

Optimal illiquidity

Abstract

John Beshears,
Harvard Business
School and NBER,
TIAA Institute Fellow

James Choi,
Yale School of
Management and NBER,
TIAA Institute Fellow

Christopher Clayton,
Harvard University

Christopher Harris,
Cambridge University

David Laibson,
Harvard University
and NBER

Brigitte Madrian,
Brigham Young University
Marriott School of
Business and NBER,
TIAA Institute Fellow

We calculate the socially optimal level of illiquidity in an economy populated by households with taste shocks and present bias (Amador, Werning, and Angeletos 2006). The government chooses mandatory contributions to respective spending/savings accounts, each with a different pre-retirement withdrawal penalty. Penalties collected by the government are redistributed through the tax system. When naive households have heterogeneous present bias, the social optimum is well approximated by a three-account system: (i) a completely liquid account, (ii) a completely illiquid account, and (iii) an account with an $\approx 10\%$ early withdrawal penalty. In some ways this resembles the U.S. system, which includes completely liquid accounts, completely illiquid Social Security and 401(k)/IRA accounts with a 10% early withdrawal penalty. The social optimum is also well approximated by an even simpler two-account system—(i) a completely liquid account and (ii) a completely illiquid account—which is the most common retirement system in the world today.

The project described received funding from the TIAA Institute and Wharton School's Pension Research Council/Boettner Center. The content is solely the responsibility of the authors and does not necessarily represent official views of the TIAA Institute or Wharton School's Pension Research Council/Boettner Center. The authors gratefully acknowledge additional financial support from the Eric M. Mindich Fund for Research on the Foundations of Human Behavior, the National Institute on Aging (P30AG034532), and the Social Security Administration. The views expressed in this paper are not endorsed by NIA, SSA, or any federal government agency. The authors are grateful for comments from Emmanuel Farhi, Mike Golosov, Francesco Lippi, Ben Lockwood, Greg Mankiw, Christian Moser, Matthew Rabin, David Richardson, John Sabelhaus, Dmitry Taubinsky and seminar participants at the ASSA, Einaudi Institute for Economics and Finance, Harvard University, MIT, the SSA RRC, and the NBER. We are very grateful to Kartik Vira for outstanding research assistance.

Any opinions expressed herein are those of the authors, and do not necessarily represent the views of TIAA, the TIAA Institute or any other organization with which the authors are affiliated.

1. Introduction

How much liquidity should be built into a socially optimal savings system? On the one hand, flexibility allows households to consume in ways that reflect their idiosyncratic preferences—i.e., households can respond to normatively legitimate taste shocks. However, liquidity allows households with self-control problems (and other types of biases or errors) to overconsume.

If some illiquidity is optimal, how should it be implemented? Possible forms of illiquidity include a perfectly illiquid retirement claim (like a typical defined-benefit pension or the U.S. Social Security system) or a partially liquid account (like an IRA or 401(k) plan, which allow penalty-based early withdrawals). In some cases, an optimal system would combine different types of illiquid accounts.

In the practical policy domain, there is a partial consensus on some of these questions. Almost all developed countries have some form of compulsory savings that is *completely* illiquid (e.g., Social Security in the United States).

Nevertheless, there are substantial differences among retirement savings systems. For example, in most developed countries, defined-contribution (DC) savings accounts have mandatory contributions and balances that are completely illiquid before age 55 (Beshears, et al., 2015). By contrast, in the United States, DC contributions are voluntary, certain types of withdrawals are allowed without penalty, and, for IRAs, withdrawals may be made for any reason if a 10% penalty is paid. Liquidity engenders significant pre-retirement “leakage”: for every \$1 contributed to the DC retirement accounts of U.S. households under age 55, \$0.30 to \$0.40 simultaneously flows out of the 401(k)/IRA system, not counting rollovers or loans (Argento, Bryant, and Sabelhaus, 2014). From a theoretical perspective, it is not clear whether allowing such leakage is good

or bad from the perspective of overall social welfare. Nevertheless, most media coverage bemoans leakage.¹

Our paper evaluates the optimality of an N -account retirement savings system comprised of liquid, partially illiquid, and/or completely illiquid accounts. The illiquidity is obtained with linear penalties for early withdrawals. Within this frame-work, we focus on systems with two accounts and systems with three accounts. However, we show that such two- and three-account systems come extremely close to delivering the welfare obtainable from a fully general optimized (non-linear) mechanism. We do this by characterizing an upper bound for social welfare in the non-linear mechanism and show that our two- and three-account systems nearly obtain this bound.

We study preferences that include both normatively legitimate taste shifters and normatively undesirable self-control problems. The self-control problems are modeled as the consequence of present bias (Phelps and Pollak, 1968, Laibson, 1997): i.e., a discount function with weights $\{1, \beta\delta, \beta\delta^2, \dots, \beta\delta^t\}$, where the degree of present bias is $1 - \beta$. Our model is an aggregate version (which adds interpersonal transfers) of the flexibility/commitment framework of Amador, Werning, and Angeletos (2006; hereafter referred to as AWA). Our model is very closely related to the model of Moser and Olea de Souza e Silva (2017), who also generalize AWA by allowing for mechanisms with inter-household transfers. In their model, households have unobservable earnings ability and unobservable β , whereas we study the case of unobservable taste shocks (with exogenous earnings) and unobservable β . Moser and Olea de Souza e Silva (2017) find that second-best optimal savings institutions have many of the properties of the U.S. retirement savings system, including forced savings, a theme that also emerges in AWA and in our paper. Like Moser and Olea de Souza e Silva, we find that optimal savings mechanisms are characterized by more mandatory savings than currently exists in the U.S. system. Our paper contributes to this literature by

¹ For example, see Anne Tergesen, “The Rising Retirement Perils of 401(k) ‘Leakage’”, *The Wall Street Journal*, April 2, 2017. For a similar industry perspective, see Hewitt Associates (2009).

showing how highly simplified retirement savings systems (e.g., two- and three-account systems with linear early-withdrawal penalties) come very close to generating welfare levels that arise under the fully general optimized non-linear mechanism.²

We divide our analysis into the cases of homogeneous present bias and heterogeneous present bias. The homogeneous case implies that all agents have the same degree of present bias—in other words, the same value of β . In the case of homogeneous β , our model implies that a single partially illiquid account with (Pigouvian) early-withdrawal penalty $\pi \simeq 1 - \beta$ approximates the fully general socially optimal mechanism.

We then relax the homogeneity assumption, and consider an economy in which agents have *heterogeneous* present bias. In this heterogeneous-preference case, we find that completely illiquid savings accounts play an important role in improving welfare. Specifically, the social optimum is well-approximated by a three-account system with a perfectly liquid savings account, a partially illiquid savings account (with an early-withdrawal penalty of approximately 13%), and a completely illiquid savings account. Even more strikingly, the social optimum is also well-approximated by an even simpler two-account system with a completely liquid savings account and a completely illiquid savings account. In both the two- and three-account systems, the completely illiquid savings account receives a substantial mandatory contribution from the household—enough to almost smooth consumption between working life and retirement even if all other wealth is consumed during working life. The completely illiquid savings account caters to the households with the low β values. Fully illiquid savings generates large welfare gains for these low- β agents, and these welfare gains swamp the welfare losses of the high- β agents (who are made only slightly worse off by shifting some of their wealth from completely liquid accounts to completely illiquid accounts).

To the extent that there is a role for partially illiquid accounts in the heterogeneous- β economy, we find that such accounts should have low early-withdrawal penalties—in most calibrations, the penalty is slightly above 10%. This implies that the partially illiquid accounts look much like a typical 401(k) account in the U.S. Moreover, these partially illiquid accounts display a high level of leakage in equilibrium. In other words, early withdrawals (i.e., pre-retirement withdrawals) are commonplace. This leakage is a double-edged sword: it results in part from legitimate taste shocks and in part from self-control problems (i.e., low β). The costs of the partially illiquid account to low- β types (who end up paying most of the early-withdrawal penalties) and benefits to high- β types (who are net recipients of these penalties) are nearly offsetting, although the net effect for all of society is slightly positive.

Section 2 describes the planner's problem—i.e., account allocations and early-withdrawal penalties that maximize social welfare subject to information asymmetries between the planner and the households. Section 2 presents the benchmark case against which we will compare all other cases: the *laissez faire* policy of completely liquid accounts.

Sections 3 and 4 analyze the case of homogeneous present bias. In Section 3, we follow AWA and assume that inter-household transfers are not permitted: i.e., any revenue collected by the government must be destroyed rather than being redistributed, which is referred to in the literature as the assumption of 'money burning'. Section 4 analyzes the solution to the planner's problem when inter-household transfers are admitted.

Section 5 analyzes the solution to the planner's problem in the case of inter-household transfers and heterogeneous present-bias. This is the most realistic benchmark that we study. We show that the fully optimal retirement savings system is well-approximated by two- and three-account systems, which have a completely liquid account and a (well-funded) completely illiquid

² There is a large literature on optimal taxation when consumers have present bias, including Laibson, Repetto and Tobacman (1998), Gruber and Köszegi (2001, 2004), O'Donoghue and Rabin (2006), Lockwood (2016), Farhi and Gabaix (2018), Allcott, Lockwood and Taubinsky (2019). See Bernheim and Taubinsky (2018) for a review of the literature on behavioral public economics.

account. In the three-account system, welfare is nearly identical to welfare in the two-account system and welfare in the fully general, non-linear system. Accordingly, little can be gained by advancing from two accounts to an arbitrary number of accounts ($N > 2$). In addition, we find that the optimal partially illiquid account in the three-account system is characterized by a high rate of leakage.

Section 6 presents robustness analysis. In Section 7, we conclude and discuss the limitations of our analysis. Three online appendices are used for proofs, including our explanation of how to calculate a tight upper bound on the welfare that can be obtained from the fully general, optimized non-linear mechanism (Appendix 2).

2. Model

We study a two-period model of consumption for a continuum of households with unit mass. Households are indexed by a taste shock θ and a present bias β . In period 1, a household consumes $c_1(\theta, \beta)$. In period 2, a household consumes $c_2(\theta, \beta)$. One can think of period 1 as working life and period 2 as retirement. We will sometimes refer only to c_1 and c_2 for notational simplicity. When we use this short-hand notation, the dependence on θ and β is implied.

In this model, we give households access to N savings accounts with initial mandatory balances $(x_n)_{n=1}^N$ and linear early-withdrawal penalties $(\pi_n)_{n=1}^N$ (which will usually turn out to be positive—i.e., we will usually find $\pi_n > 0$ —but our framework does allow for negative penalties). We allow households to withdraw from these accounts in any order; in equilibrium, they will choose to withdraw from the low-penalty accounts first. This N -account model is a special case of the fully general mechanism-design problem, which we also study. We show that the welfare that arises from the N -account framework with $N \leq 2$ is very close to the welfare for the optimal general (non-linear) mechanism. In

Appendix B, we present a method for characterizing the optimal general (non-linear) mechanism. We choose to focus most of our non-appendix text on the N -account framework because of its similarity to the actual retirement savings systems that are currently in use around the world.

2.1 Preferences of households

We now describe the preferences of households. Let $\theta u_1(c_1)$ represent the utility flow in period 1, where θ is a stochastic taste shifter that is realized in period 1.³ Let $u_2(c_2)$ represent the utility flow in period 2. Without loss of generality, there is no taste shifter in period 2.⁴

From the perspective of the self in period 1, utility flows in period 2 are discounted by a standard (normative) discount factor δ as well as an additional discount factor, β , that represents present bias, with

$$\begin{aligned} 0 < \delta &\leq 1, \\ 0 < \beta &\leq 1. \end{aligned}$$

The agent has no present bias when $\beta = 1$.

Putting these elements together, after realization of θ the agent's preferences in period 1 are given by

$$\theta u_1(c_1) + \beta \delta u_2(c_2).$$

We assume that the flow utility functions, u_1 and u_2 , are both strictly increasing and strictly concave. We assume that u'_1 and u'_2 converge to ∞ as their respective arguments fall to zero (see Ambrus and Egorov, 2013). Following AWA and Beshears, et al. (2015), we assume that the taste shifter θ has bounded support with closure $[\underline{\theta}, \bar{\theta}]$, where $0 < \underline{\theta} < \bar{\theta} < \infty$.

³ See Atkeson and Lucas (1992) for an earlier use of such taste shifters. There are also other ways of modeling taste shifters. For example, one could assume that the utility function is $u(c-\theta)$, where θ is an additive taste shifter inside the utility function. Analyzing this case is beyond the scope of the current paper, but is part of our ongoing work.

⁴ Specifically, including a multiplicative taste-shifter in period 2 would not change any of our results as long as the taste shifter had expected value of one.

2.2 Information structure

We assume that households are naive in the sense that they don't anticipate their own present bias and hence won't use commitment strategies (see Strotz, 1956; O'Donoghue and Rabin, 1999a, 1999b).

We assume that taste shifters, θ , and present bias, β , are private information of each household in the economy. The social planner knows the aggregate distributions of these (independent) parameters. We assume that the distribution function of θ is $F(\cdot)$ and the distribution function of β is $G(\cdot)$. We will make assumptions on these distribution functions in the theorems that follow.

2.3 Preferences of the social planner

The *social planner* and the *household* (with taste shifter θ) have nearly identical preferences over consumption in periods 1 and 2. The only difference is that the social planner does not endorse present bias, implying that the planner's preferences for an individual household are summarized by

$$\theta u_1(c_1) + \delta u_2(c_2).$$

The social planner chooses policies that maximize the social objective function:

$$\int_{\theta, \beta} \left(\theta u_1(c_1(\theta, \beta)) + \delta u_2(c_2(\theta, \beta)) \right) dF(\theta) dG(\beta).$$

Note that the social planner must take account of the (endogenous) equilibrium policy functions of the households, c_1 and c_2 . The social planner can influence these policy functions (as we explain below), but can't control them directly because the social planner doesn't directly observe θ and β for each household (though it can generate a mechanism that leads agents to reveal θ/β). Finally, the social planner must set up a mechanism that uses total resources less than or equal to the aggregate endowment Y .

2.4 Timing

Time 0: The planner sets up N accounts with interest rate R , where N is a constraint that we discuss in the next section. Each of the N accounts is characterized by two variables: an initial allocation x_n and a linear withdrawal penalty π_n , which applies only to withdrawals in period 1. Because penalties apply only in period 1, we refer to them as *early-withdrawal penalties*. Specifically, if a consumer withdraws ω dollars from account n in period 1 with withdrawal penalty π_n , then the consumer receives $(1 - \pi_n)\omega$ dollars.⁵ Without loss of generality, we assume that there are no withdrawal penalties in period 2. From the planner's perspective, the choice variables are the allocations to the N accounts, $(x_n)_{n=1}^N$, and the early (i.e., period-1) withdrawal penalties on those accounts, $(\pi_n)_{n=1}^N$.

In this framework, a completely liquid account has $\pi_n = 0$, a partially liquid account has an early-withdrawal penalty such that $0 < \pi_n < 1$, and a completely illiquid account has an early-withdrawal penalty $\pi_n = 1$.

The planner must satisfy intertemporal, economy-wide budget balance. We state the budget constraint in two equivalent ways. First, the integral of equilibrium consumption over states must equal the overall resources in the economy, Y . Our framework assumes a continuum population of consumers (with measure one), so integrating over taste-parameters, θ and β , is the same as integrating over consumers. Assuming a linear storage technology, the aggregate budget constraint can be written:

$$\int \left(c_1(\theta, \beta) + \frac{c_2(\theta, \beta)}{R} \right) dF(\theta) dG(\beta) \leq Y. \quad (1)$$

An equivalent way of describing budget balance is to relate allocations to resources. Allocations are the accounts given to each consumer. Resources are both the initial endowment and the revenue raised from penalties paid in equilibrium. Let $\omega_n(\theta, \beta)$ be equilibrium

⁵ The framework admits *negative* penalties for period 1 consumption (i.e., subsidies for period 1 consumption).

period-1 withdrawals from account n for consumers with taste shifter θ and present bias β . Then the household-level budget constraints can be written:

$$\begin{aligned} c_1(\theta, \beta) &= \sum_{n=1}^N (1 - \pi_n) \omega_n(\theta, \beta), \\ c_2(\theta, \beta) &= R \sum_{n=1}^N (x_n - \omega_n(\theta, \beta)). \end{aligned}$$

The societal budget constraint can be written

$$\sum_{n=1}^N x_n = Y + \sum_{n=1}^N \left(\pi_n \int \omega_n(\theta, \beta) dF(\theta) dG(\beta) \right).$$

Note that the account allocates resources from two sources: the societal endowment, Y , and the revenue from penalties paid from early withdrawals:

$$\sum_{n=1}^N \left(\pi_n \int \omega_n(\theta, \beta) dF(\theta) dG(\beta) \right).$$

Time 1: Self 1 maximizes her perceived welfare from the perspective of time 1 (which includes present bias). This will generate withdrawals from the accounts established at date 0. Her consumption will be $c_1(\theta, \beta)$.

Time 2: Self 2 spends any remaining funds in the accounts. Her consumption will be $c_2(\theta, \beta)$.

2.5 Summary of the N -account mechanism-design problem

We can now jointly express both the planner's problem and the consumer's problem. We begin with the consumer's problem, since consumer behavior is an input to the planner's problem. In essence, the consumer has only one decision to make.

In period 1, the consumer with parameters θ and β faces the problem

$$\max_{(\omega_n)_{n=1}^N} \theta u_1(c_1) + \beta \delta u_2(c_2), \quad (2)$$

subject to the constraints

$$c_1 = \sum_{n=1}^N (1 - \pi_n) \omega_n, \quad (3)$$

$$c_2 = R \sum_{n=1}^N (x_n - \omega_n). \quad (4)$$

Conditional on the policy vectors $(x_n)_{n=1}^N$ and $(\pi_n)_{n=1}^N$, this generates consumption levels $c_1(\theta, \beta)$ and $c_2(\theta, \beta)$, where we have suppressed the dependency on $(x_n)_{n=1}^N$ and $(\pi_n)_{n=1}^N$.

In period 0, the planner faces the problem

$$\max_{(x_n)_{n=1}^N, (\pi_n)_{n=1}^N} \int \left(\theta u_1(c_1(\theta, \beta)) + \delta u_2(c_2(\theta, \beta)) \right) dF(\theta) dG(\beta) \quad (5)$$

subject to the constraints that (i) $c_1(\theta, \beta)$ and $c_2(\theta, \beta)$ are given by the consumer's problem (equations 2-4) and (ii) economy-wide budget balance is satisfied:

$$\int \left(c_1(\theta, \beta) + \frac{c_2(\theta, \beta)}{R} \right) dF(\theta) dG(\beta) \leq Y, \quad (6)$$

which is equation 1.

In other words, the planner chooses the account allocation vector, $(x_n)_{n=1}^N$, and the penalty vector, $(\pi_n)_{n=1}^N$, to maximize social surplus (equation 5) subject to the constraints that agents will exhibit present bias in their choices (equations 2-4) and that total consumption does not exceed social resources (equation 6).

To simplify notation, and without loss of generality, henceforth we set $\delta = 1$, $R = 1$ and $Y = 1$.

The N -account problem summarized here is a restricted version of a completely general (non-linear) mechanism-design problem (our solution method is explained in Appendix A). We compare our results to the solution of the general mechanism-design problem below.

2.6 Laissez faire reference case: $\pi = 0$

In the analysis that follows, we will always compare social welfare to a reference case in which there are no early-withdrawal penalties—in other words, the agent has access to only one account ($x_1 = Y$), and this account has no penalty for early withdrawal ($\pi_1 = 0$). This is a pure laissez-faire system, in which the government does nothing to distort the decisions of households.

3. Optimal liquidity with homogeneous present bias and no inter-household transfers

In this section, we consider a first deviation from the (laissez faire) reference case. We allow the government to intervene by setting up multiple accounts and imposing early-withdrawal penalties, but we do not allow any inter-household transfers. This is equivalent to saying that any penalty revenue that is collected must be discarded/burned (instead of being transferred to other households through the government budget constraint). Such money burning is a case of theoretical interest and it has been characterized by AWA. This restriction on inter-household transfers is equivalent to assuming that

$$\sum_{n=1}^N x_n = 1.$$

In other words, the sum of the resources allocated to households (account by account) will equal the total sum of resources in society, which is $Y = 1$. (In the next section, we eliminate the money-burning restriction and accordingly allow inter-household transfers to occur through the tax/penalty system.)

In this section, we assume that all agents share a common value of β — i.e., a common degree of present bias. Hence, the distribution function G is degenerate.

With the assumption of no inter-household transfers, our problem can be expressed using our standard notation

with the aggregate budget constraint replaced by a household-level budget constraint:

$$c_1 + c_2 \leq 1 \text{ for each household.}$$

Assuming a degenerate G , no inter-household transfers, and a weak restriction on F , the socially optimal allocation is achieved with only two accounts (AWA, 2006): one account that is completely liquid, and a second account that is completely illiquid in period 1 and completely liquid in period 2. Any additional accounts (with intermediate levels of liquidity) do not have any value to the planner.

Assume that F is differentiable⁶ and define $\Gamma(\theta) \equiv (1-\beta)\theta F'(\theta) + F(\theta)$. Assume there exists $\theta_M \in [\underline{\theta}, \bar{\theta}]$ such that: (i) $\Gamma' \geq 0$ on $(0, \theta_M)$; and (ii) $\Gamma' < 0$ on (θ_M, ∞) . Assume that F' is bounded away from zero on $[\underline{\theta}, \bar{\theta}]$. We will maintain these assumptions on F throughout the paper.

Theorem 1 (AWA, 2006; see also Ambrus and Egorov, 2013) *Assume there is a homogeneous population-wide value of β . Assume that no inter-household transfers are possible (but money-burning is possible). Welfare is maximized by giving self 1 two accounts: a completely liquid account and a completely illiquid account.*

This result is a corollary of a result in AWA and is proven in Appendix A.⁷ The result depends critically on two assumptions: homogeneous β and no inter-household transfers (i.e., money burning). Under these two assumptions (and our maintained regularity assumptions), intermediate penalties (i.e., $0\% < \pi < 100\%$) are socially inefficient because they force resources to be destroyed—the revenue from penalties can't be transferred to other households and must, therefore, be wasted.

This proposition implies that no welfare benefits are achieved by increasing the number of accounts beyond $N = 2$ in the N -account mechanism-design problem (equations 2-6). But the theorem relies on a strong assumption—money burning—which prevents inter-household transfers.

⁶ We assume that F and its derivative have bounded variation. See Appendix A.

⁷ See Ambrus and Egorov (2013), for related information about this argument. The theorem is stated here with a slightly stronger condition on F than the condition used by AWA (see Beshears, et al., 2015).

4. Optimal liquidity with homogeneous present bias and inter-household transfers

In this section, we further relax the restrictions on government by allowing inter-household transfers. Specifically, we now impose *overall* budget balance rather than *consumer-by-consumer* budget balance. With overall budget balance (so that inter-household transfers are allowed), we will show that a perfectly liquid and a perfectly illiquid account are not jointly sufficient to maximize social surplus.

Theorem 2 *Suppose that inter-household transfers are possible. A two-account system with one completely liquid account and one completely illiquid account does not maximize welfare.*

Intuitively, when inter-household transfers are allowed, it is now optimal to use early-withdrawal penalties, because the revenue from those penalties does not need to be destroyed (money burning), but can instead be reallocated to relax the planner's budget constraint. This theorem is proven in Appendix A.

4.1 Optimal policy with quasi-linear utility

Theorem 2 tells us what is not socially optimal (namely a two-account system comprised of a completely liquid account and a completely illiquid account), but it does not provide guidance on what *is* optimal. To gain intuition about socially *optimal* mechanisms, it is helpful to begin by studying the quasi-linear limit case of our model. Specifically, we consider the case in which utility is linear in period 2, i.e. $u_2(c_2) = c_2$. In this way, we obtain a useful exact result that captures the intuition behind the general case in which utility is concave in both periods.

Theorem 3 *Suppose that all households have the same value of β . Suppose that inter-household transfers are possible. Assume that utility is strictly concave in the first period, linear in the second period, and the solution is interior. Then the socially optimal retirement system is a one-account system with a Pigouvian tax on consumption in period 1:*

$$\pi = 1 - \beta$$

This one-account system is also first-best efficient.

Proof. The wedge between the welfare criterion of the planner and the choice function of the agent, which is generated by present bias $\beta < 1$, can be exactly offset by the early-withdrawal penalty $\pi = 1 - \beta$. This Pigouvian tax corrects the negative externality generated by overconsumption. With this penalty, the household's (present-biased) Euler Equation reduces to:

$$(1 - \pi) \theta u'_1(c_1) = \beta \theta u'_1(c_1) = \beta u'_2(c_2).$$

Crossing out identical terms, we obtain

$$\theta u'(c_1) = u'_2(c_2),$$

which is the planner's Euler Equation (if the planner observed θ).

To this point, the argument does *not* rely on quasi-linearity, which we now deploy to prove that the resulting allocation is also first-best. At the margin, all agents are doing some consumption in period 2 (because we assume an interior solution), so for all households the value of a marginal dollar of wealth is $u'_2(c_2) = 1$. Accordingly, social welfare cannot be raised by changing the level of inter-household transfers.

The special case of quasi-linear utility turns out to be a good proxy for the general case in which the utility function is concave in both periods. We next study that case.

4.2 Optimal policy with N accounts

We now return to the general case in which the utility functions in periods 1 and 2, namely u_1 and u_2 , are both strictly concave (as opposed to the quasi-linear case, in which u_2 is linear). We continue to assume that inter-household transfers are possible. Theorem 2 establishes that a perfectly liquid account and a perfectly illiquid account do *not* jointly obtain the social optimum for this case. We now study other account structures using simulation results. Each simulation has a different assumption on the number of accounts and the scope that the planner has to set withdrawal penalties on those accounts. In our benchmark simulations, we make the following assumptions.

A1. The utility functions in periods 1 and 2 are

$$u_1(c) = u_2(c) = \ln(c);$$

A2. The density of the multiplicative taste shocks is a truncated normal distribution. Specifically: we start with a normal distribution (mean $\mu = 1$ and standard deviation $\sigma = 0.25$); truncate it at the symmetrically placed points $1 - \chi$ and $1 + \chi$ (where $\chi = 2/3$, resulting in a distribution with support $[1 - \chi, 1 + \chi]$); and rescale it so that it integrates to one.

We begin with Table 1, which reports the *improvement* in total welfare for different systems of accounts where the planner chooses the optimal x_n and π_n . The entries in this table are percentage welfare improvements with the laissez-faire case as the benchmark. Specifically,

each entry tells us how much social welfare improves expressed as the equivalent percentage improvement in the societal resource endowment; this is typically referred to as a money metric welfare criterion. We use this welfare reporting framework throughout the rest of the paper (with the laissez-faire case as our benchmark in all analyses).

The first row of Table 1 reports the case of one (flexible) account. We refer to the account as flexible to emphasize that the planner sets the penalty-level, π_1 , for this account (as well as the mandatory initial balance x_1). The second row of Table 1 reports the case of two (flexible) accounts: now the planner sets π_1 and π_2 (as well as x_1 and x_2).

Table 1

	Value of β				
	0.1	0.2	0.3	0.4	0.5
1 Flexible	69.658	31.775	17.124	9.630	5.347
2 Flexible	71.648	32.698	17.605	9.882	5.478
General Mechanism (Relaxed)	71.674	32.748	17.659	9.929	5.511
1 Liquid, 1 Illiquid	71.633	32.648	17.482	9.671	5.196
	Value of β				
	0.6	0.7	0.8	0.9	1.0
1 Flexible	2.794	1.283	0.446	0.067	0.012
2 Flexible	2.860	1.314	0.458	0.070	0.012
General Mechanism (Relaxed)	2.881	1.325	0.462	0.071	0.014
1 Liquid, 1 Illiquid	2.542	1.018	0.256	0.015	0.000

Table 1: The welfare gain from four mechanisms (namely a single account with an endogenous penalty (row 1); two accounts with endogenous penalties (row 2); the general mechanism (relaxed), which allows for an arbitrary non-linear budget set and does not impose the monotonicity restriction (row 3); and a two-account system with one completely liquid account and one completely illiquid account (row 4)) calculated for 10 different values of β (namely 0.1, 0.2, ..., 1.0) in the homogeneous- β model. The welfare gain is calculated as the percentage increase in household wealth that would produce the same average welfare in the laissez-faire case. Welfare is calculated using the planner's welfare criterion (i.e., without present bias in the welfare objective).

The third row of Table 1 reports an *upper bound* for welfare improvements that can be achieved with the completely general non-linear mechanism. This general non-linear mechanism does *not* restrict the planner to use a finite set of accounts that each have a linear early-withdrawal penalty. The general non-linear mechanism allows the planner to offer households a non-linear budget set from which each household can pick a consumption pair, (c_1, c_2) . This non-linear mechanism is equivalent to the case in which the planner has a finite set of accounts, can dictate the order in which those accounts are depleted, and lets the number of finite accounts get large ($N \rightarrow \infty$).

Our calculation of welfare improvements in the general non-linear mechanism is an *upper bound* on welfare, because we omit a monotonicity restriction when we calculate the optimum, leading us to refer to this as a ‘relaxed’ case of the general mechanism. In practice, this monotonicity restriction has little or even no effect on welfare. See Appendix B for a complete description of our analysis of the general non-linear mechanism. We know that our reported upper bound for welfare improvements in the general non-linear mechanism is tight in practice because it is nearly identical to the welfare improvements that we calculate for highly restricted *finite-account* cases with a small number of linear accounts (e.g., $N = 2$), as we will show below.

The fourth row of Table 1 reports the case of two accounts: a completely liquid account and a completely illiquid account.

The columns of Table 1 represent different cases of *homogeneous* β , starting with $\beta = 0.1$ and progressing to $\beta = 1.0$.⁸

Table 1 reveals that a simple single-account system generates most of the obtainable welfare gains. For example, consider the column $\beta = 0.6$ (a natural value for a homogeneous calibration in light of current estimates in the empirical literature—see Cohen, et al., 2020). In this column, one flexible account generates a social-welfare gain equal to 2.794% of the endowment (relative to the laissez-faire reference case). Two flexible accounts generate a social-welfare gain equal to 2.860% of the endowment. The general non-linear mechanism generates a welfare gain that is weakly bounded above by 2.881% of the endowment.

This analysis also reveals another important feature of the homogeneous case: the optimal penalties are essentially Pigouvian corrections to present bias. We can see this in Figure 1, where we report the optimal penalty for the one-account case, which turns out to be nearly identical to $(1 - \beta)$, both of which are plotted in Figure 1. This near-Pigouvian result echoes the exact Pigouvian correction that arises in the quasi-linear case (subsection 4.1).⁹

⁸ There is a growing literature on estimation of present bias (e.g., Dellavigna and Paserman, 2005; Shapiro, 2005; Dellavigna and Malmendier, 2006; Gine, Karlan, and Zinman, 2010; Meier and Sprenger, 2010; Augenblick, Niederle, and Sprenger, 2015; see Cohen, et al., 2020 for a review of this literature).

⁹ Similar Pigouvian taxes also arise in the cases with more than one account. For example, with $\beta = 0.6$ and two accounts, the penalties on those two accounts are respectively 0.32 and 0.42, straddling the exact Pigouvian correction of $1 - \beta = 0.4$.

Figure 1

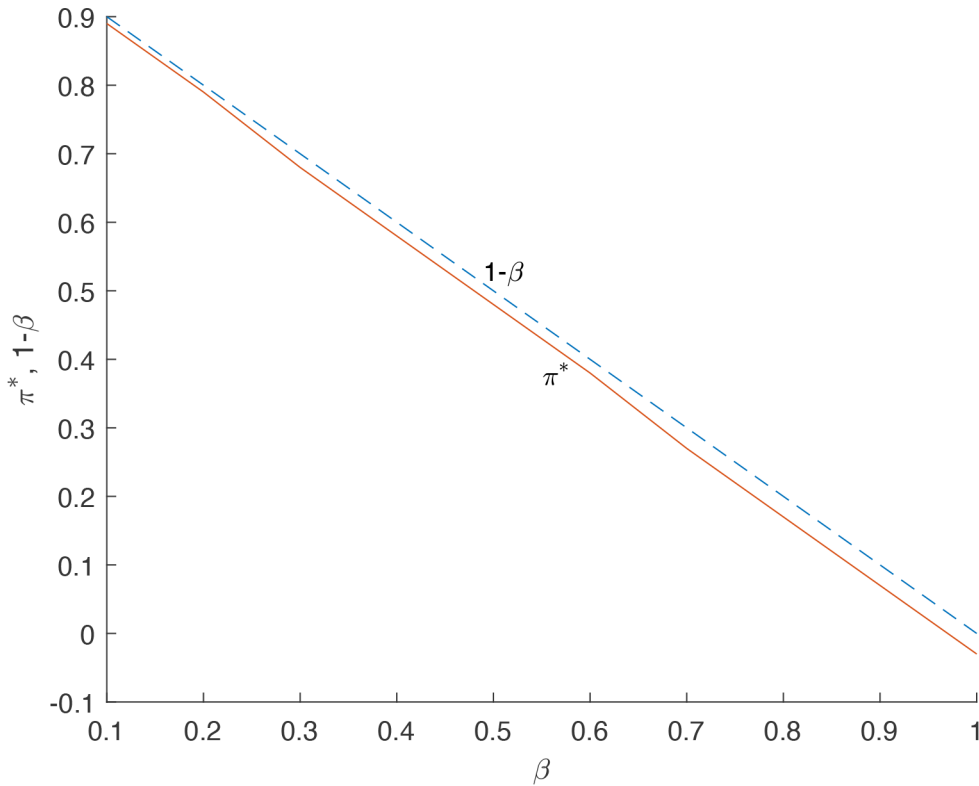


Figure 1: The optimal penalty π^* and the notional Pigouvian tax $1-\beta$ as a function of β in the case in which: (i) the population has homogeneous β ; (ii) the planner is confined to a mechanism with a single account, with penalty π . Note that π^* is always lower than $1-\beta$. In particular, π^* is negative at $\beta=1$. This is due to the redistributive motives of the planner: she wishes to redistribute from types with low θ to types with high θ .

Finally, note that an exact Pigouvian correction (which did arise in the quasilinear case) is *not* generally socially optimal, because there is an inter-household redistributive motive for the planner when both u_1 and u_2 are concave. In this general case, the planner would like to reallocate resources from low- θ types to high- θ types. This redistributive motive is reflected in the fact that the socially optimal penalties in the one-account case (for any given value of β) are all strictly below the corresponding value of $(1-\beta)$. Intuitively,

the households who will be paying the penalties are those households with the higher θ values. To transfer resources to these households, the planner lowers the socially optimal penalty below the $(1-\beta)$ benchmark value. However, as one can see in Figure 1, this downward adjustment is small in magnitude. Accordingly, the Pigouvian correction is the dominant force in these simulations.

5. Optimal liquidity with heterogeneous present bias and inter-household transfers

In this section, we continue to allow inter-household transfers. In addition, we now relax the assumption that consumers have homogeneous β . As in the previous section, we begin with the quasi-linear case and then provide quantitative simulations.

5.1 Optimal policy with heterogeneous present bias and quasi-linear utility

The efficacy of Pigouvian taxation generalizes to an economy with heterogeneous present bias and quasi-linear utility, even though the planner does not directly observe each household's β . To prove this, we exploit the revelation principle and study mechanisms in which agents reveal their intertemporal preferences (between periods 1 and 2). Note that a household's preferences are

$$\theta u_1(c_1) + \beta u_2(c_2).$$

This representation is observationally equivalent (in choices) to the preferences,

$$\frac{\theta}{\beta} u_1(c_1) + u_2(c_2).$$

Accordingly, intertemporal preferences have a sufficient statistic, $\phi = \theta/\beta$, and the revelation principle can be implemented with the variable ϕ . We will study general (non-linear) mechanisms in which the agents each report ϕ and receive a consumption pair (c_1, c_2) which depends on their reported value of ϕ .

Theorem 4 *Suppose that inter-household transfers are possible. Assume that utility is strictly concave in the first period and linear in the second period, that the solution is interior and that $E[\theta | \phi]$ is non-decreasing in $\phi = \theta/\beta$. Then the optimal allocation is characterized by*

$$E[\theta | \phi] u'_1(c_1(\phi)) = 1,$$

and the implied (local) marginal penalty rate for period 1 withdrawals is

$$\pi(\phi) = E[1 - \beta | \phi].$$

Note that this penalty is an 'average Pigouvian correction,' in the sense that the marginal dollar of consumption in period 1 is penalized with the *conditional* expected value of $1 - \beta$, where the conditioning is done with respect to the (truthfully) reported value of ϕ . The proof of Theorem 4 is in Appendix C.

Theorem 4 tells us that an optimal mechanism will generate penalties ranging from $\pi = 1 - \bar{\beta}$ to $\pi = 1 - \underline{\beta}$. Intuitively, households who report the lowest feasible value of $\phi = \theta/\beta$ are revealing that they have the highest feasible value of $\beta = \bar{\beta}$ and the lowest feasible value of θ . They will consume relatively little in period 1 and will face a penalty of $\pi = 1 - \bar{\beta}$ for this consumption, which is an exact Pigouvian correction because their $\bar{\beta}$ value has been fully revealed. An identical argument applies for households who report the *highest* feasible value of $\phi = \theta/\beta$. They have revealed their $\beta = \underline{\beta}$ value, and accordingly will receive an exact Pigouvian correction embodied by a penalty of $\pi = 1 - \underline{\beta}$.

5.2 Optimal policy with N accounts

We now switch from the case of quasi-linear utility to the case in which the consumer has log utility in both periods. In the current and the following subsections, we study optimal mechanisms using numerical solutions. We focus on cases in which β is uniformly distributed on the interval $[0.2, 1]$, implying a mean value of 0.6.

As before, each simulation has a different assumption on the (finite) number of accounts and the scope that the planner has to set withdrawal penalties on those accounts. In all of these cases, we make assumptions A1 and A2 from the previous section, pinning down the utility functions u_1 and u_2 and the distribution function of the taste shock F . We explore the robustness of these particular cases in Section 6.

Table 2 reports the welfare improvements (again using a money metric) that are obtained when the planner shifts from the laissez-faire reference system to an N -account system with (second-best) optimal disincentives to overconsume in period 1. The first row of Table 2 reports the case of one flexible account. The second row reports the case of two flexible accounts. The third row reports the case of the general non-linear mechanism (see

the earlier discussion in Subsection 4.2 and the full derivation in Appendix B). The fourth row reports the case of two accounts: a completely liquid account and a completely illiquid account. The fifth row reports the case of three accounts: a completely liquid account, a partially illiquid account, and a completely illiquid account.

Table 2 reveals that a one-account system no longer obtains most of the feasible welfare gains: one flexible account generates a social-welfare gain of only 3.569% of the endowment. (Following the approach in Table 1, all welfare gains in Table 2 are reported relative to the laissez faire reference case.) With just one account, the system cannot generate the necessary heterogeneity in marginal penalties for early consumption. However, a two-account system does get very close to this efficiency bound: two flexible accounts generate a

social-welfare gain equal to 6.136% of the endowment. (For the two account case, we find that one penalty is close to zero and the other is close to one.) The general mechanism generates a welfare gain that is bounded above by 6.144% of the endowment. A completely liquid and a completely illiquid account generate a welfare gain of 6.105% of the endowment. Finally, the three-account system (one completely liquid, one partially liquid and one completely illiquid) generates a welfare gain of 6.137% of the endowment. The (money-metric) differences among the mechanisms with more than a single account are small in economic magnitude. To summarize, a very simple two-account system—one perfectly liquid and one perfectly illiquid—generates approximately optimal welfare gains. Such a two-account system is commonplace in most countries in the developed world (Beshears, et al., 2015).

Table 2

1 Flexible	3.569
2 Flexible	6.136
General Mechanism (Relaxed)	6.144
1 Liquid, 1 Illiquid	6.105
1 Liquid, 1 Flexible, 1 Illiquid	6.137

Table 2: The welfare gain from five mechanisms (namely a single account with an endogenous penalty (row 1); two accounts with endogenous penalties (row 2); the general mechanism (relaxed), which allows for an arbitrary non-linear budget set and does not impose the monotonicity restriction (row 3); a two-account system with one completely liquid account and one completely illiquid account (row 4); and a three-account system with one completely liquid account, one account with an endogenous penalty and one completely illiquid account (row 5)) in the heterogeneous- β model (with β distributed uniformly between 0.2 and 1).

To gain intuition for this result, we report a related set of analyses in Figure 2. Here, we study a two-account system. One account is completely liquid (i.e., $\pi_1 = 0$) and the other account has varying illiquidity (i.e., π_2 varies). As we vary the penalty π_2 from 0 to 1, we re-optimize the allocations x_1 and x_2 to the liquid and the partially illiquid accounts. The horizontal axis shows the penalty π_2 , and the vertical axis shows the average welfare of the cross sections of the population obtained by fixing β at one of the five values 0.2, 0.4, 0.6, 0.8 and 1.0. In this context it should be emphasized that all households are treated identically ex ante and, therefore, receive the same

allocations and face the same early-withdrawal penalties. Also, the allocations are chosen to maximize the welfare of the population as a whole. The welfare of a given β cross section is then obtained by averaging over the welfare of the households with that particular value of β .

For the $\beta = 0.2$ households, welfare as perceived by the planner rises dramatically as the early-withdrawal penalty increases (Figure 2). Indeed, as π_2 rises from 0 to 1, the increase in welfare for the $\beta = 0.2$ households is equivalent, using a money metric, to a 30% increase of wealth (the money-metric welfare improvement is

equivalent to half of the vertical distance on Figure 2). From the planner's perspective, there is a substantial gain from discouraging these low- β households from excessive consumption in period 1.

Households with other β values experience increasing and then decreasing welfare as π_2 increases from 0 to 1. However, conditional on β , all households experience a rise in expected welfare as π_2 rises from zero. For low- β households, this rise occurs because higher penalties prevent low- β households from overconsuming in period

1. For high- β households, this rise occurs because higher penalties generate larger cross-subsidies from low- β households to high- β households. Specifically, these cross-subsidies occur because higher penalty revenue relaxes the planner's budget constraint, thereby enabling the planner to give agents higher endowments in period 1. High- β households are net recipients of cross-subsidies because they tend to make smaller early withdrawals and, therefore, pay fewer penalties than low- β households (look ahead to Figure 4).

Figure 2

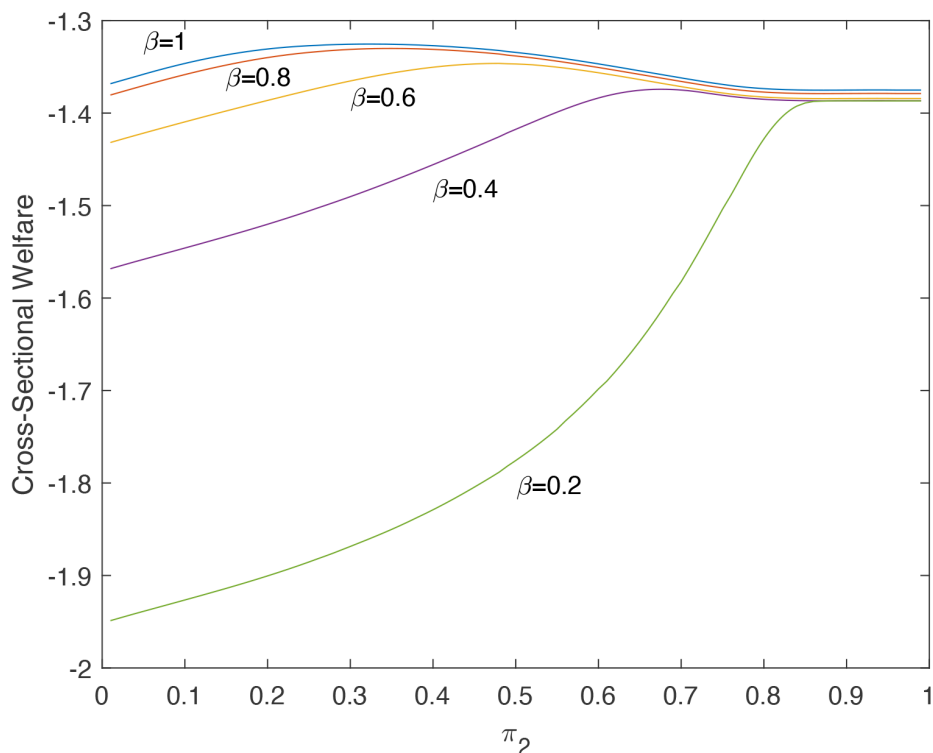


Figure 2: The welfare of various β cross sections of the population as a function of π_2 in the case in which: (i) the population has heterogeneous β ; (ii) the planner is confined to a mechanism with two accounts, with penalties π_1 and π_2 respectively; (iii) $\pi_1 = 0$ (i.e., the first account is completely liquid); (iv) the account allocations are chosen to maximize the welfare of the population as a whole. Note that the cross section of the population with $\beta = 0.82$ (not shown) is almost indifferent between the system with $\pi_2 = 0$ and the system with $\pi_2 = 1$.

Unlike the welfare of low- β households, which rises monotonically as π_2 rises, the welfare of high- β households eventually peaks and thereafter falls with π_2 . This single-peaked property arises because, while initial rises in π_2 simply result in greater revenue from the early-withdrawal penalties paid by low- β households, later rises tend to eliminate early withdrawals altogether. Hence the cross-subsidy to high- β households first rises and then falls. By the time π_2 reaches 1, the cross-subsidy has been completely eliminated, and high- β households are now facing a binding constraint (if they have a sufficiently high θ value) that limits their ability to adjust consumption in period 1, so high- β households are slightly worse off on average than they were when π_2 was 0. On a money-metric basis, the $\beta = 1$ households experience a welfare loss equivalent to 0.23% of their income as the planner moves from $\pi_2 = 0$ to $\pi_2 = 1$ in Figure 2. However, this welfare loss is swamped by the welfare gain experienced by the $\beta = 0.2$ types (which is two orders of magnitude larger).

Figure 3 shows the welfare of the population as a whole as a function of the early-withdrawal penalty π_2 . It confirms that—as one would expect—the enormous

welfare gains for low- β households swamp the modest welfare losses for high- β households, an example of asymmetric paternalism (Camerer, et al., 2003). Although it appears that total social welfare rises monotonically and asymptotes, social welfare actually reaches a global maximum at $\pi_2 = 0.85$ and then falls very slightly. However, the fall in welfare between $\pi_2 = 0.85$ and $\pi_2 = 1$ is insignificant: it is 0.00002% of wealth using a money metric. Accordingly, the social optimum is effectively obtained with one completely liquid account and one completely illiquid account.

Figure 4 reports the gross penalties paid by households with different values of β (again integrating over θ). As anticipated above, the penalties are hump shaped in π_2 , with the hump occurring for larger values of π_2 the lower the value of β . For example, for $\beta = 1$, the hump peaks at $\pi_2 = 0.07$. For $\beta = 0.2$, the hump peaks at $\pi_2 = 0.51$. Intuitively, high- β households stop making early withdrawals even when the penalty is low (because they value utils in period 2 almost as much as they value utils in period 1). By contrast, low- β households do not stop making early withdrawals until the penalty reaches a relatively high level.

Figure 3

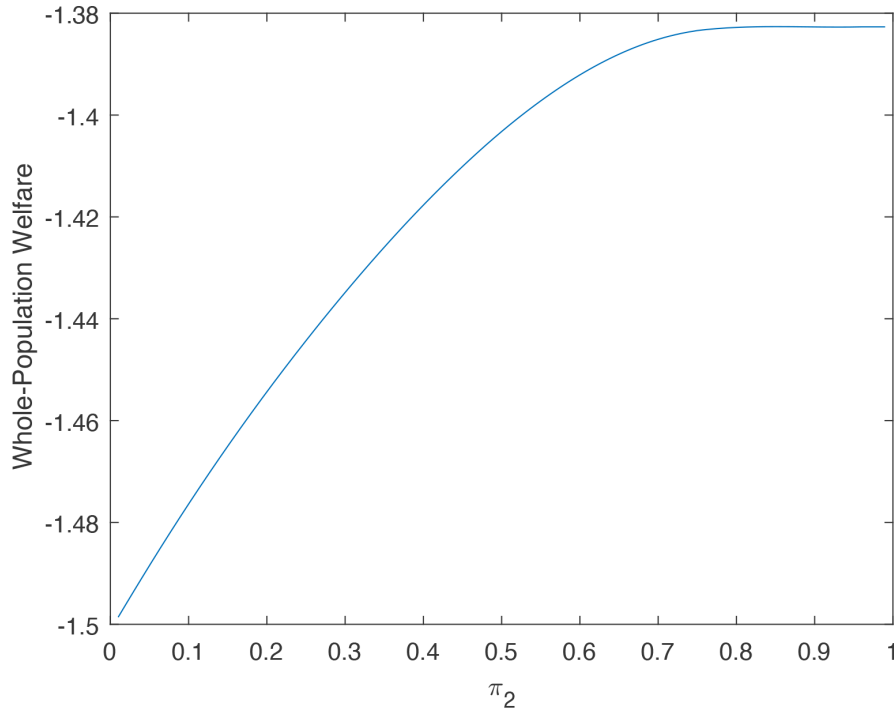


Figure 3: The welfare of the population as a whole as a function of π_2 in the case in which: (i) the population has heterogeneous β ; (ii) the planner is confined to a mechanism with two accounts, with penalties π_1 and π_2 , respectively; (iii) $\pi_1 = 0$ (i.e., the first account is completely liquid); (iv) the account allocations are chosen to maximize the welfare of the population as a whole. Note that: (i) while this is not immediately apparent from the figure, the function in question is non-monotone; (ii) the optimal penalty π_2^* is approximately 85%; (iii) π_2^* yields a proportional increase of approximately 0.00002% in money-metric welfare relative to the case in which $\pi_2 = 1$ (i.e., the case in which the second account is completely illiquid). In particular, the welfare cost of setting the penalty on the second account too low far exceeds that of setting it too high.

Figure 4

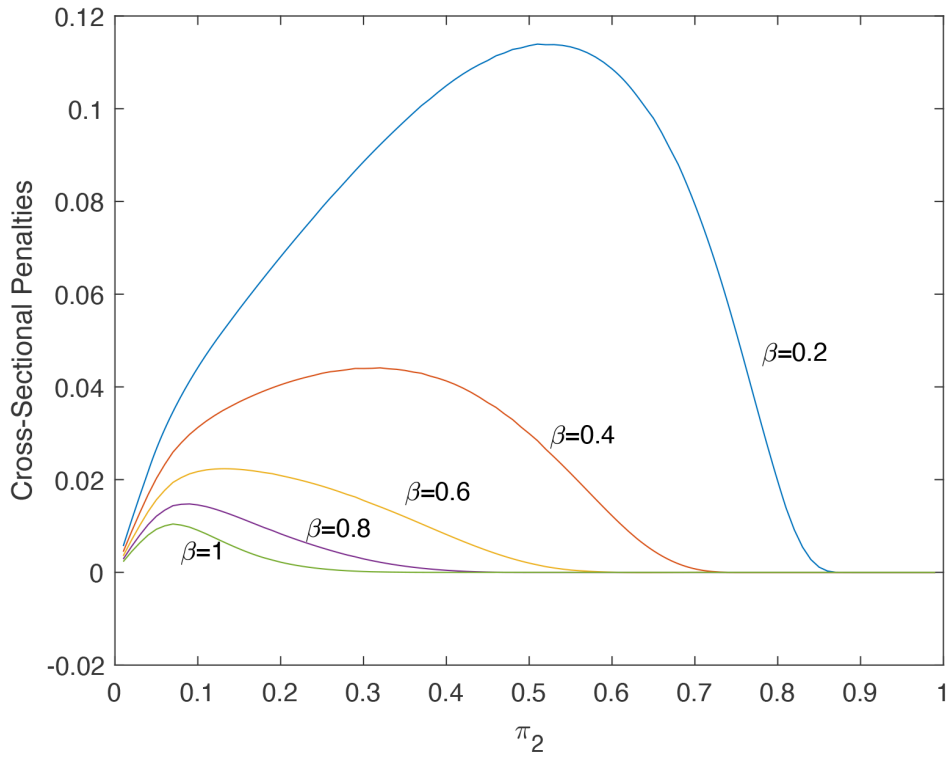


Figure 4: The total penalties paid by various β cross sections of the population as a function of π_2 in the case in which: (i) the population has heterogeneous β ; (ii) the planner is confined to a mechanism with two accounts, with penalties π_1 and π_2 , respectively; (iii) $\pi_1 = 0$ (i.e., the first account is completely liquid); (iv) the account allocations are chosen to maximize the welfare of the population as a whole.

5.3 A three-account system that approximates the U.S. retirement savings system

The fifth row in Table 2 reports the welfare gains for a three-account system ($N = 3$). We will see that this analysis reproduces some of the features of the U.S. system.

We constrain the first account to be completely liquid ($\pi_1 = 0$) and the third account to be completely illiquid ($\pi_3 = 1$). Think of this third account—the illiquid account—as Social Security or a defined-benefit pension. The planner picks the penalty on the “middle” account ($0 < \pi_2 < 1$) and the values of the three endowments (x_1, x_2 and x_3) to optimize social welfare (while satisfying the budget constraint). The “middle” account turns out to have an optimal penalty of $\pi_2 = 0.13$, which is close to the actual penalty associated with a 401(k) or IRA account, namely 0.10.

Adding this optimized “middle” account to the constrained two-account system (row 4 in Table 2) only slightly raises consumer welfare relative to the two-account system with a completely liquid and a completely illiquid account. Comparing rows 4 and 5 of Table 2, we see that the addition of the middle account increases social welfare by $6.137\% - 6.105\% = 0.032\%$ of wealth (using a money metric).

Our simulations reveal that the middle account is characterized by a very high degree of leakage in equilibrium. Ninety percent of the assets in the middle account are withdrawn to fund consumption in period 1. Figure 5 disaggregates this result, by plotting the cumulative distribution function of the ratio c_2/c_1 . Figure 5 shows that 76% of households choose the maximal value of c_1 (corresponding to a full withdrawal of the funds in the partially liquid account), generating a consumption ratio of $c_2/c_1 = 0.94$. Another 22% of households withdraw at least some (though not all) of the funds in the partially liquid account. Another 1% of households choose to save all of the funds in the partially liquid account for retirement (but no more than that). Finally,

1% of households choose to save some of the funds in the completely liquid account for retirement (in addition to all of the funds in the partially liquid account).

In summary, our analysis finds that welfare is nearly as high in the two-account system with a completely liquid account and a completely illiquid account as it is in the three-account system that adds a partially illiquid account.¹⁰ When a third account is added, it looks and performs somewhat like a U.S. 401(k) plan: the third account has an optimized penalty of 0.13 and generates a very high rate of leakage in equilibrium. This high leakage rate is even higher than the empirical leakage rate observed in the U.S. system.

One explanation for the difference between the model-predicted leakage rate (90%) and the empirically observed leakage rate (40%) is that initial account balances in the model are generated by government fiat, whereas almost all of the dollars in real-world 401(k)/IRA accounts are voluntarily deposited, implying that they are coming from households with higher β values and lower θ values in the first place. In this sense, one can't directly compare the leakage rate in the model (which is the aftermath of universal *forced* savings in a DC system) and the leakage rate in the US economy (which is the aftermath of *voluntary* savings in a DC system). Accordingly, differential selection makes this an apples to oranges comparison.

Another key factor that explains the high model-predicted leakage rate is the fact that the planner endogenously (optimally) chooses to put a large fraction of each household's resources into the completely illiquid account, thereby reducing the marginal value of retirement consumption, and implicitly encouraging pre-retirement withdrawals of balances in the partially illiquid account. However, we find that leakage rates from the partially illiquid account are substantially lower under a system in which contributions to the completely illiquid account (e.g., Social Security) are exogenously set at a lower level than is recommended by our normative framework. Our mechanism-design framework generates

¹⁰ The third account offers the welfare benefit of additional separation for high- θ households and low- θ households. However, the third account has two effects that jointly offset the welfare gains from separation. First, the third account enables low- β households to increase their period 1 over-consumption. Second, withdrawals from the third account generate (socially inefficient) transfers of resources from low- β and high- θ households to high- β and low- θ households because of the penalties that are paid for period 1 withdrawals from the third account. These tax revenues are redistributed in the mechanism.

nearly complete consumption smoothing between period 1 and period 2, even if the household relies only on the resources from the completely illiquid account to finance period-2 consumption. By implication, our framework recommends far more funding for the completely illiquid account than we actually observe in the United States. If

we adopt an empirically more realistic (modest) funding rule for the completely illiquid account, then we observe much lower levels of leakage from the partially illiquid account because its balances are more urgently needed to support consumption in period 2 (i.e., retirement). We return to these issues in Subsection 6.1.

Figure 5

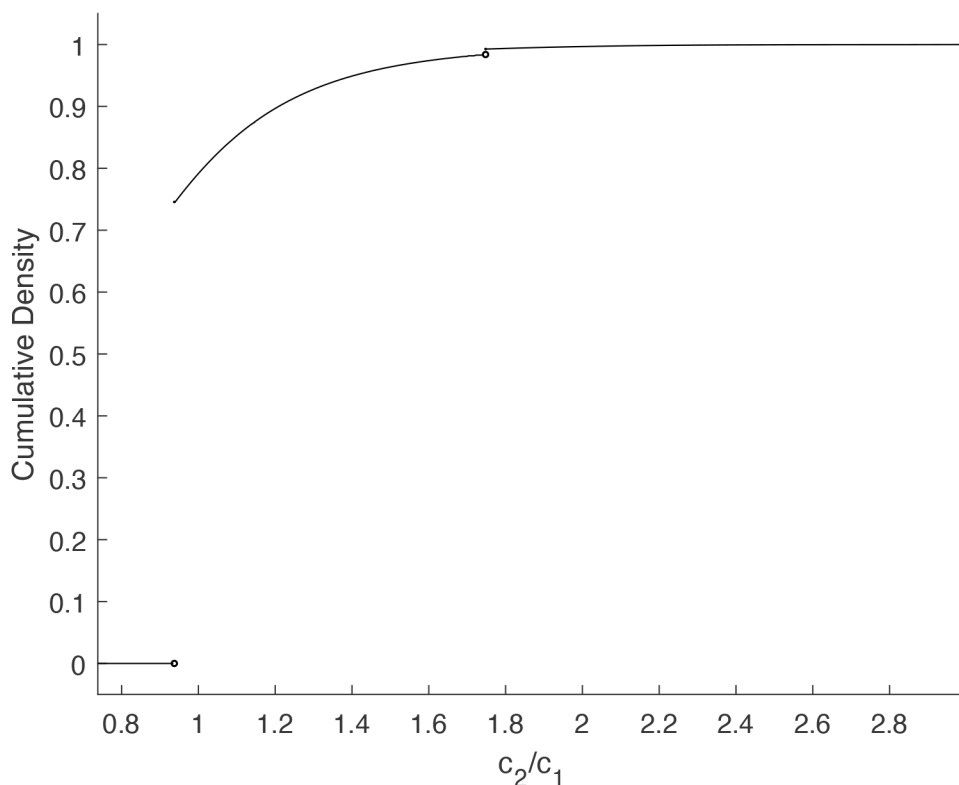


Figure 5: The distribution function of the ratio c_2/c_1 of period-2 consumption to period-1 consumption in the population as a whole in the case in which: (i) the population has heterogeneous β ; (ii) the planner is confined to a mechanism with three accounts, with penalties π_1 , π_2 and π_3 , respectively; (iii) $\pi_1 = 0$ (i.e., the first account is completely liquid); (iv) $\pi_3 = 1$ (i.e., the third account is completely illiquid); (v) both π_2 and the account allocations are chosen to maximize the welfare of the population as a whole. There are two atoms in the distribution: a large atom accounting for about 76% of the total mass near $c_2/c_1 = 0.94$; and a small atom accounting for about 1% of the total mass near $c_2/c_1 = 1.70$. Individuals at the second atom have withdrawn the entire balance from the first (liquid) account, but have not yet touched the second account. Individuals at the first atom have withdrawn the entire balance from both the first and the second accounts. In particular, they have paid the penalty π_2 on the entire balance of the second account.

6. Optimal policy with transfers and heterogeneous present bias: Robustness

In the previous section, which studied the case in which inter-household transfers are allowed and present bias is heterogeneous in the population, three key findings emerged:

1. The constrained-efficient social optimum is approximated by a two-account system, with one account that is completely liquid and a second account that is completely illiquid. Little welfare gain is obtained by moving beyond this simple two-account system.
2. If a third account is added, its optimized early-withdrawal penalty is 13%.
3. The equilibrium leakage rate from this third account is 90%.

In the current section, we document the robustness of these three findings when the distribution of β is heterogeneous.¹¹ With respect to the first finding, the largest incremental welfare gain that we generate in our robustness checks by extending the system of savings accounts *beyond* one completely liquid and one completely illiquid account is 0.081% of income. Hence, we continue to find that a simple system with one completely liquid account and one completely illiquid account is approximately socially optimal.

With respect to the second finding, the optimized penalty on the partially illiquid account ranges from 11% to 14% across our calibrated economies. Hence, we continue to find that the partially illiquid account has a penalty that is similar to the penalties on 401(k)s and IRAs.

With respect to the third finding, the equilibrium leakage rate ranges from 84% to 99%. Hence, we continue to find that our simulations generate very high equilibrium leakage rates.

The detailed results are reported in the three panels of Table 3, which report the welfare gain (relative to the laissez faire case) for (i) the two-account system $\pi_1 = 0$ and $\pi_2 = 1$, (ii) the three-account system with $\pi_1 = 0$, $0 < \pi_2 < 1$, and $\pi_3 = 1$, and (iii) the upper bound of the general mechanism (i.e., the ‘relaxed’ non-linear case described in Appendix B). For case (ii), in addition to the welfare gain, we also report the penalty π_2 and the leakage rate. Note that the upper bound on the welfare gain—case (iii)—is always tight in the sense that it is economically close to the constrained N -account cases that we study.

Table 3a varies the value of the coefficient of relative risk aversion (γ). In our benchmark calibration, we set $\gamma = 1$. In Table 3a, we study the cases $\gamma = 1/2$, $\gamma = 1$ (for comparison), and $\gamma = 2$.

Table 3b varies the shape of the density of θ . Recall that in our benchmark calibration, the density of the multiplicative taste shocks is truncated normal. Specifically, we start with a normal distribution, with mean $\mu = 1$ and standard deviation $\sigma = 0.25$; the resulting density is truncated (and reweighted) with support $[1 - X, 1 + X]$, with $X = 2/3$. In Table 3b, we study the cases $\sigma = 0.30$, $\sigma = 0.25$ (for comparison), and $\sigma = 0.20$.

Table 3c varies the standard deviation of the distribution of β values (holding the mean fixed). In our benchmark calibration, we studied the case of a uniform distribution of β between 0.2 and 1.0. In Table 3c, we study truncated normal distributions of β , with 0.2 and 1.0 serving as the truncation points. Our original benchmark is equivalent to the (truncated) normal case with $\sigma_\beta = \infty$ and $\mu_\beta = 0.6$. We now reduce σ_β to 1, 1/2 and 0 (holding the truncation points and μ_β fixed). The case $\sigma_\beta = 0$ is the degenerate case in which all agents have the same value of $\beta = 0.6$. Our results do not generalize to the degenerate case (the last column of Table 3c). As we showed in Section 5, the homogeneous- β case engenders a Pigouvian tax as the approximately optimal policy.

¹¹ When β takes on a degenerate distribution — i.e., β is homogeneous in the population — these results no longer apply (see the last column in Table 3c).

Table 3

	Value of γ		
	0.5	1.0	2.0
1 Liquid, 1 Illiquid	8.851	6.105	3.261
1 Liquid, 1 Flexible, 1 Illiquid	8.919	6.137	3.274
---- Penalty π_2^*	0.13	0.13	0.11
---- Leakage Rate	0.89	0.90	0.99
General Mechanism (Relaxed)	8.932	6.144	3.278

(a) Variation of the coefficient of relative risk aversion γ

	Value of σ_μ		
	0.30	0.25	0.20
1 Liquid, 1 Illiquid	5.918	6.105	6.323
1 Liquid, 1 Flexible, 1 Illiquid	5.958	6.137	6.344
---- Penalty π_2^*	0.14	0.13	0.12
---- Leakage Rate	0.84	0.90	0.89
General Mechanism (Relaxed)	5.966	6.144	6.349

(b) Variation of the standard deviation σ_μ of the taste shock

	Value of σ_β			
	$+\infty$	1.0	0.5	0.0
1 Liquid, 1 Illiquid	6.105	6.019	5.772	2.542
1 Liquid, 1 Flexible, 1 Illiquid	6.137	6.053	5.810	2.841
---- Penalty π_2^*	0.13	0.13	0.14	0.36
---- Leakage Rate	0.90	0.90	0.90	0.73
General Mechanism (Relaxed)	6.144	6.060	5.819	2.881

(c) Variation of the standard deviation σ_β of the present bias distribution

Table 3: Robustness checks for welfare gains, optimal penalties and leakage rates. In each subtable: row 1 contains welfare gains for a two-account system with one completely liquid account and one completely illiquid account; row 2 contains welfare gains for a three-account system with one completely liquid account, one account with an endogenous penalty and one completely illiquid account; rows 3 and 4 contain the optimal penalty and leakage rate from the endogenous-penalty account associated with the system in row 2; and row 5 contains welfare gains for the general mechanism (relaxed), which allows for an arbitrary non-linear budget set and does not impose the monotonicity restriction. Table 3a varies the value of the coefficient of relative risk aversion γ . Table 3b varies the parameter σ_θ of the truncated-normal distribution of θ . Table 3c varies the parameter σ_β of the truncated-normal distribution of β .

6.1 How our normative framework deviates from the U.S. retirement savings system

Our model predicts that the planner should set up a completely illiquid account and populate it with enough assets so that, in equilibrium, there is at most a small drop in consumption at retirement. For example, under the optimal policy with a three-account system—one completely liquid, one partially illiquid (with an optimized 0.13 early-withdrawal penalty), and one completely illiquid—the respective allocations are 36.4%, 16.2%, and 47.4% of assets. Accordingly, the completely illiquid account *alone* (with 47.4% of total assets) is sufficient to generate nearly equal consumption in periods 1 and 2, even if the household consumes all of its completely liquid and partially illiquid assets in period 1. The high level of completely illiquid retirement savings explains the high level of equilibrium leakage from the partially illiquid account (in period 1). The partially illiquid account is a source of retirement consumption that can be used to supplement the consumption that will be generated by the assets in the completely illiquid account. Because the mandatory, completely illiquid retirement savings is so large (at the social optimum), households are not strongly motivated to preserve the assets in the partially illiquid account until retirement. Accordingly, the equilibrium leakage rate from the partially illiquid account is 90.2%.¹²

In the United States, the *actual* allocation to completely illiquid accounts is far lower than our optimized policy implies (e.g., mandatory savings is not sufficient to generate approximate consumption smoothing on its own in the United States), implying that the partially illiquid account plays a more important role in practice than it does in our model. To illustrate this point, we report another calibration of the model where we exogenously set the account balance allocations (rather than endogenously optimizing them) to better reflect the operation of the status quo system in the United States. When we exogenously allocate 50% of assets to the liquid account, 25% of assets to the partially illiquid account (with a 0.10 early withdrawal penalty to match

the U.S. retirement savings system), and 25% of assets to the completely illiquid account, we obtain a leakage rate from the partially illiquid account of 45.6%, which is close to the leakage rate in the United States (see Argento, Bryant and Sabelhaus, 2014).

7. Conclusions and directions for future work

To summarize, we focus on the case in which agents have heterogeneous present bias and the planner can implement mechanisms with inter-household transfers. Three findings emerge from our analysis:

1. The constrained-efficient social optimum is well-approximated by a two-account system, with one account that is completely liquid and a second account that is completely illiquid. Little welfare gain is obtained by moving beyond this simple two-account system. Accordingly, the two-account system identified in AWA (in a model with homogeneous β and no inter-household transfers) turns out to be approximately optimal in our new setting (with heterogeneous β and inter-household transfers).
2. If a third account is added, its optimized early-withdrawal penalty is only slightly above 10%.
3. In equilibrium, the leakage rate from this (partially illiquid) third account is high. We report a range of equilibrium leakage rates, depending on the calibration. With optimal allocations to all three accounts—completely liquid, partially illiquid, and completely illiquid—equilibrium leakage rates from the partially illiquid account range from 73% to 99%. By contrast, when we calibrate the model to match actual empirical allocations to the completely illiquid account (e.g., treating Social Security as the *empirical* analog of the model's completely illiquid account), the implied equilibrium leakage rate from the partially illiquid account drops to 46%.

These properties have analogs in the U.S. retirement savings system. The United States has completely liquid

¹² The high leakage rate implies that the partially illiquid account has very little impact on almost all households (relative to a world in which the funds from the partially illiquid account were instead put in the liquid account). This explains why the partially illiquid account has such a small effect on total social welfare relative to the two-account benchmark, with a completely liquid account and a completely illiquid account.

accounts (e.g., a standard checking account), completely illiquid accounts (e.g., Social Security), and a partially illiquid defined-contribution system with a 10% penalty for early withdrawals (e.g., an IRA or a 401(k)). This partially illiquid DC system has a leakage rate of approximately 40% (see Argento, Bryant and Sabelhaus, 2014).

Despite these similarities, it is inappropriate to conclude that our findings demonstrate the social optimality of the U.S. system. Most importantly, our theoretical model is highly stylized. First, we assume only two periods (e.g., working life and retirement).¹³ Second, we assume a particular form of multiplicative taste shifter¹⁴ and a limited set of distributions of that taste shifter. Third, we assume that households are naive with respect to their present bias parameter, β . Fourth, we study a limited set of distributions of this parameter.¹⁵ Fifth, we assume a particular conceptual formulation of self-defeating behavior (present bias).¹⁶

Moreover, our simulations imply that retirement consumption should not be allowed to fall far below working life consumption (recall that the illiquid account has a high funding level when we calculate the socially optimal system). In the actual data on U.S. households,

consumption proxies appear to decline between working life and retirement,¹⁷ raising the normative possibility that mandatory savings might be underutilized in the U.S.¹⁸ However, there is an active debate about both the existence and normative interpretation of the observed distribution of consumption changes for households transitioning from work life into retirement.¹⁹

Our normative result on underutilization of mandatory savings is closely related to a similar result reported in Moser and Olea de Souza e Silva (2017). Though many elements of the two models differ, both models assume that agents are present biased, and they both imply that optimal savings mechanisms are characterized by more mandatory savings than currently exists in the U.S. system.

Much more robustness work is needed to interrogate the three findings that we summarized above, as well as the additional finding that more mandatory savings would be socially optimal. It is not yet clear whether these results will continue to hold as future research enriches and expands this type of analysis.

¹³ Infinite horizon problems introduce technical challenges with respect to multiple equilibria. However, there has been progress on this issue. For example, see Harris and Laibson (2012) and Cao and Werning (2018).

¹⁴ We assume $\theta u(c)$, but we could have instead assumed $u(c - \theta)$.

¹⁵ Research is only beginning on the distribution of present bias. For analysis of this issue, see Lockwood (2016), Moser and Olea de Souza e Silva (2017), and Cohen, et al., (2019).


¹⁶ Other models of self-control include Thaler and Shefrin (1981), Gul and Pesendorfer (2001), Bernheim and Rangel (2004), O'Donoghue and Loewenstein (2004), Fudenberg and Levine (2006). See Ericson and Laibson (2019) for a review of this wider class of 'present-focused' models.

¹⁷ See Stephens and Toohy (2018).

References

- Alcott, Hunt, Benjamin B. Lockwood, and Dmitry Taubinsky, 2019. "Regressive Sin Taxes, with an Application to the Optimal Soda Tax," *Quarterly Journal of Economics* 134, no. 3: 1557-1626.
- Amador, Manuel, Ivan Werning, and George-Marios Angeletos, 2006. "Commitment vs. Flexibility." *Econometrica* 74, pp. 365-396.
- Ambrus, Attila, and Georgy Egorov. 2013. "Comment on 'Commitment vs. Flexibility'" *Econometrica* 81.5: 2113-2124.
- Argento, Robert, Victoria L. Bryant and John Sabelhaus, 2014, "Early Withdrawals from Retirement Accounts During the Great Recession," *Contemporary Economic Policy*.
- Atkeson, Andrew, and Robert E. Lucas. "On efficient distribution with private information." *The Review of Economic Studies* 59.3 (1992): 427-453.
- Augenblick, Ned, Muriel Niederle, and Charles Sprenger, 2015. "Working over time: Dynamic inconsistency in real effort tasks." *The Quarterly Journal of Economics* 130, no. 3: 1067-1115.
- Bernheim, B. Douglas, and Antonio Rangel. 2004. "Addiction and cue-triggered decision processes." *American Economic Review* 94, no. 5: 1558-1590.
- Bernheim, B. Douglas and Dmitry Taubinsky. 2018. "Behavioral Public Economics" in B. Douglas Bernheim, Stefano DellaVigna, and David Laibson (eds.), *Handbook of Behavioral Economics*, Volume 1, New York: Elsevier: 381-516.
- Beshears, John, James J. Choi, Joshua Hurwitz, David Laibson, and Brigitte C. Madrian. 2015. "Liquidity in retirement savings systems: An international comparison." *American Economic Review* 105, no. 5: 420-25.
- Beshears, John, James J. Choi, David Laibson, & Brigitte Madrian, 2018. "Behavioral household finance" in B. Douglas Bernheim, Stefano Dellavigna, and David Laibson Eds. *Handbook of Behavioral Economics - Foundations and Applications 1*. (Volume 1) Elsevier.
- Beshears, John, James Choi, Christopher Harris, David Laibson, Brigitte Madrian, Jung Sakong. 2020. "Which Early Withdrawal Penalty Attracts the Most Deposits to a Commitment Savings Account?" *Journal of Public Economics*.
- Camerer, Colin, Samuel Issacharoff, George Loewenstein, Ted O'Donoghue & Matthew Rabin. 2003. "Regulation for Conservatives: Behavioral Economics and the Case for 'Asymmetric Paternalism'." 151 *University of Pennsylvania Law Review* 101: 1211—1254.
- Cao, Dan, and Iván Werning, 2018. "Saving and dissaving with hyperbolic discounting." *Econometrica* 86.3: 805-857.
- Cohen, Jonathan D., Keith Marzilli Ericson, David Laibson, and John Myles White, 2020. "Measuring Time Preferences," *Journal of Economic Literature* 58, no. 2: 299-347.
- DellaVigna, Stefano and Ulrike Malmendier, 2006. "Paying Not to Go to the Gym." *American Economic Review* 96, pp. 694-719.
- DellaVigna, Stefano and Daniele Paserman, 2005. "Job Search and Impatience." *Journal of Labor Economics* 23, pp. 527-588.
- Ericson, Keith Marzilli, and David Laibson. 2019. "Intertemporal choice," in B. Douglas Bernheim, Stefano DellaVigna, and David Laibson (eds.), *Handbook of Behavioral Economics*, Volume 2, New York: Elsevier.
- Farhi, Emmanuel, and Xavier Gabaix, 2018. "Optimal Taxation with Behavioral Agents." Working paper.

- Fudenberg, Drew and David K. Levine, 2006. "A Dual-Self Model of Impulse Control." *American Economic Review* 96, pp. 1449-1476.
- Gine, Xavier, Dean Karlan, and Jonathan Zinman, 2010. "Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation." *American Economic Journal: Applied Economics* 2, pp. 213-235.
- Gruber, Jonathan, and Botond Köszegi, 2001. "Is addiction 'rational'? Theory and evidence." *The Quarterly Journal of Economics* 116.4: 1261-1303.
- Gruber, Jonathan, and Botond Köszegi, 2004. "Tax incidence when individuals are time-inconsistent: the case of cigarette excise taxes." *Journal of Public Economics* 88.9-10: 1959-1987.
- Gul, Faruk and Wolfgang Pesendorfer, 2001. "Temptation and Self-Control." *Econometrica* 69, pp. 1403-1435.
- Harris, Christopher, and David Laibson, 2012. "Instantaneous gratification." *The Quarterly Journal of Economics* 128.1 (2012): 205-248.
- Hewitt Associates, 2009. "The Erosion of Retirement Security from Cash-outs: Analysis and Recommendations." Available at http://www.aon.com/humancapital-consulting/thought-leadership/retirement/reports-pubs_retirement_cash_outs.jsp, accessed April 18, 2011.
- Laibson, David, 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112, pp. 443-477.
- Laibson, David I., Andrea Repetto, Jeremy Tobacman, 1998. "Self-control and saving for retirement." *Brookings Papers on Economic Activity*, no. 1: 91-196.
- Lockwood, Benjamin B., 2016. "Optimal income taxation with present bias." University of Pennsylvania Working Paper.
- Loewenstein, George and Ted O'Donoghue, 2004. "Animal spirits: Affective and deliberative processes in economic behavior." Available at SSRN 539843.
- Meier, Stephan and Charles Sprenger, 2010. "Present-Biased Preferences and Credit Card Borrowing." *American Economic Journal: Applied Economics* 2, pp. 193-210.
- Moser, Christian and Olea de Souza e Silva, Pedro, 2017. "Optimal Paternalistic Savings Policies" Columbia Business School Research Paper No. 17-51. Available at SSRN: <https://ssrn.com/abstract=2959844> or <http://dx.doi.org/10.2139/ssrn.2959844>.
- O'Donoghue, Ted and Matthew Rabin, 1999a. "Doing It Now or Later." *American Economic Review* 89, pp. 103-124.
- O'Donoghue, Ted and Matthew Rabin, 1999b. "Incentives for procrastinators." *The Quarterly Journal of Economics* 114:3, pp. 769-816.
- O'Donoghue, Ted, and Matthew Rabin, 2006. "Optimal sin taxes." *Journal of Public Economics* 90.10-11 (2006): 1825-1849.
- Phelps, E. S. and R. A. Pollak, 1968. "On Second-Best National Saving and Game-Equilibrium Growth." *Review of Economic Studies* 35, pp. 185-199.
- Shapiro, Jesse, 2005. "Is There a Daily Discount Rate? Evidence from the Food Stamp Nutrition Cycle." *Journal of Public Economics* 89, pp. 303-325.
- Stephens Jr, Melvin and Toohey, Desmond, 2018. "Changes in Nutrient Intake at Retirement" (No. w24621). National Bureau of Economic Research.
- Steyer, Robert, 2011. "DC Plan Leakage Problem Alarming, Solutions Evasive." *Pensions & Investments*, April 4, 2011.



Strotz, R. H., 1956. "Myopia and Inconsistency in Dynamic Utility Maximization." *Review of Economic Studies* 23, pp. 165-180.

Thaler, Richard H. and H. M. Shefrin, 1981. "An Economic Theory of Self-Control." *Journal of Political Economy* 89, pp. 392-406.

Online Appendices

A Proofs of Theorems 1 and 2

A.1 Introduction

We assume that: $u_1, u_2 : (0, \infty) \rightarrow (-\infty, \infty)$; $u'_1, u'_2 > 0$ and $u''_1, u''_2 < 0$ on $(0, \infty)$; and $u'_1(0+) = u'_2(0+) = \infty$. Next, in order to formulate our assumptions on the distribution function F of the taste shock θ , we introduce the function Γ defined by the formula

$$\Gamma(\theta) = (1 - \beta) \theta F'(\theta) + F(\theta).$$

Our assumptions on F are then:

A1 Both F and F' are functions of bounded variation on $(0, \infty)$.

A2 The support of F' is contained in $[\underline{\theta}, \bar{\theta}]$, where $0 < \underline{\theta} < \bar{\theta} < \infty$.

A3 There exists $\theta_M \in [\underline{\theta}, \bar{\theta}]$ such that: (i) $\Gamma' \geq 0$ on $(0, \theta_M)$; and (ii) $\Gamma' \leq 0$ on (θ_M, ∞) .

A4 F' is bounded away from 0 on $(\underline{\theta}, \bar{\theta})$. In particular, $F'(\underline{\theta}+) > 0$, $F'(\bar{\theta}-) > 0$.

Finally, we assume that $0 < \beta < 1$.

Under these assumptions, we have:

Theorem 5 *Suppose that interpersonal transfers are not possible. Then welfare is maximized by giving self 1 two accounts: a completely liquid account and a completely illiquid account.*

This result is proved (under slightly different assumptions) in Amador et al (2006), and in Beshears et al (2020).

We also have:

Theorem 6 *Suppose that interpersonal transfers are possible. Then a two-account system with one completely liquid account and one completely illiquid account never maximizes welfare.*

This result is proved in the current appendix.

A.2 The Optimization Problem of the Planner

If self 1 is presented with two accounts, a perfectly liquid account containing the amount $x_{\text{liquid}} > 0$ and a perfectly illiquid account containing the amount $x_{\text{illiquid}} \geq 0$, then the outcome will depend on her type θ . There will exist $\theta_1 \in (0, \infty)$ such that: if $\theta < \theta_1$, then she consumes less than the balance x_{liquid} in her liquid account: and, if $\theta \geq \theta_1$, then she consumes the whole of x_{liquid} . The cutoff θ_1 need not lie in $[\underline{\theta}, \bar{\theta}]$. It could be that $\theta_1 < \underline{\theta}$, in which case there will be perfect pooling: all types will consume the whole of x_{liquid} and both c_1 and c_2 will be constant. Or it could be that $\theta_1 > \bar{\theta}$, in which case there will be perfect separation: all types will consume less than x_{liquid} , c_1 will be strictly increasing in θ and c_2 will be strictly decreasing in θ .

More generally, we will obtain consumption allocations $c_1, c_2 : \Theta \rightarrow (0, \infty)$ and associated utility allocations $r_1, r_2 : \Theta \rightarrow \mathbb{R}$, where the latter are given by the formulae $r_1(\theta) = u_1(c_1(\theta))$ and $r_2(\theta) = u_2(c_2(\theta))$. The overall utility allocation (r_1, r_2) will be a smooth function of θ for $\theta < \theta_1$, have a kink at θ_1 , and be constant for $\theta > \theta_1$. The idea behind the proof is to find necessary conditions for utility allocations of this type to be optimal, and to use these necessary conditions to derive a contradiction.

The first step is to formulate the optimization problem of the planner. We do this in terms of general utility allocations $v_1, v_2 : \Theta \rightarrow \mathbb{R}$, reserving the notation r_1, r_2 for the specific allocations arising from two-account systems with one completely liquid account and one completely illiquid account. Accordingly, the planner seeks to maximize social welfare

$$\int (\theta v_1(\theta) + v_2(\theta)) dF(\theta)$$

over utility allocations

$$(v_1, v_2) : [\underline{\theta}, \bar{\theta}] \rightarrow (u_1(0+), u_1(\infty-)) \times (u_2(0+), u_2(\infty-))$$

subject to aggregate budget balance and incentive compatibility. Aggregate budget balance can be expressed in the form

$$\int (Y - C_1(v_1(\theta)) - C_2(v_2(\theta))) dF(\theta) \geq 0, \quad (\text{BC})$$

where $C_t = u_t^{-1}$ for $t \in \{1, 2\}$. Incentive compatibility breaks down into two parts,

a linear part

$$\theta v'_1 + \beta v'_2 = 0 \quad (\text{ICL})$$

and a monotonic part

$$v'_2 \leq 0. \quad (\text{ICM})$$

Remark 7 *The two conditions (ICL) and (ICM) are simply the differential counterpart of the usual integral representation of incentive compatibility in a mechanism-design problem.*

Remark 8 *Since incentive-compatible utility allocations v_1, v_2 are monotonic, in the sense that v_1 is non-decreasing and v_2 is non-increasing, it makes sense to restrict v_1 and v_2 to the space of functions of bounded variation. If this is done, then v'_1 and v'_2 are both measures, and (ICL) and (ICM), therefore, both make sense.*

A.3 The Case $\theta_1 \in (\underline{\theta}, \bar{\theta})$

Consider the case in which x_{liquid} and x_{illiquid} are such that $\theta_1 \in (\underline{\theta}, \bar{\theta})$. In this case, the second step is to parameterize candidate solutions to planner's problem in terms of $v_1(\bar{\theta})$, $v_2(\bar{\theta})$ and continuous functions $v'_{1L} : [\underline{\theta}, \theta_1] \rightarrow \mathbb{R}$ and $v'_{1R} : [\theta_1, \bar{\theta}] \rightarrow \mathbb{R}$. More precisely, we can put:

1. $v_1(\theta) = v_1(\bar{\theta}) - \int_{\theta}^{\bar{\theta}} v'_{1R}(t) dt$ for $\theta \in [\theta_1, \bar{\theta}]$;
2. $v_1(\theta) = v_1(\theta_1) - \int_{\theta}^{\theta_1} v'_{1L}(t) dt$ for $\theta \in [\underline{\theta}, \theta_1]$;
3. $v'_{2R}(\theta) = -\frac{\theta}{\beta} v'_{1R}(\theta)$ for $\theta \in [\theta_1, \bar{\theta}]$;
4. $v'_{2L}(\theta) = -\frac{\theta}{\beta} v'_{1L}(\theta)$ for $\theta \in [\underline{\theta}, \theta_1]$;
5. $v_2(\theta) = v_2(\bar{\theta}) - \int_{\theta}^{\bar{\theta}} v'_{2R}(t) dt$ for $\theta \in [\theta_1, \bar{\theta}]$;
6. $v_2(\theta) = v_2(\theta_1) - \int_{\theta}^{\theta_1} v'_{2L}(t) dt$ for $\theta \in [\underline{\theta}, \theta_1]$.

In other words: v_1 is the function with derivative v'_{1L} on $[\underline{\theta}, \theta_1)$, derivative v'_{1R} on $(\theta_1, \bar{\theta}]$ and value $v_1(\bar{\theta})$ at $\bar{\theta}$; and v_2 is the function with derivative v'_{2L} on $[\underline{\theta}, \theta_1)$, derivative v'_{2R} on $(\theta_1, \bar{\theta}]$ and value $v_2(\bar{\theta})$ at $\bar{\theta}$.

Remark 9 *The logic behind this parameterization is to consider v'_1 as a function of bounded variation on $[\underline{\theta}, \bar{\theta}]$. In this case, the left- and right-hand limits of v'_1 are well defined at all points of $(\underline{\theta}, \bar{\theta})$. In particular, $v'_1(\theta_1+)$ and $v'_1(\theta_1-)$ are both well defined, and we can interpret $v'_{1R}(\theta_1)$ as $v'_1(\theta_1+)$ and $v'_{1L}(\theta_1)$ as $v'_1(\theta_1-)$. Analogous remarks apply to v'_2 .*

The third step is to formulate the Lagrangean. This can be written

$$\begin{aligned}
L(v_1(\bar{\theta}), v_2(\bar{\theta}), v'_{1L}, v'_{1R}, \lambda, \zeta_L, \zeta_R) &= \int (\theta v_1(\theta) + v_2(\theta)) dF(\theta) \\
&+ \lambda \int (Y - C_1(v_1(\theta)) - C_2(v_2(\theta))) dF(\theta) \\
&- \int_{[\underline{\theta}, \theta_1]} v'_{2L}(\theta) d\zeta_L(\theta) \\
&- \int_{[\theta_1, \bar{\theta}]} v'_{2R}(\theta) d\zeta_R(\theta), \tag{7}
\end{aligned}$$

where:

1. the arguments of L are the parameters $v_1(\bar{\theta})$, $v_2(\bar{\theta})$, v'_{1L} and v'_{1R} , and the multipliers λ , ζ_L and ζ_R ;
2. λ is a scalar (namely the multiplier on the aggregate budget constraint);
3. ζ_L is a finite non-negative Borel measure on $[\underline{\theta}, \theta_1]$ (namely the multiplier associated with the non-positivity constraint on v'_{2L});
4. ζ_R is a finite non-negative Borel measure on $[\theta_1, \bar{\theta}]$ (namely the multiplier associated with the non-positivity constraint on v'_{2R});
5. the variables v_1 , v_2 , v'_{2L} and v'_{2R} on the right-hand side are determined by the parameters $v_1(\bar{\theta})$, $v_2(\bar{\theta})$, v'_{1L} and v'_{1R} as explained above.

Remark 10 *The Lagrangean does not include a term corresponding to (ICL). This is because we have used (ICL) to solve for v'_{2L} and v'_{2R} in terms of v'_{1L} and v'_{1R} .*

The fourth step is to note that we can associate parameters $(r_1(\bar{\theta}), r_2(\bar{\theta}), r'_{1L}, r'_{1R})$ with the reference utility curve (r_1, r_2) in the obvious way, and take the derivative of the Lagrangean at the parameter values $(r_1(\bar{\theta}), r_2(\bar{\theta}), r'_{1L}, r'_{1R})$ in the direction of the alternative parameter values $(s_1(\bar{\theta}), s_2(\bar{\theta}), s'_{1L}, s'_{1R})$. Furthermore, this

calculation can be simplified by noting that the variables $(v_1, v_2, v'_{2L}, v'_{2R})$ in the RHS of the equation for the Lagrangean are linear in the underlying parameters $(v_1(\bar{\theta}), v_2(\bar{\theta}), v'_{1L}, v'_{1R})$. Hence we can simply take the derivative of the RHS at the point $(r_1, r_2, r'_{2L}, r'_{2R})$ in the direction of the point $(s_1, s_2, s'_{2L}, s'_{2R})$ and only then substitute for $(s_1, s_2, s'_{2L}, s'_{2R})$ in terms of $(s_1(\bar{\theta}), s_2(\bar{\theta}), s'_{1L}, s'_{1R})$.

Taking the derivative of the RHS at the point $(r_1, r_2, r'_{2L}, r'_{2R})$ in the direction of the point $(s_1, s_2, s'_{2L}, s'_{2R})$, we obtain

$$0 = \int (\theta s_1 + s_2) dF - \lambda \int (C'_1(r_1) s_1 + C'_2(r_2) s_2) dF - \int_{[\underline{\theta}, \theta_1]} s'_{2L}(\theta) d\zeta_L(\theta) - \int_{[\theta_1, \bar{\theta}]} s'_{2R}(\theta) d\zeta_R(\theta) \quad (8)$$

for all feasible $(s_1, s_2, s'_{2L}, s'_{2R})$. Moreover, the constraints must all be satisfied. That is,

$$\begin{aligned} 0 &= \int (Y - C_1(r_1(\theta)) - C_2(r_2(\theta))) dF(\theta), \\ 0 &\geq r'_{2L}, \\ 0 &\geq r'_{2R}. \end{aligned}$$

Finally, constraint qualification must hold. That is,

$$0 = \int_{[\underline{\theta}, \theta_1]} r'_{2L}(\theta) d\zeta_L(\theta), \quad (9)$$

$$0 = \int_{[\theta_1, \bar{\theta}]} r'_{2R}(\theta) d\zeta_R(\theta). \quad (10)$$

Furthermore, a variation $(s_1, s_2, s'_{2L}, s'_{2R})$ is feasible iff it can be expressed in terms of the underlying parameters $(s_1(\bar{\theta}), s_2(\bar{\theta}), s'_{1L}, s'_{1R})$. We therefore substitute for the variation $(s_1, s_2, s'_{2L}, s'_{2R})$ in terms of the underlying parameters $(s_1(\bar{\theta}), s_2(\bar{\theta}), s'_{1L}, s'_{1R})$ and manipulate the RHS in such a way as to expose the linear dependence of the RHS on $s_1(\bar{\theta}), s_2(\bar{\theta}), s'_{1L}$ and s'_{1R} .

The first contribution to the RHS is $\int \theta s_1 dF(\theta)$. Putting $\bar{F}(\theta) = \int_{[\underline{\theta}, \theta]} F(t) dt$, and noting that $\theta F - \bar{F}$ and s_1 are both continuous, we can integrate this contri-

bution by parts to obtain

$$\begin{aligned}
\int \theta s_1 dF(\theta) &= [(\theta F - \bar{F}) s_1]_{\underline{\theta}-}^{\bar{\theta}} - \int (\theta F - \bar{F}) s'_1 d\theta \\
&= (\bar{\theta} F(\bar{\theta}) - \bar{F}(\bar{\theta})) s_1(\bar{\theta}) - \int (\theta F - \bar{F}) s'_1 d\theta \\
&= (\bar{\theta} F(\bar{\theta}) - \bar{F}(\bar{\theta})) s_1(\bar{\theta}) \\
&\quad - \int_{[\underline{\theta}, \theta_1]} (\theta F - \bar{F}) s'_{1L} d\theta - \int_{[\theta_1, \bar{\theta}]} (\theta F - \bar{F}) s'_{1R} d\theta.
\end{aligned}$$

The second contribution to the RHS is $\int s_2 dF(\theta)$. For this contribution, we have

$$\begin{aligned}
\int s_2 dF(\theta) &= [F s_2]_{\underline{\theta}-}^{\bar{\theta}} - \int F s'_2 d\theta \\
&= F(\bar{\theta}) s_2(\bar{\theta}) - \int F s'_2 d\theta \\
&= F(\bar{\theta}) s_2(\bar{\theta}) - \int_{[\underline{\theta}, \theta_1]} F s'_{2L} d\theta - \int_{[\theta_1, \bar{\theta}]} F s'_{2R} d\theta \\
&= F(\bar{\theta}) s_2(\bar{\theta}) + \int_{[\underline{\theta}, \theta_1]} F \frac{\theta}{\beta} s'_{1L} d\theta + \int_{[\theta_1, \bar{\theta}]} F \frac{\theta}{\beta} s'_{1R} d\theta.
\end{aligned}$$

Next, putting $\Lambda_1(\theta) = \int_{[\underline{\theta}, \theta]} C'_1(r_1(t)) dF(t)$, we have

$$\begin{aligned}
-\lambda \int C'_1(r_1) s_1 dF &= - \int s_1 \lambda \Lambda'_1 d\theta \\
&= -[s_1 \lambda \Lambda_1]_{\underline{\theta}-}^{\bar{\theta}} + \int \lambda \Lambda_1 s'_1 d\theta \\
&= -s_1(\bar{\theta}) \lambda \Lambda_1(\bar{\theta}) + \int \lambda \Lambda_1 s'_1 d\theta \\
&= -s_1(\bar{\theta}) \lambda \Lambda_1(\bar{\theta}) \\
&\quad + \int_{[\underline{\theta}, \theta_1]} \lambda \Lambda_1 s'_{1L} d\theta + \int_{[\theta_1, \bar{\theta}]} \lambda \Lambda_1 s'_{1R} d\theta.
\end{aligned}$$

Similarly, putting $\Lambda_2(\theta) = \int_{[\underline{\theta}, \theta]} C'_2(r_2(t)) dF(t)$,

$$\begin{aligned}
-\lambda \int C'_2(r_2) s_2 dF &= - \int s_2 \lambda \Lambda'_2 d\theta \\
&= - [s_2 \lambda \Lambda_2]_{\underline{\theta}}^{\bar{\theta}} + \int \lambda \Lambda_2 s'_2 d\theta \\
&= -s_2(\bar{\theta}) \lambda \Lambda_2(\bar{\theta}) + \int \lambda \Lambda_2 s'_2 d\theta \\
&= -s_2(\bar{\theta}) \lambda \Lambda_2(\bar{\theta}) \\
&\quad + \int_{[\underline{\theta}, \theta_1]} \lambda \Lambda_2 s'_{2L} d\theta + \int_{[\theta_1, \bar{\theta}]} \lambda \Lambda_2 s'_{2R} d\theta \\
&= -s_2(\bar{\theta}) \lambda \Lambda_2(\bar{\theta}) \\
&\quad - \int_{[\underline{\theta}, \theta_1]} \lambda \Lambda_2 \frac{\theta}{\beta} s'_{1L} d\theta - \int_{[\theta_1, \bar{\theta}]} \lambda \Lambda_2 \frac{\theta}{\beta} s'_{1R} d\theta.
\end{aligned}$$

Finally, we have

$$- \int_{[\underline{\theta}, \theta_1]} s'_{2L}(\theta) d\zeta_L(\theta) = \int_{[\underline{\theta}, \theta_1]} \frac{\theta}{\beta} s'_{1L}(\theta) d\zeta_L(\theta)$$

and

$$- \int_{[\theta_1, \bar{\theta}]} s'_{2R}(\theta) d\zeta_R(\theta) = \int_{[\theta_1, \bar{\theta}]} \frac{\theta}{\beta} s'_{1R}(\theta) d\zeta_R(\theta).$$

The fifth step is to equate the coefficients of $s_1(\bar{\theta})$, $s_2(\bar{\theta})$, s'_{1L} and s'_{1R} to 0. Doing so yields:

$$0 = \bar{\theta} F(\bar{\theta}) - \bar{F}(\bar{\theta}) - \lambda \Lambda_1(\bar{\theta}), \quad (11)$$

$$0 = F(\bar{\theta}) - \lambda \Lambda_2(\bar{\theta}), \quad (12)$$

$$0 = -(\theta F - \bar{F}) d\theta + \frac{\theta}{\beta} F d\theta + \lambda \Lambda_1 d\theta - \frac{\theta}{\beta} \lambda \Lambda_2 d\theta + \frac{\theta}{\beta} d\zeta_L, \quad (13)$$

$$0 = -(\theta F - \bar{F}) d\theta + \frac{\theta}{\beta} F d\theta + \lambda \Lambda_1 d\theta - \frac{\theta}{\beta} \lambda \Lambda_2 d\theta + \frac{\theta}{\beta} d\zeta_R. \quad (14)$$

Now, we certainly have $r'_{2L} < 0$ on $[\underline{\theta}, \theta_1]$. (This is because, if $\theta < \theta_1$, then self 1 consumes less than x_{liquid} . Hence $r'_{1L} > 0$ and $r'_{2L} < 0$.) It therefore follows from constraint qualification (namely (9)) that $\zeta_L = 0$. Equation (13) therefore implies that

$$\lambda(\theta \Lambda_2 - \beta \Lambda_1) = \theta F - \beta(\theta F - \bar{F}) = (1 - \beta)\theta F + \beta \bar{F} = \bar{\Gamma} \quad (15)$$

almost everywhere on $[\underline{\theta}, \theta_1]$, where $\Gamma = (1 - \beta) \theta F' + F$ and $\bar{\Gamma}(\theta) = \int_{[\underline{\theta}, \theta]} \Gamma(t) dt$. Furthermore, since F' is of bounded variation,

$$\begin{aligned} \frac{\theta \Lambda_2(\theta)}{\theta - \underline{\theta}} &\rightarrow \underline{\theta} C'_2(r_2(\underline{\theta})) F'(\underline{\theta}+), \\ \frac{\beta \Lambda_1(\theta)}{\theta - \underline{\theta}} &\rightarrow \beta C'_1(r_1(\underline{\theta})) F'(\underline{\theta}+), \\ \frac{\bar{\Gamma}}{\theta - \underline{\theta}} &\rightarrow \Gamma(\underline{\theta}+) = (1 - \beta) \underline{\theta} F'(\underline{\theta}+) \end{aligned}$$

as $\theta \downarrow \underline{\theta}$. But, since $(r_1(\underline{\theta}), r_2(\underline{\theta}))$ is chosen freely from the ambient budget line by the $\underline{\theta}$ type, we must have

$$\frac{C'_1(r_1(\underline{\theta}))}{\underline{\theta}} = \frac{C'_2(r_2(\underline{\theta}))}{\beta}.$$

We, therefore, have

$$\frac{\theta \Lambda_2(\theta) - \beta \Lambda_1(\theta)}{\theta - \underline{\theta}} \rightarrow 0$$

as $\theta \downarrow \underline{\theta}$. On the other hand,

$$\frac{\bar{\Gamma}}{\theta - \underline{\theta}} \rightarrow (1 - \beta) \underline{\theta} F'(\underline{\theta}+) > 0$$

as $\theta \downarrow \underline{\theta}$. Passing to the limit in equation (15), we therefore obtain

$$0 = (1 - \beta) \underline{\theta} F'(\underline{\theta}+).$$

Since the RHS is strictly positive, this is a contradiction. This establishes that we cannot have $\theta_1 \in (\underline{\theta}, \bar{\theta})$.

Remark 11 *We have used the assumption that $\beta < 1$.*

A.4 The Case $\theta_1 \in [\bar{\theta}, \infty)$

Consider now the case in which x_{liquid} and x_{illiquid} are such that $\theta_1 \in [\bar{\theta}, \infty)$. In this case, we can still derive equations (11, 12 and 13). In particular, we can still derive equation (13). We can therefore derive a contradiction by essentially the same argument.

A.5 The Case $\theta_1 \in (0, \underline{\theta}]$

Consider now the case in which x_{liquid} and x_{illiquid} are such that $\theta_1 \in (0, \underline{\theta}]$. In this case, we can still derive equations (11, 12 and 14). However, we can no longer derive equation (13). We therefore need new arguments. The first point to note is that, since $\theta_1 \leq \underline{\theta}$, all types $\theta \in [\underline{\theta}, \bar{\theta}]$ choose the point that a hypothetical θ_1 type would choose from the ambient budget set. We therefore have

$$\Lambda_1(\bar{\theta}) = \int_{[\underline{\theta}, \bar{\theta}]} C'_1(r_1(t)) dF(t) = F(\bar{\theta}) C'_1(r_1(\theta_1)), \quad (16)$$

$$\Lambda_2(\bar{\theta}) = \int_{[\underline{\theta}, \bar{\theta}]} C'_2(r_2(t)) dF(t) = F(\bar{\theta}) C'_2(r_2(\theta_1)). \quad (17)$$

Furthermore, since the θ_1 type chooses freely from the ambient budget set, we have

$$\frac{C'_1(r_1(\theta_1))}{\theta_1} = \frac{C'_2(r_2(\theta_1))}{\beta}.$$

Using (11) and (12), we therefore obtain

$$\frac{\bar{\theta} F(\bar{\theta}) - \bar{F}(\bar{\theta})}{F(\bar{\theta})} = \frac{\Lambda_1(\bar{\theta})}{\Lambda_2(\bar{\theta})} = \frac{C'_1(r_1(\theta_1))}{C'_2(r_2(\theta_1))} = \frac{\theta_1}{\beta}. \quad (18)$$

Hence

$$\begin{aligned} (\bar{\theta} - \theta_1) F(\bar{\theta}) &= \bar{\theta} F(\bar{\theta}) - \beta (\bar{\theta} F(\bar{\theta}) - \bar{F}(\bar{\theta})) \\ &= (1 - \beta) \bar{\theta} F(\bar{\theta}) + \beta \bar{F}(\bar{\theta}) \\ &= \bar{\Gamma}(\bar{\theta}), \end{aligned} \quad (19)$$

where Γ and $\bar{\Gamma}$ are as above.

Remark 12 *Bearing in mind that $\theta_1 \leq \underline{\theta}$, so that $\bar{\Gamma}(\theta_1) = 0$, this equation can also be written*

$$(\bar{\theta} - \theta_1) F(\bar{\theta}) = \bar{\Gamma}(\bar{\theta}) - \bar{\Gamma}(\theta_1)$$

or

$$\frac{1}{\bar{\theta} - \theta_1} \int_{[\theta_1, \bar{\theta}]} \Gamma(t) dt = F(\bar{\theta}).$$

That is, θ_1 satisfies the equation for the optimal cutoff in the problem without inter-

personal transfers. Cf. Amador et al (2006) and Beshears et al (2020).

Remark 13 That θ_1 should satisfy the equation for the optimal cutoff in the problem without interpersonal transfers makes perfect sense. If $\theta_1 \leq \underline{\theta}$ then all θ types make the same choice, and the optimum with interpersonal transfers happens not to involve any transfers. It must therefore also be the optimum without interpersonal transfers, and must therefore satisfy the equation for the optimal cutoff in the problem without interpersonal transfers.

However, we have not yet used equation (14). It follows from this equation that

$$d\zeta_R = \frac{\beta}{\theta} (\theta F - \bar{F}) d\theta - F d\theta + \lambda (\Lambda_2 - \frac{\beta}{\theta} \Lambda_1) d\theta.$$

In other words, ζ_R is absolutely continuous w.r.t. Lebesgue measure, with density

$$\zeta'_R = \frac{\beta}{\theta} (\theta F - \bar{F}) - F + \lambda (\Lambda_2 - \frac{\beta}{\theta} \Lambda_1).$$

Furthermore:

$$\begin{aligned} \Lambda_1(\theta) &= \int_{[\underline{\theta}, \theta]} C'_1(r_1(t)) dF(t) = F(\theta) C'_1(r_1(\theta_1)) = \frac{F(\theta)}{F(\bar{\theta})} \Lambda_1(\bar{\theta}) \\ &= \frac{F(\theta)}{F(\bar{\theta})} \frac{\theta_1}{\beta} \Lambda_2(\bar{\theta}) = \frac{F(\theta)}{F(\bar{\theta})} \frac{\theta_1}{\beta} \frac{F(\bar{\theta})}{\lambda} = \frac{\theta_1}{\beta} \frac{F(\theta)}{\lambda} \end{aligned}$$

(where the last line follows from (18) and (12)); and

$$\begin{aligned} \Lambda_2(\theta) &= \int_{[\underline{\theta}, \theta]} C'_2(r_2(t)) dF(t) = F(\theta) C'_2(r_2(\theta_1)) = \frac{F(\theta)}{F(\bar{\theta})} \Lambda_2(\bar{\theta}) \\ &= \frac{F(\theta)}{F(\bar{\theta})} \frac{F(\bar{\theta})}{\lambda} = \frac{F(\theta)}{\lambda} \end{aligned}$$

(where the last line follows from (12)). Hence

$$\lambda (\theta \Lambda_2 - \beta \Lambda_1) = (\theta - \theta_1) F(\theta)$$

and

$$\begin{aligned} \theta \zeta'_R &= \beta (\theta F - \bar{F}) - \theta F + (\theta - \theta_1) F \\ &= (\theta - \theta_1) F(\theta) - \bar{\Gamma}. \end{aligned}$$

Now, $F(\underline{\theta}) = \bar{\Gamma}(\underline{\theta}) = 0$. Hence $\underline{\theta} \zeta'_R(\underline{\theta}) = 0$. Furthermore, we must have $\theta \zeta'_R \geq 0$ on $(\underline{\theta}, \bar{\theta})$. Hence

$$\frac{\theta \zeta'_R(\theta) - \underline{\theta} \zeta'_R(\underline{\theta})}{\theta - \underline{\theta}} \geq 0.$$

Letting $\theta \rightarrow \underline{\theta}+$, we therefore obtain

$$(\theta \zeta'_R)'(\underline{\theta}+) = (\beta \underline{\theta} - \theta_1) F'(\underline{\theta}+) \geq 0.$$

Since $F'(\underline{\theta}+) > 0$, it follows that $\theta_1 \leq \beta \underline{\theta}$. Similarly, (19) implies that $(\bar{\theta} - \theta_1) F'(\bar{\theta}) - \bar{\Gamma}(\bar{\theta}) = 0$. Hence $\bar{\theta} \zeta'_R(\bar{\theta}) = 0$. Hence

$$\frac{\bar{\theta} \zeta'_R(\bar{\theta}) - \theta \zeta'_R(\theta)}{\bar{\theta} - \theta} \leq 0.$$

Letting $\theta \rightarrow \bar{\theta}-$, we therefore obtain

$$(\theta \zeta'_R)'(\bar{\theta}-) = (\beta \bar{\theta} - \theta_1) F'(\bar{\theta}-) \leq 0.$$

Since $F'(\bar{\theta}-) > 0$, it follows that $\theta_1 \geq \beta \bar{\theta}$. These two inequalities on θ_1 are inconsistent, so we have a contradiction.

Remark 14 *This is where we use the assumption $\beta > 0$: we would not obtain a contradiction by combining two inequalities $\theta_1 \leq \beta \underline{\theta}$ and $\theta_1 \geq \beta \bar{\theta}$ if we had $\beta = 0$.*

Remark 15 *We can use the preceding analysis to obtain some perspective on why a pooling mechanism in which all resources are placed in the illiquid account is never optimal. Suppose that we replace the inequality constraint $0 \geq r'_{2R}$ with an equality constraint and choose the multiplier ζ_R in such a way that this constraint is respected. Then, proceeding almost exactly as above, we will obtain*

$$\begin{aligned} (\theta \zeta'_R)' &= (\theta - \theta_1) F' + F - \Gamma \\ &= (\theta - \theta_1) F' - (1 - \beta) \theta F' \\ &= (\beta \theta - \theta_1) F'. \end{aligned}$$

Moreover we will have the boundary conditions $\underline{\theta} \zeta'_R(\underline{\theta}) = 0$ and $\bar{\theta} \zeta'_R(\bar{\theta}) = 0$. It follows that $\theta_1 \in (\beta \underline{\theta}, \beta \bar{\theta})$ and $\theta \zeta'_R < 0$ on $(\underline{\theta}, \bar{\theta})$. Hence a small change in the direction of any incentive-compatible and fully separating mechanism is desirable.

(This would have the effect of reducing r'_2 from 0—and increasing r'_1 from 0—at all points in the range $(\underline{\theta}, \bar{\theta})$.) In other words, it is always desirable to allow some flexibility to the decision maker to respond to the information contained in θ .

B Analysis of a General (Non-Linear) Mechanism

B.1 The Mechanism-Design Problem

In the general mechanism-design problem, the planner chooses a budget set

$$C \subset (0, \infty)^2$$

and consumption allocations $c_1, c_2 : \Theta \times B \rightarrow (0, \infty)$ to maximize welfare

$$\int \int (\theta u_1(c_1(\theta, \beta)) + u_2(c_2(\theta, \beta))) f(\theta) g(\beta) d\theta d\beta$$

subject to the resource constraint

$$\int \int (Y - c_1(\theta, \beta) - \frac{1}{R} c_2(\theta, \beta)) f(\theta) g(\beta) d\theta d\beta \geq 0$$

and the incentive-compatibility constraint

$$(c_1(\theta, \beta), c_2(\theta, \beta)) \in \operatorname{argmax}_{(\tilde{c}_1, \tilde{c}_2) \in C} \{\theta u_1(\tilde{c}_1) + \beta u_2(\tilde{c}_2)\}.$$

Here, f is the density of θ (associated with distribution function F in the main text); g is the density of β (associated with distribution function G in the main text); Y is the per capita endowment; and R is the gross rate of return. Furthermore, we assume that: $\Theta = [\underline{\theta}, \bar{\theta}]$; $B = [\underline{\beta}, \bar{\beta}]$; $0 < \underline{\theta} < \bar{\theta} < \infty$; $0 < \underline{\beta} < \bar{\beta} < \infty$; f is continuous and bounded away from 0 on Θ ; g is continuous and bounded away from 0 on B .

Remark 16 For example: f might take the form

$$f(\theta) = \frac{\exp\left(-\frac{1}{2} \left(\frac{\theta - \mu}{\sigma}\right)^2\right)}{\int_{\underline{\theta}}^{\bar{\theta}} \exp\left(-\frac{1}{2} \left(\frac{\theta - \mu}{\sigma}\right)^2\right) d\theta} \quad \text{for } \theta \in [\underline{\theta}, \bar{\theta}]$$

and $f(\theta) = 0$ otherwise, i.e., f might be the density of the univariate normal distribution with mean μ and variance σ^2 truncated to the interval $[\underline{\theta}, \bar{\theta}]$; and g might take the form

$$g(\beta) = \frac{1}{\bar{\beta} - \underline{\beta}} \quad \text{for } \beta \in [\underline{\beta}, \bar{\beta}]$$

and $g(\beta) = 0$ otherwise, i.e., g might be the density of the uniform distribution on the interval $[\underline{\beta}, \bar{\beta}]$.

B.2 Transforming the Problem

The first step in solving this problem is to note that

$$(c_1, c_2) \in \operatorname{argmax}_{(\tilde{c}_1, \tilde{c}_2) \in C} \{\theta u_1(\tilde{c}_1) + \beta u_2(\tilde{c}_2)\}$$

iff

$$(c_1, c_2) \in \operatorname{argmax}_{(\tilde{c}_1, \tilde{c}_2) \in C} \left\{ \frac{\theta}{\beta} u_1(\tilde{c}_1) + u_2(\tilde{c}_2) \right\}.$$

The set of optimal choices of the individual therefore depends only on $\phi = \theta / \beta$. Combining this fact with the assumed continuity of the distribution functions F and G of θ and β implies that, if we put $\Phi = [\underline{\phi}, \bar{\phi}]$ where $\underline{\phi} = \underline{\theta} / \bar{\beta}$ and $\bar{\phi} = \bar{\theta} / \underline{\beta}$, then the planner can work with consumption allocations $c_1, c_2 : \Phi \rightarrow (0, \infty)$ instead of with consumption allocations $c_1, c_2 : \Theta \times B \rightarrow (0, \infty)$.

The second step is to note that we can work with utility allocations $v_1, v_2 : \Phi \rightarrow \mathbb{R}$ instead of with consumption allocations $c_1, c_2 : \Phi \rightarrow (0, \infty)$. The former are related to the latter via the formulae $v_1(\phi) = u_1(c_1(\phi))$ and $v_2(\phi) = u_2(c_2(\phi))$. We can also invert these formulae to get $c_1(\phi) = C_1(v_1(\phi))$ and $c_2(\phi) = C_2(v_2(\phi))$.

The third step is to note that we can change variables in the integral defining welfare and in the integral giving the resource constraint, replacing (θ, β) with (ϕ, β) .

At this point, the planner's problem can be expressed as that of choosing $v_1, v_2 : \Phi \rightarrow \mathbb{R}$ to maximize welfare

$$\int \int (\beta \phi v_1(\phi) + v_2(\phi)) \beta f(\beta \phi) g(\beta) d\phi d\beta$$

subject to the resource constraint

$$\int \int \left(Y - C_1(v_1(\phi)) - \frac{1}{R} C_2(v_2(\phi)) \right) \beta f(\beta \phi) g(\beta) d\phi d\beta \geq 0$$

and the incentive-compatibility constraint, which now has two parts, namely a linear part,

$$0 = \phi v_1'(\phi) + v_2'(\phi) \quad (\text{ICL})$$

and a monotonic part,

$$0 \leq -v_2'(\phi). \quad (\text{ICM})$$

Remark 17 Notice that, whenever c_1 and c_2 are chosen from a budget set C , v_1 will be non-decreasing and v_2 will be non-increasing. However, neither function need be differentiable (or even continuous). Hence the derivatives v_1' and v_2' might in principle be a non-negative and a non-positive measure respectively. This does not invalidate (IC1) or (IC2), both of which make sense for measures. However, in what follows, we will sometimes reason as if v_1' and v_2' exist in the usual sense.

The fourth step is to introduce the marginal density h of ϕ and the conditional density j of β given ϕ , namely

$$h(\phi) = \int \beta f(\beta \phi) g(\beta) d\beta \quad (20)$$

and

$$j(\beta | \phi) = \frac{\beta f(\beta \phi) g(\beta)}{h(\phi)}. \quad (21)$$

We can also introduce the conditional expectation of β , namely

$$b(\phi) = \int \beta j(\beta | \phi) d\beta. \quad (22)$$

Remark 18 The limits of integration in the definition of h (namely (20)) are implicit in the definitions of f and g . Since the integrand will only be non-zero if both $f(\beta \phi)$ and $g(\beta)$ are non-zero, these limits are $\max\{\underline{\beta}, \underline{\theta}/\phi\}$ and $\min\{\bar{\beta}, \bar{\theta}/\phi\}$. In particular, the support of the conditional distribution of β varies with ϕ :

1. For $\phi \in [\underline{\phi}, \min\{\underline{\theta}/\underline{\beta}, \bar{\theta}/\bar{\beta}\}]$, the support of β is $[\underline{\theta}/\phi, \bar{\beta}]$. In other words: the range of β types that is consistent with ϕ is increasing in ϕ , and

this range always includes $\bar{\beta}$. By the same token, the range of θ types that is consistent with ϕ is increasing in ϕ , and this range always includes $\underline{\theta}$.

2. For $\phi \in [\max\{\underline{\theta}/\underline{\beta}, \bar{\theta}/\bar{\beta}\}, \bar{\phi}]$, the support of β is $[\underline{\beta}, \bar{\theta}/\phi]$. In other words: the range of β types that is consistent with ϕ is decreasing in ϕ , and this range always includes $\underline{\beta}$.
3. If $\underline{\theta}/\underline{\beta} < \bar{\theta}/\bar{\beta}$ then, for $\phi \in [\min\{\underline{\theta}/\underline{\beta}, \bar{\theta}/\bar{\beta}\}, \max\{\underline{\theta}/\underline{\beta}, \bar{\theta}/\bar{\beta}\}]$, the support of β is $[\underline{\beta}, \bar{\beta}]$. In other words, if the range of θ types is large relative to the range of β types, then all β types are consistent with intermediate values of ϕ .
4. If $\underline{\theta}/\underline{\beta} > \bar{\theta}/\bar{\beta}$ then, for $\phi \in [\min\{\underline{\theta}/\underline{\beta}, \bar{\theta}/\bar{\beta}\}, \max\{\underline{\theta}/\underline{\beta}, \bar{\theta}/\bar{\beta}\}]$, the support of β is $[\underline{\theta}/\phi, \bar{\theta}/\phi]$. In other words, if the range of θ types is small relative to the range of β types, then there is no value of ϕ for which all β types are consistent with that value.

Armed with b and h , the integral defining welfare and the integral giving the resource constraint can be expressed

$$\int \left(b(\phi) \phi v_1(\phi) + v_2(\phi) \right) h(\phi) d\phi \quad (\text{W})$$

and

$$\int \left(Y - C_1(v_1(\phi)) - \frac{1}{R} C_2(v_2(\phi)) \right) h(\phi) d\phi \geq 0. \quad (\text{R})$$

We have therefore completed the transformation of our initial two-dimensional problem into a purely one-dimensional problem.

The Lagrangean for the one-dimensional problem can be written

$$\begin{aligned} & \int \left(b(\phi) \phi v_1(\phi) + v_2(\phi) \right) h(\phi) d\phi \\ & + \lambda \int \left(Y - C_1(v_1(\phi)) - \frac{1}{R} C_2(v_2(\phi)) \right) h(\phi) d\phi \\ & - \int \left(\phi v_1'(\phi) + v_2'(\phi) \right) \mu(\phi) h(\phi) d\phi \\ & - \int v_2'(\phi) \nu(\phi) h(\phi) d\phi, \end{aligned}$$

where the Lagrange multipliers on the resource constraint, the incentive-compatibility

constraint (ICL) and the incentive-compatibility constraint (ICM) take the form $\lambda \in \mathbb{R}$, $\mu : \Phi \rightarrow \mathbb{R}$ and $\nu : \Phi \rightarrow \mathbb{R}$.

B.3 The First-Order Conditions

In order to derive first-order conditions from this Lagrangean, we must first eliminate v'_1 and v'_2 . We can do this by integrating by parts. Taking the third term of the Lagrangean, we obtain

$$\begin{aligned} - \int (\phi v'_1 + v'_2) \mu h d\phi &= - \int ((\phi v_1)' - v_1 + v'_2) \mu h d\phi \\ &= \int v_1 \mu h d\phi - \int ((\phi v_1)' + v'_2) \mu h d\phi, \end{aligned}$$

where we have dropped the dependence of v_1 , v_2 , μ and h on ϕ . Moreover

$$\begin{aligned} - \int ((\phi v_1)' + v'_2) \mu h d\phi &= - [((\phi v_1) + v_2) \mu h]_{\underline{\phi}}^{\bar{\phi}} + \int ((\phi v_1) + v_2) (\mu h)' d\phi \\ &= \int ((\phi v_1) + v_2) (\mu h)' d\phi \end{aligned}$$

(since $h(\underline{\phi}) = h(\bar{\phi}) = 0$). Similarly, taking the fourth term,

$$\begin{aligned} - \int v'_2 \nu h d\phi &= - [v_2 \nu h]_{\underline{\phi}}^{\bar{\phi}} + \int v_2 (\nu h)' d\phi \\ &= \int v_2 (\nu h)' d\phi. \end{aligned}$$

The Lagrangean can therefore be written

$$\begin{aligned} &\int \left(\left((b\phi + \mu) v_1 + v_2 + \lambda \left(Y - C_1(v_1) - \frac{1}{R} C_2(v_2) \right) \right) \right) h \\ &+ (\phi v_1 + v_2) (\mu h)' + v_2 (\nu h)' d\phi. \end{aligned}$$

Differentiating the latter Lagrangean with respect to v_1 and v_2 , we obtain the first-order conditions

$$0 = \left(b\phi + \mu - \lambda C'_1(v_1) \right) h + \phi (\mu h)'$$

and

$$0 = \left(1 - \lambda \frac{1}{R} C_2'(v_2)\right) h + (\mu h)' + (\nu h)'$$

We also have: (IC1), namely

$$0 = \phi v_1' + v_2';$$

the complementary slackness condition associated with the resource constraint, namely

$$\left. \begin{array}{l} 0 \leq \int \left(Y - C_1(v_1) - \frac{1}{R} C_2(v_2)\right) h d\phi \\ 0 \leq \lambda \end{array} \right\};$$

and the complementary slackness condition associated with (IC2), namely

$$\left. \begin{array}{l} 0 \leq -v_2' \\ 0 \leq \nu \end{array} \right\}.$$

B.4 The Relaxed Problem

We focus on the relaxed version of the problem, in which we do not impose (IC2). Furthermore, we look for a solution of the relaxed problem in which the resource constraint holds as an equality. We therefore drop ν from the equations and tackle the three differential equations

$$0 = \left(b\phi + \mu - \lambda C_1'(v_1)\right) h + \phi (\mu h)', \quad (23)$$

$$0 = \left(1 - \lambda \frac{1}{R} C_2'(v_2)\right) h + (\mu h)', \quad (24)$$

$$0 = \phi v_1' + v_2' \quad (25)$$

and the integral equation

$$0 = \int \left(Y - C_1(v_1) - \frac{1}{R} C_2(v_2)\right) h d\phi. \quad (26)$$

The first step is to make v_1 and v_2 the subjects of equations (23) and (24). Putting $U_1 = (C_1')^{-1}$ and $U_2 = (C_2')^{-1}$, we obtain

$$v_1 = U_1\left(\frac{a_1}{\lambda}\right), \quad (27)$$

$$v_2 = U_2\left(\frac{a_2}{\lambda}\right), \quad (28)$$

where

$$a_1 = b\phi + \mu + \frac{\phi(\mu h)'}{h}, \quad (29)$$

$$a_2 = R \left(1 + \frac{(\mu h)'}{h} \right). \quad (30)$$

B.5 Solving (23-25) where b and h are Smooth

Consider the equations (23-25) in the open region $\mathring{\Phi} = \Phi \setminus \{\underline{\phi}, \underline{\theta}/\underline{\beta}, \bar{\theta}/\bar{\beta}, \bar{\phi}\}$. In this region, both b and h are smooth. Hence we may differentiate (27,28) to obtain

$$v'_1 = U'_1 \left(\frac{a_1}{\lambda} \right) \frac{a'_1}{\lambda}, \quad (31)$$

$$v'_2 = U'_2 \left(\frac{a_2}{\lambda} \right) \frac{a'_2}{\lambda} \quad (32)$$

and, substituting (31,32) in (25),

$$0 = \phi U'_1 \left(\frac{a_1}{\lambda} \right) \frac{a'_1}{\lambda} + U'_2 \left(\frac{a_2}{\lambda} \right) \frac{a'_2}{\lambda}.$$

Next, provided that u_1 and u_2 have the same coefficient of relative risk aversion γ , the latter equation is homogeneous in λ . It therefore simplifies further to

$$0 = \phi U'_1(a_1) a'_1 + U'_2(a_2) a'_2.$$

(If u_1 and u_2 have coefficient of relative risk aversion γ , then $U'_1(x) = U'_2(x) = \frac{1}{\gamma} x^{\frac{1}{\gamma}-2}$.)

Next, substituting for a'_1 and a'_2 and collecting terms in μ'' , μ' and μ , we obtain

$$\begin{aligned} 0 &= (\phi^2 U'_1(a_1) + R U'_2(a_2)) h^2 \mu'' \\ &+ (\phi(\phi h' + 2h) U'_1(a_1) + h' R U'_2(a_2)) h \mu' \\ &+ (\phi(h(\phi h'' + h') - \phi h'^2) U'_1(a_1) + (h h'' - h'^2) R U'_2(a_2)) \mu \\ &+ \phi(\phi b' + b) U'_1(a_1) h^2. \end{aligned} \quad (33)$$

In other words, in the region $\mathring{\Phi}$, equations (23-25) reduce to a second-order ordinary differential equation for μ .

B.6 Solving (23-25) where b and h have Kinks

Now consider the equations (23-25) at the points $\phi_1 = \underline{\theta}/\underline{\beta}$ and $\phi_2 = \bar{\theta}/\bar{\beta}$, where both b and h have kinks. We cannot differentiate (27,28) at these points. However, we do have

$$\begin{aligned}\Delta v_1(\phi_i) &= U_1\left(\frac{a_1(\phi_i+)}{\lambda}\right) - U_1\left(\frac{a_1(\phi_i-)}{\lambda}\right), \\ \Delta v_2(\phi_i) &= U_2\left(\frac{a_2(\phi_i+)}{\lambda}\right) - U_2\left(\frac{a_2(\phi_i-)}{\lambda}\right)\end{aligned}$$

where

$$\begin{aligned}a_1(\phi_i+) &= b\phi_i + \mu(\phi_i+) + \frac{\phi(\mu h)'(\phi_i+)}{h(\phi_i)}, \\ a_1(\phi_i-) &= b\phi_i + \mu(\phi_i-) + \frac{\phi(\mu h)'(\phi_i-)}{h(\phi_i)}, \\ a_2(\phi_i+) &= R\left(1 + \frac{(\mu h)'(\phi_i+)}{h(\phi_i)}\right), \\ a_2(\phi_i-) &= R\left(1 + \frac{(\mu h)'(\phi_i-)}{h(\phi_i)}\right).\end{aligned}$$

Hence, at ϕ_i , we can impose the value-matching condition

$$0 = \Delta\mu(\phi_i) = \mu(\phi_i+) - \mu(\phi_i-) \quad (34)$$

and the incentive condition

$$0 = \phi_i(U_1(a_1(\phi_i+)) - U_1(a_1(\phi_i-))) + (U_2(a_2(\phi_i+)) - U_2(a_2(\phi_i-))). \quad (35)$$

B.7 Solving (23-25) at the Endpoints

Assuming for concreteness that $\phi_1 < \phi_2$, we now have the second-order ordinary differential equation (33) in the three open intervals $(\underline{\phi}, \phi_1)$, (ϕ_1, ϕ_2) and $(\phi_2, \bar{\phi})$. Moreover, we have two boundary conditions at each of ϕ_1 and ϕ_2 . (Cf. (34) and (35).) The obvious way of completing the equation would therefore be to require that μ take on appropriate values at the boundaries $\underline{\phi}$ and $\bar{\phi}$. However, h decays linearly to 0 at both $\underline{\phi}$ and $\bar{\phi}$. Moreover, inspection of (33) shows that:

1. the coefficient of μ'' is positive and of order h^2 near $\underline{\phi}$ and $\bar{\phi}$;

2. the coefficient of μ' is positive and of order h near $\underline{\phi}$, and negative and of order h near $\bar{\phi}$;
3. the coefficient of μ is negative and of order 1 near $\underline{\phi}$ and $\bar{\phi}$.

Hence μ will not take on boundary values at $\underline{\phi}$ and $\bar{\phi}$ in the usual way.²⁰ On the other hand, the inhomogeneous term, namely

$$\phi(\phi b' + b)U_1'(a_1)h^2,$$

is of order h^2 near $\underline{\phi}$ and $\bar{\phi}$. In particular, it is bounded. Hence the relevant solution of the equation is the one that is bounded near $\underline{\phi}$ and $\bar{\phi}$.²¹

B.8 Solving for λ

As we have seen, we can find μ by solving the second-order o.d.e. (33) with the required boundary conditions at the internal boundaries ϕ_1 and ϕ_2 and the required boundedness properties at the endpoints $\underline{\phi}$ and $\bar{\phi}$. Like b and h , μ can be expected to have kinks at ϕ_1 and ϕ_2 . The next step is to solve for λ . This can be done using the resource equation (26).

Indeed, if u_1 and u_2 have the same coefficient of relative risk aversion γ , then we have

$$C_i(v_i) = C_i\left(U_i\left(\frac{a_i}{\lambda}\right)\right) = \left(\frac{a_i}{\lambda}\right)^{\frac{1}{\gamma}}.$$

Hence, substituting in (26),

$$0 = \int \left(Y - \left(\frac{a_1}{\lambda}\right)^{\frac{1}{\gamma}} - \frac{1}{R} \left(\frac{a_2}{\lambda}\right)^{\frac{1}{\gamma}} \right) h d\phi = \lambda^{-\frac{1}{\gamma}} \int \left(\lambda^{\frac{1}{\gamma}} Y - a_1^{\frac{1}{\gamma}} - \frac{1}{R} a_2^{\frac{1}{\gamma}} \right) h d\phi,$$

or

$$\lambda^{\frac{1}{\gamma}} = \frac{\int \left(a_1^{\frac{1}{\gamma}} + \frac{1}{R} a_2^{\frac{1}{\gamma}} \right) h d\phi}{\int Y h d\phi}.$$

²⁰Intuitively speaking, the dynamics of ϕ move away from the endpoints $\underline{\phi}$ and $\bar{\phi}$.

²¹Since the inhomogeneous term is of order h^2 near $\underline{\phi}$ and $\bar{\phi}$, the solution can in fact be expected to decay quadratically to 0 at both $\underline{\phi}$ and $\bar{\phi}$. In particular, we would expect that it would satisfy $\mu(\underline{\phi}) = \mu'(\underline{\phi}) = 0$ and $\mu(\bar{\phi}) = \mu'(\bar{\phi}) = 0$. These equations cannot, however, be used as boundary conditions. For one thing, there are too many of them! (There are 4 instead of 2.) They are simply additional properties that we would expect the unique bounded solution to possess.

Bearing in mind that a_1 and a_2 are given in terms of μ by equations (29) and (30), this gives us a formula for λ in terms of μ .

B.9 Completing the Solution

It is then a straightforward matter to find the remaining unknowns in the model: v_1 and v_2 are given in terms of λ and μ by (27) and (28); and c_1 and c_2 are given in terms of v_1 and v_2 by the formulae $c_1 = C_1(v_1)$ and $c_2 = C_2(v_2)$.

C Analysis of the Quasi-Linear Limit Case for a Population of Agents with Heterogeneous β

C.1 Introduction

In Subsections 4.1 and 5.1, we discuss the quasi-linear limit case of our model: i.e., the case in which the utility function in the second period is linear (i.e., $u_2(c_2) = c_2$). In this case, the planner's problem can be written

$$\max \int (\theta u_1(c_1) + u_2(c_2)) dF(\theta) dG(\beta) = \max \int (\theta u_1(c_1) + c_2) dF(\theta) dG(\beta),$$

subject to

$$\int (c_1 + c_2) dF(\theta) dG(\beta) = Y,$$

$$\phi \in \arg \max_{\phi' \in \Phi} \{\phi u_1(c_1(\phi')) + u_2(c_2(\phi'))\} \quad (\text{IC})$$

for $\phi \equiv \theta/\beta$.

We study equilibria that satisfy the revelation principle, and, following the literature, refer to these as direct mechanisms. When we talk about ϕ , we refer to the true value of ϕ elicited from each agent in an equilibrium that satisfies the revelation principle.

Proposition 19 *Suppose that $E[\theta | \phi]$ is non-decreasing in ϕ . Then, the allocation rule under the optimal mechanism is*

$$E[\theta | \phi] u_1'(c_1(\phi)) = 1.$$

Accordingly the implied marginal penalty is:

$$\pi(\phi) = 1 - \mathbf{E}[\beta | \phi].$$

Accordingly, in the case of homogeneous β , we get a simple Pigouvian tax, $\pi = 1 - \beta$, because $\mathbf{E}[\beta | \phi]$ is equal to the homogeneous value of β in the population. In the case of heterogeneous β , we get a ϕ -conditional Pigouvian tax: $\pi(\phi) = 1 - \mathbf{E}[\beta | \phi]$.

C.2 Derivations

C.2.1 Implementability

Given the representation of the problem in the space of ϕ , we now effectively have a single-type mechanism-design problem. We begin by transforming the problem into the promised utility space, $v_1(\phi) = u_1(c_1(\phi))$ and $v_2(\phi) = u_2(c_2(\phi)) = c_2(\phi)$. We invoke the standard equivalence between global incentive compatibility and the combination of integral incentive compatibility and monotonicity. Monotonicity implies $v_1'(\phi) \geq 0$, and in the standard way we solve the relaxed mechanism (not subject to monotonicity) and verify that the solution satisfies monotonicity.

Integral incentive compatibility is the standard condition, derived from the Envelope Theorem. In particular, the Envelope Theorem implies $\frac{d}{d\phi} (\phi v_1(\phi) + v_2(\phi)) = v_1(\phi)$, and we obtain integral incentive compatibility by integrating:

$$\phi v_1(\phi) + v_2(\phi) = \bar{\phi} v_1(\bar{\phi}) + v_2(\bar{\phi}) + \int_{\bar{\phi}}^{\phi} v_1(\zeta) d\zeta.$$

We then use integral incentive compatibility to define the function v_2 in terms of the function v_1 and the constant $v_2(\bar{\phi})$, which gives us the implementing function v_2 that guarantees integral incentive compatibility:

$$v_2(\phi) = \bar{\phi} v_1(\bar{\phi}) + v_2(\bar{\phi}) + \int_{\bar{\phi}}^{\phi} v_1(\zeta) d\zeta - \phi v_1(\phi).$$

We then characterize $v_2(\bar{\phi})$ from v_1 using the resource constraint. Rewriting the resource constraint over promised utility in the ϕ space:

$$\int (u_1^{-1}(v_1(\phi)) + v_2(\phi)) dH(\phi) = Y.$$

Rearranging:

$$\int v_2(\phi) dH(\phi) = Y - \int u_1^{-1}(v_1(\phi)) dH(\phi).$$

Or, in other words, given a specification of a function v_1 , we can use this condition plus the implementability condition to pin down v_2 . In other words, if we substitute in the implementability condition for v_2 , we get an equation for $v_2(\bar{\phi})$ in terms of v_1 :

$$v_2(\bar{\phi}) = Y - \int u_1^{-1}(v_1(\phi)) dH(\phi) - \bar{\phi} v_1(\bar{\phi}) - \int \left(\int_{\bar{\phi}}^{\phi} v_1(\zeta) d\zeta - \phi v_1(\phi) \right) dH(\phi).$$

C.2.2 Completing the Model

Lastly, let us rewrite the objective function in terms of ϕ and v_1 . The contribution of type- ϕ agents to social welfare is $E[\theta | \phi] v_1(\phi) + v_2(\phi)$. Therefore, the planner objective function is:

$$\int \left(E[\theta | \phi] v_1(\phi) + v_2(\phi) \right) dH(\phi).$$

Substituting in the characterization of v_2 above, we get:

$$\max_{v_1} \left\{ \int \left(E[\theta | \phi] v_1(\phi) - u_1^{-1}(v_1(\phi)) \right) dH(\phi) + Y \right\} \quad \text{s.t. (Monotonicity).}$$

That is, the planner chooses a non-decreasing function v_1 , with the implementability conditions above defining the function v_2 that implements this outcome.

Proposition 20 *Suppose that $E[\theta | \phi]$ is non-decreasing in ϕ . Then, the optimal allocation rule is:*

$$E[\theta | \phi] u_1'(c_1(\phi)) = 1.$$

Proof. This condition is simply the first order condition from choosing the control v_1 , and substituting back in $v_1(\phi) = u_1(c_1(\phi))$. Provided that $E[\theta | \phi]$ is non-decreasing, the above first order condition yields a monotone function c_1 (and hence monotone v_1), and so is implementable. ■

C.3 The Optimal Penalty

Consider the implied marginal penalty $\pi(\phi)$ that implements the above allocation rule. The marginal trade-off of a private agent is then:

$$(1 - \pi(\phi)) \phi u'_1(c_1(\phi)) = 1.$$

Therefore, the marginal penalty is:

$$1 - \pi(\phi) = \frac{\mathbb{E}[\theta | \phi]}{\phi} = \mathbb{E}\left[\frac{\theta}{\phi} | \phi\right] = \mathbb{E}[\beta | \phi].$$

C.4 Homogeneous β

If β is homogeneous, then $\mathbb{E}[\beta | \phi] = \beta$, and we have:

$$\pi(\phi) = 1 - \beta.$$

That is, we simply have a Pigouvian tax. Note that homogeneous β always satisfies the “regularity condition” of Proposition 1 needed to guarantee monotonicity. In fact, this result is first-best.

Proposition 21 *With homogeneous β , the Pigouvian tax yields a first-best allocation.*

Proof. With $\pi(\phi) = 1 - \beta$, every agent sets $\theta u'_1(c_1) = 1$, so the allocation rule is first-best. The resource distribution of c_2 is irrelevant to social welfare, and therefore we have the first-best social welfare. ■

It is simply a social first-best allocation, since as emphasized above, any zero-sum perturbation Δ of c_2 is also a first-best allocation rule. However, the allocation rule above is the *unique first-best implementable allocation rule*. To understand why the allocation rule is unique, note that the optimal rule v_1 is unique (the solution to the simple first order condition). The resource constraint then uniquely pins down $v_2(\bar{\phi})$, given v_1 . Finally, integral incentive compatibility uniquely pins down the full function v_2 , given v_1 and $v_2(\bar{\phi})$.

C.5 Heterogeneous β

If β is heterogeneous *and* the regularity condition of Proposition 1 is satisfied, then as mentioned before we have:

$$\pi(\phi) = 1 - \mathbf{E}[\beta \mid \phi].$$

That is, we have an “average Pigouvian tax”: the optimal tax rate on the margin for a type- ϕ agent is the average tax rate in that population.

We know that $\pi(\phi)$ must be close to $1 - \bar{\beta}$ near $\underline{\phi}$, where the highest β types are the only ones with that ϕ type. Similarly, we know that $\pi(\phi) \simeq 1 - \underline{\beta}$ near $\bar{\phi}$. This suggests a large degree of flexibility over initial withdrawals, and much tighter restrictions on flexibility for households withdrawing a lot.