t+1}^i\left(x_{t+1}, s_{t+1}\right)\right\}u'\left(c_{t+1}^i\left(x_{t+1}^i, s_{t+1}^i\right)\right)|s_t^i\right]$$

[47] 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.

where

$$MPC^i_{t+1}\left(x_{t+1}, s_{t+1}\right) = \frac{\partial c^i_{t+1}\left(x^i_{t+1}, s^i_{t+1}\right)}{\partial x^i_{t+1}}$$

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 4 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:

$$\tau_\beta = 1 - \left(1 - \beta\right) MPC^i_{t+1}\left(x_{t+1}, s_{t+1}\right)$$

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 $\tau_\beta = 1.0$.

Statistic	Data	Value
δ		0.9856
$avg\ \beta$		0.80
$std\ \beta$		0.62
avg		0.96
$std\ b$		0.058
$corr\left(b, LT\ Labor\ Earnings\right)$		-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

Table 4. Calibration Results