Optimal Illiquidity*

John Beshears † James J. Choi‡ Christopher Clayton§

Christopher Harris¶ David Laibson‖ Brigitte C. Madrian**

June 14, 2023

Abstract

We study the socially optimal level of illiquidity in an economy populated by households with taste shocks and present bias with naive beliefs. The government chooses mandatory contributions to accounts, each with a different pre-retirement withdrawal penalty. Collected penalties are rebated lump sum. When households have homogeneous present bias, β, the social optimum is well approximated by a single account with an early-withdrawal penalty of $1 - \beta$. When households have heterogeneous present bias, the social optimum is well approximated by a two-account system: (i) an account that is completely liquid and (ii) an account that is completely illiquid until retirement.

*The authors gratefully acknowledge financial support from from the TIAA Institute, the Pension Research Council, the Eric M. Mindich Fund for Research on the Foundations of Human Behavior, the Pershing Square Fund for Research on the Foundations of Human Behavior, and the Social Security Administration. The views expressed in this paper are not endorsed by the Social Security Administration or other funders. The authors received insightful comments from Manuel Amador, Marios Angeletos, Emmanuel Farhi, Mike Golosov, Francesco Lippi, Ben Lockwood, Greg Mankiw, Christian Moser, Matthew Rabin, David Richardson, John Sabelhaus, Dmitry Taubinsky, Iván Werning, 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 Santiago Medina Pizarro and Kartik Vira for outstanding research assistance.

†Beshears: Harvard. Email: jbeshears@hbs.edu

‡Choi: Yale. Email: james.choi@yale.edu

§Clayton: Yale. Email: christopher.clayton@yale.edu

¶Harris: Cambridge. Email: cjharris@econ.cam.ac.uk

‖Laibson (Corresponding Author): Harvard. Email: dlaibson@harvard.edu

**Madrian: BYU. Email: brigitte_madrian@byu.edu

1 Introduction

What are the liquidity characteristics of a socially optimal savings system? Almost all developed economies have some form of compulsory savings that is *completely* illiquid (e.g., U.S. Social Security). In many countries, defined-contribution (DC) savings accounts have mandatory contributions and balances that are completely illiquid during most of working life (Beshears et al., 2015). In the United States, by contrast, DC contributions are almost always voluntary (in IRA's and 401(k)'s), 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, $0.40 simultaneously flows out of the DC system for the same age group, not counting rollovers or loans (Argento, Bryant, and Sabelhaus, 2015).[1] It is not clear whether allowing such leakage is consistent with overall social welfare maximization, although most media coverage bemoans leakage.[2]

Our paper evaluates the optimality of an N-account system comprised of liquid, partially illiquid, and completely illiquid accounts. The illiquidity is obtained with compulsory deposits and linear penalties for pre-retirement withdrawals. We show that simple two- or three-account systems come extremely close to delivering the welfare obtainable from a fully general (non-linear) mechanism. We find an upper bound for social welfare and show that two- and three-account systems nearly attain this bound.

We study preferences that include both normative taste shocks and non-normative self-control problems due to present bias: i.e., the discount function $\{1, \beta\,\delta, \beta\,\delta^2, \ldots, \beta\,\delta^t\}$, where the degree of present bias is $1-\beta$ (Phelps and Pollak, 1968; Laibson, 1997). Our model builds on the commitment vs. flexibility framework of Amador, Werning, and Angeletos (2006), hereafter AWA. AWA features households with homogeneous present bias $1-\beta$ and heterogeneous taste shocks θ, with each household's θ being private information. AWA does

[1] About half of these withdrawals (dollar-weighted) are made in a category that avoids the 10% penalty.

[2] 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).

not allow for policies that admit inter-household transfers, and finds that minimum savings is optimal. In other words, the optimal system consists of two accounts, where one account is fully liquid and the other is completely illiquid until period 2.[3]

In the current paper, we revisit AWA's setting and make two changes that reflect considerations faced by a policymaker. First, we allow the government to make inter-household transfers. Second, we allow for heterogeneous β, with each household's β being private information. We show that, absent heterogeneity of β, a simple Pigouvian tax is approximately optimal (i.e., a one-account system is approximately optimal). However, with sufficient heterogeneity of β, the optimal account mechanism is well-approximated by a two-account system, with one completely liquid account and one completely illiquid account (like AWA, but arising for different reasons).

The central contribution of the current paper is therefore to show that a simple (e.g., 1- or 2-account) mechanism that mirrors actual institutions achieves social welfare that is very close to the welfare that would be achieved by a fully general (non-linear) mechanism. We now provide a map/summary of the argument that runs through the rest of this paper. Our analysis is based on three nested classes of mechanisms.

The first class is the class of mechanisms with N accounts in which households are free to draw down the accounts in whatever order they prefer. The class of N-Account Mechanisms is restrictive (from the point of view of the planner) in two ways. First, an N-Account Mechanism has a finite number of accounts, each of which has a linear penalty for early (i.e., period 1) withdrawals. Second, an N-Account Mechanism restricts the planner by allowing households to draw down the accounts in the order that the households prefer. Because this system is closest to the actual set of institutions/accounts that exist in almost all countries, this first class of mechanisms is the focus of our numerical analysis. We denote the planner welfare achieved under an N-Account Mechanism as W^N.

[3] Halac and Yared (2014) study the commitment vs. flexibility tradeoff with persistent shocks and show that the second-best optimal mechanism features history dependence. Bond and Sigurdsson (2018) study the commitment vs. flexibility trade-off in three periods, identifying conditions that produce a first-best allocation. See also Beshears et al. (2020).

The second class of consists of general non-linear mechanisms satisfying the standard Global Incentive Compatibility constraint, re-expressed as usual as a Local Incentive Compatibility constraint and a Monotonicity constraint. We denote the planner welfare achieved under this class of General Mechanisms as W^G. This is the theoretical (feasible) social optimum. This second class of mechanisms allows for arbitrary non-linear budget sets rather than requiring piecewise linear, convex budget sets (which were assumed in the first class of mechanisms).[4]

The third class of mechanisms consists of general non-linear mechanisms satisfying only the Local Incentive Compatibility Constraint (i.e., omitting the Monotonicity constraint from the second class). As is standard, we refer to this as the Relaxed Problem. Because Monotonicity is required for Global Incentive Compatibility, the optimum in this third class of mechanisms is not necessarily feasible. The third class is used in our paper as the source for an upper bound on welfare. We denote the planner welfare achieved under this Relaxed Problem as W^R.

Because of the progressive relaxation of constraints, the three classes of mechanism generate ranked levels of welfare (where welfare is evaluated from the perspective of the planner). To provide a welfare benchmark, we compare all three mechanisms to an autarkic economy (with a household storage technology) in which all households are given a single liquid account and inter-household transfers are not possible. We denote planner welfare under this autarkic benchmark by W^A.[5] We then have

$$W^A \leq W^N \leq W^G \leq W^R.$$

So far we have summarized the autarkic benchmark and three interventionist policy regimes: N-Account Mechanism, General Mechanism, and Relaxed Problem. Our analysis

[4]Our first class of mechanisms – N accounts with a flexible draw-down rule – is mathematically equivalent to assuming a piecewise linear, convex budget set. Piecewise linearity follows from the assumption that each account has a specific (constant) early withdrawal penalty. Convexity follows from the flexible draw-down rule, which implies that households will draw down the accounts in order of increasing penalty.

[5]Because we assume naivete, the autarky case does not generate self-imposed illiquidity.

finds the *minimum* number of accounts under N Accounts such that the welfare improvement from N Accounts to the Relaxed Problem is *de minimis*. Under this minimum number of accounts, the welfare improvement from the N-Account Mechanism to the General Mechanism must also be *de minimis* (because $W^G - W^N$ is necessarily weakly smaller than $W^R - W^N$). In other words, the number of accounts that makes welfare under an N-Account Mechanism close to welfare under the Relaxed Problem, must also make welfare under that N-Account Mechanism close to welfare under the General Mechanism.

This analytic approach is divided into two fundamental cases: homogeneous β and heterogeneous β. When we assume that β is heterogeneous across households, we assume that the social planner knows the population distribution of β, but does not know the value of β for each household. The interventionist mechanisms that we study and the bounding argument that we use, are summarized in the following 3×2 matrix:

	Homogeneous β (Section 3)	Heterogeneous β (Section 4)
N-account Mechanisms (W^N)	$N = 1$	$N \in \{2, 3\}$
General Mechanism (W^G)	$W^G - W^N \leq W^R - W^N$	$W^G - W^N \leq W^R - W^N$
Relaxed Problem (W^R)	Numerical ODE	Numerical ODE

The N-account Mechanism (first row) is approximately optimal in the homogeneous-β case (first column; $N = 1$) and the heterogeneous-β case (second column; $N \in \{2, 3\}$). In the homogeneous-β case, a single account suffices to achieve a close welfare approximation to the General Mechanism. In the heterogeneous-β case, two accounts suffice to achieve a close welfare approximation to the General Mechanism, although some further improvement – as well as greater realism – can be obtained by adding a third account. These results derive from the simple bound stated in the second row of the table, namely $W^G - W^N \leq W^R - W^N$. In the homogeneous-β case, the welfare gap $W^R - W^N$ between the Relaxed Problem and an N-account Mechanism is *de minimis* when $N = 1$; and, in the heterogeneous-β case, the welfare gap $W^R - W^N$ is *de minimis* when $N = 2$ and becomes even smaller when $N = 3$. Hence, the welfare gap $W^G - W^N$ is likewise *de minimis* in these cases. The final row of the table refers to our solution method for the Relaxed Problem: we solve the ordinary differential equation (ODE) that characterizes the planner's relaxed optimization problem numerically (see Appendix D).

In the homogeneous-β case, the 1-account system that approximates the theoretically optimal welfare (achievable using a general mechanism) consists of partially illiquid account with (Pigouvian) early-withdrawal penalty $\pi \simeq 1 - \beta$.

In the heterogeneous-β case, we find that completely illiquid accounts play an important role in improving welfare. Specifically, the theoretically optimal welfare is well-approximated by a 3-account system with: (1) a perfectly liquid savings account; (2) a partially illiquid savings account (with an early-withdrawal penalty of approximately 13%); and (3) a completely illiquid savings account. More strikingly, the social optimum is also well-approximated by an even simpler 2-account system with a completely liquid savings account and a completely illiquid savings account. In both the 2- and 3-account systems the completely illiquid 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.[6] The completely illiquid savings account caters to the households with relatively 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 being forced to shift some of their wealth from completely liquid accounts to completely illiquid accounts).

To the extent that there is a role for low-balance 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%. Hence, the partially illiq-

[6]In our model the planner collects all of the resources from households and then deposits funds into accounts. This is isomorphic to a model in which households are themselves required to deposit funds into accounts (e.g., an illiquid account functions like a DB pension, Social Security, or a typical (non-US) DC account that does not allow pre-retirement withdrawals).

uid accounts look like a 401(k) account in the U.S. Moreover, these partially illiquid accounts display a high level of leakage in equilibrium: pre-retirement withdrawals are commonplace. This leakage results in part from normative 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 benefit from a fiscal externality) are nearly offsetting.

There is a growing literature that studies how present bias effects retirement savings and how governments should optimally respond.[7] Our model is related to the independent and contemporaneous work of Moser and Olea de Souza e Silva (2019), who study an environment with unobservable earnings ability, unobservable β, and inter-household transfers. Moser and Olea de Souza e Silva (2019) find that optimal savings institutions include some forced savings, a result that also emerges in AWA and in our own paper. Like Moser and Olea de Souza e Silva (2019), we find that optimal savings mechanisms are characterized by more mandatory savings than currently exists in the U.S. system. Most importantly, our paper is the first to show 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 with transfers.[8] We contribute to the literature that identifies settings in which very simple mechanisms provide good welfare approximations to arbitrarily complex, optimal mechanisms.[9]

Finally, a large literature studies how firms attempt to exploit agents with present bias.[10]

[7]For example, Laibson, Repetto, and Tobacman (1998, 2003) study the design of U.S. 401(k)'s, Galperti (2015) studies optimal screening among agents with different levels of present bias, Paluszynski and Yu (2019) study the effects of preference heterogeneity across educational groups, Yu (2021) studies screening between sophisticates and naives, Pavoni and Yazici (2016) study optimal lifecycle taxation, Maxted (2022) identifies isomorphisms between optimal policies with time consistent and present-biased agents (in economies in which agents are always in the interior of their action space). See also O'Donoghue and Rabin (1999b).

[8]There is a 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), Allcott, Lockwood, and Taubinsky (2019), Lockwood (2020), Farhi and Gabaix (2020). See Bernheim and Taubinsky (2018) for a review of behavioral public economics.

[9]For example, see Reichelstein (1992), Bower (1993), Sappington and Weisman (1996), Gasmi, Laffont, and Sharkey (1999), McAfee (2002), Rogerson (2003), and Chu and Sappington (2007).

[10]For example, see DellaVigna and Malmendier (2004, 2006), Heidhues and Kőszegi (2010), Sulka (2022), and several literature reviews: Heidhues and Kőszegi (2018), Ericson and Laibson (2019), and Cohen et al.

By contrast, our paper studies how a benevolent planner would set up a simple socially optimal pension scheme.

Our paper proceeds as follows. 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 households. Section 2 also analyzes the case of homogeneous present bias without inter-household transfers (AWA): i.e., resources collected by the government must be destroyed rather than redistributed.

Sections 3, 4, and 5 all incorporate inter-household transfers, which is a generalization from AWA. Sections 3 and 4 respectively analyze the economy with homogeneous and heterogeneous present bias. Section 5 presents robustness analysis. Section 6 highlights the many strong assumptions that we make and raises questions of generalizability. Five online appendices contain proofs, including a method for calculating welfare for the Relaxed Problem (Appendix D).

2 Model

We study a two-period model of consumption for a continuum of households, with idiosyncratic taste shock θ and idiosyncratic 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; 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). In equilibrium, households choose to withdraw from the low-penalty accounts first. An N-account Mechanism is equivalent to a budget set that is piecewise linear and convex, whereas the General Non-linear Mechanism imposes neither of these restrictions. To preview the results to come, we show that the welfare that arises from the N-account

(2020).

Mechanism with $N \leq 2$ is very close to the welfare for the General Non-linear Mechanism. We focus most of this paper on N-account Mechanisms because of their similarity to the actual retirement savings systems that are currently in use globally.

2.1 Preferences of households

Preferences in period 1 are given by

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

where θ is a stochastic taste shifter,[11] $u_t : (0, \infty) \to \mathbb{R}$ is the period-t utility function, c_t is period-t consumption, β is the present-bias discount factor, and δ is the standard discount factor.[12] Preferences in period 2 are given by $u_2(c_2)$.

2.2 Information structure

We assume households are naive: they do not anticipate present bias (see Strotz 1955; O'Donoghue and Rabin 1999a). The assumption of naivite is broadly supported by the empirical literature (see reviews in Ericson and Laibson 2019; Cohen et al. 2020), although there are a range of results (e.g., see Allcott et al. 2022). The assumption of naivite eliminates the opportunity for screening in a hypothetical 'pre-period'.[13]

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* distribution of (θ, β) across households. We denote the distribution function of θ by $F(\cdot)$ and of β by $G(\cdot)$. In

[11] See Atkeson and Lucas Jr. (1992) for use of such taste shifters. There are also other ways of modeling taste shifters. For example, consider $u(c - \theta)$, where θ is a taste shifter. This case is beyond the scope of the current paper.

[12] This framework can be generalized WLOG by including a second independent stochastic taste shifter (with mean 1, which is realized in period 2) that multiplies period 2's utility function.

[13] Galperti (2015) studies screening in a contracting setting where agents are sophisticated, have private information about their degree of present bias, and contract with a firm. See also Moser and Olea de Souza e Silva (2019) and Yu (2021).

our analysis, we assume that θ and β are independent, but we could generalize our analysis to allow for a joint distribution.

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 normatively endorse present bias, implying that the planner's objective for a household is

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

The assumption that the social planner maximizes an objective without present bias, is a common assumption in the literature (AWA). The social planner chooses policies that maximize the utilitarian[14] social objective:

$$\int \int \Big(\theta\, u_1(c_1(\theta, \beta)) + \delta\, u_2(c_2(\theta, \beta))\Big)\, dF(\theta)\, dG(\beta). \tag{1}$$

The social planner takes account of the (endogenous) equilibrium policy functions of the households, c_1 and c_2. The social planner creates incentives that influence these policy functions, but can't control them directly because the planner doesn't directly observe θ and β for each household. The social planner's mechanism uses total resources bounded by the aggregate endowment Y.

Equation (1) implies that the planner has two motives in changing the allocations that emerge in an autarkic system. First, the planner would like to generate more savings, because only households, and not the planner, have present bias. Second, the planner would like to generate inter-personal reallocations from agents with low θ values to agents with high θ values. The first motive is an inter-temporal reallocation (within a household) and the second motive is an inter-personal redistribution.

[14]We can generalize this framework to incorporate Pareto weights, but omit this step to simplify exposition.

2.4 Timing

Time 0: The planner sets up N accounts, each with gross rate of return R, where N is a constraint discussed 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 (i.e., an early-withdrawal penalty).[15] If a consumer withdraws ω dollars from account n in period 1, the consumer actually receives $(1-\pi_n)\,\omega$ dollars.[16] A completely liquid account has $\pi_n = 0$, a partially liquid account has $0 < \pi_n < 1$, and a completely illiquid account has $\pi_n = 1$. For the planner, the choice variables are the allocations to the N accounts, $(x_n)_{n=1}^N$, and the respective early withdrawal penalties, $(\pi_n)_{n=1}^N$. The planner chooses the account allocations in a way that respects the economy's overall budget balance: $\sum_{n=1}^N x_n$ will equal Y plus the aggregate value of the early withdrawal penalties collected in equilibrium.

Time 1: Self 1 maximizes welfare from the perspective of time 1 (including present bias). This generates withdrawals from the accounts established at time 0. Consumption is $c_1(\theta, \beta)$.

Time 2: Self 2 spends any remaining funds in their accounts. Consumption is $c_2(\theta, \beta)$.

2.5 Summary of the N-account Mechanism

We begin with the consumer's problem, since consumer behavior is an input to the planner's problem. In period 1, the consumer with parameters θ and β maximizes

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

[15] WLOG, there are no withdrawal penalties in period 2.

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

where consumption is given by

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

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

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 $(\pi_n)_{n=1}^N$.

We assume a continuum of consumers (with measure one), so integrating over taste-parameters, θ and β, is the same as integrating over consumers. In period 0, the planner faces the problem

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

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:

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

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). Although we assume the planner implements the N-account allocation through *involuntary* contributions, the planner could implement the same allocation under *voluntary* contributions through appropriate use of contribution subsidies (e.g., matching contributions).[17] We choose to use an *involuntary*

[17] For example, if the planner sets an account-specific match threshold of z (i.e., the maximum voluntary contribution that can be matched) and an account-specific match rate of m (i.e., the match per dollar of voluntary contributions), then for all m greater than some match rate m^*, the equilibrium account contribution will produce a total account balance of $x = (1 + m)\, z$.

framing in our model presentation because it is without loss of generality and notationally simpler (avoiding matching notation) *and* almost all developed countries have some involuntary retirement savings (e.g., Social Security in the United States, superannuation in Australia, the Central Provident Fund in Singapore, and the public pension system in Sweden, to pick a few examples).[18]

The N-account Mechanism summarized here is a restricted version of the General Non-linear Mechanism. We compare our welfare results to bounds on the General Non-linear Mechanism below.

2.6 Autarky reference case: $\pi = 0$

In the analysis that follows, we 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 an autarkic system with a household storage technology, in which the government does nothing to distort the decisions of each household (implicitly ruling out redistribution).

2.7 Special case of no transfers: AWA (2006)

We consider a first deviation from the autarkic reference case. We allow the government to intervene by offering households a nonlinear budget set. As in autarky, we continue to assume that each budget constraint holds at the household level (instead of economy-wide),

$$c_1 + \frac{c_2}{R} \leq Y \text{ for each household.} \tag{7}$$

ruling out inter-household transfers. As in AWA, households have homogeneous β.

In Appendix A, we prove a version of a proposition by AWA (2006). In particular, we show that under a set of assumptions about u_1, u_2, and F, an optimal mechanism is a two-

[18]Some of these systems are funded, some of are unfunded, and some are hybrid. The key unifying feature (for the purposes of our model) is that they are involuntary.

account system consisting of a completely liquid account (that can be used in both period 1 and period 2) and a completely illiquid account (that can be used only in period 2). This system does not feature money burning, so, in equilibrium we have $c_1(\theta) + c_2(\theta) = Y$ for all households.[19]

We now embed the AWA result in the conceptual framework described in the introduction: i.e., the three classes of welfare-ranked mechanisms.[20] The AWA result implies that the General Non-linear Mechanism, turns out to be piecewise linear and unconstrained with respect to the order of account depletion. Accordingly, the N-account Mechanism and the General Non-linear Mechanism are identical,

$$W^N = W^G.$$

3 Optimal Liquidity with Homogeneous Present Bias and Inter-Household Transfers

We now study the case in which present bias β is homogeneous across households, but the government can make inter-household transfers. Specifically, we now replace *household-by-household* budget balance (Equation 7) with *overall* budget balance (Equation 6). With overall budget balance, we show in Appendix C that a combination of a perfectly liquid and a perfectly illiquid account is not sufficient to maximize social surplus. Intuitively, when inter-household transfers are possible (in the interior case, with partial separation), we can use an incentive compatible mechanism to redistribute c_1 away from low-θ types (i.e., households with low marginal utility, ceteris paribus). To simplify notation, we set $R = \delta = Y = 1$ for the remainder of the paper.[21] We now turn to studying socially optimal

[19] See Ambrus and Egorov (2013) for cases (that do not satisfy our assumptions) in which money burning arises.

[20] Because we directly know the General Non-linear Mechanism, we do not need to discuss the welfare bound provided by the Relaxed Problem.

[21] This involves no loss of generality because utility functions can be rescaled.

mechanisms in this environment.

3.1 Optimal policy with quasi-linear utility

To gain intuition about socially optimal mechanisms, it is helpful to begin by studying the special case of quasi-linear utility: $u_2(c_2) = c_2$. To anticipate our results for this case, we find that the General Non-linear Mechanism is a linear mechanism: i.e., a single account. Accordingly, the N-account Mechanism and the General Non-linear Mechanism are identical.[22]

With quasi-linear utility, we obtain a useful exact result that captures the intuition behind the general case in which utility is concave in both periods.[23]

Proposition 1 *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 1-account system with a Pigouvian tax on consumption in period 1:*

$$\pi = 1 - \beta.$$

This 1-account system is also first-best efficient.

The proof appears in Appendix E.

Quasi-linear utility in period 2 implies that all agents have the same period-2 marginal utility (regardless of their period-2 consumption). Because marginal transfers to period 2 have the same marginal value for all agents, and because all agents have the same degree of present bias, a homogeneous Pigouvian correction achieves the first best allocation. Although

[22] Because we directly know the General Non-linear Mechanism, we do not need to discuss the welfare bound provided by the Relaxed Problem.

[23] This result is a version of the well known Pigouvian logic from consumption externalities (Diamond, 1973) and present bias (DellaVigna and Malmendier, 2004; Galperti, 2015). DellaVigna and Malmendier (2004) and Galperti (2015) study a setting in which a single household contracts with a firm, subject to a participation constraint. Diamond (1973) studies a population of households with consumption externalities, subject to an aggregate resource constraint.

this is not exactly true in the general case in which the utility function is concave in both periods, the special case of quasi-linear utility turns out to be a good proxy for the case with strictly concave utility in both periods. We next study that case.

3.2 Optimal policy with strictly concave utility

We now return to the 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). We explicitly solve for welfare in the N-account Mechanism and the Relaxed Problem, thereby bounding welfare in the General Non-linear Mechanism.

We begin by discussing the General Non-linear Mechanism and the Relaxed Problem. 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), rather than restricting to an N-account system. Formally, this problem is transformed into a selection of a utility pair, (v_1, v_2), where $v_t = u_t(c_t)$.

In the General Non-linear Mechanism, the planner's problem can be expressed as that of choosing $v_1, v_2 : \Theta \to \mathbb{R}$ to maximize welfare

$$\int \left(\theta\, v_1(\theta) + v_2(\theta)\right)\, f(\theta)\, d\theta \qquad \text{(Planner Objective)}$$

subject to the resource constraint

$$\int \left(Y - C_1(v_1(\theta)) - \frac{1}{R}\, C_2(v_2(\theta))\right)\, f(\theta)\, d\theta \geq 0 \qquad \text{(Budget Constraint)}$$

where $C_t = u_t^{-1}$, that is $c_t(\theta) = C_t(v_t(\theta))$, and the incentive-compatibility constraint, which now has two parts, namely a linear part,

$$0 = \theta\, v_1'(\theta) + \beta\, v_2'(\theta) \qquad \text{(Local IC)}$$

and a monotonic part,

$$0 \leq -v_2'(\theta). \qquad \text{(Monotonicity)}$$

This completes our description of the General Non-linear Mechanism.

The Relaxed Problem is obtained from the General Non-linear Mechanism by *removing* the Monotonicity constraint. The Relaxed Problem generates an upper bound on welfare under the General Non-linear Mechanism.[24] We characterize the solution to the Relaxed Problem using a system of differential equations (see Appendix D). For the homogeneous-β cases that we solve, Monotonicity is actually satisfied, implying that our calculations for the Relaxed Problem generate the exact solution to the General Non-linear Mechanism.

We also numerically solve for N-account Mechanism with $N = 1$ account (i.e., the linear tax case). To preview our results, there is a very small welfare gap between the N-account Mechanism with $N = 1$ account and the Relaxed Problem, implying that the N-account Mechanism with $N = 1$ account is approximately optimal.

In our benchmark simulations, we make the following functional form assumptions (which are motivated in the paragraph that follows and evaluated for robustness in Section 5).

S1. The utility functions in periods 1 and 2 are $u_1(c) = u_2(c) = \ln(c)$;

S2. The density of the multiplicative taste shocks is a truncated[25] 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.

Assumption **S1** implies that the coefficient of relative risk aversion is one, a magnitude that often (approximately) emerges in estimates of lifecycle savings models.[26] Assumption

[24] We are interested in deriving a bound for welfare, not deriving the exact solution. As we explain below, it turns out that our bound is exact for the homogeneous-β case and close for the heterogeneous-β case.

[25] We truncate the tails to generate a strictly positive density on a bounded support and to avoid a lifetime utility function with a negative weight on u_1.

[26] For example, see Gourinchas and Parker (2002) and Laibson et al. (2021).

S2 implies that a one standard deviation taste shock will induce marginal utility in period 1 to change by ±24.2%.[27] We view this as a plausible assumption given the many uninsurable shocks that buffet households, but we are not aware of formal estimates of this parameter. In Section 5, we evaluate the robustness of our paper's findings with respect to variations in our parametric assumptions.

We begin with Table 1, which reports the *improvement* in total welfare, relative to the autarkic benchmark, for different systems of accounts where the planner chooses the optimal x_n and π_n. 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 autarky case as our benchmark in all analyses). The columns of Table 1 represent different cases of *homogeneous* β, starting with $\beta = 0.1$ and progressing to $\beta = 1.0$.[28]

The first two rows of Table 1 report welfare analyses for N-account Mechanisms with, respectively $N = 1$ and $N = 2$ accounts. In N-account Mechanisms, the planner sets the penalty-level for each account, π_n, and the mandatory initial balance for each account, x_n. The third row of Table 1 reports welfare analyses for the Relaxed Problem. Because the Relaxed Problem drops the Monotonicity restriction in the General Non-linear Mechanism, welfare for the Relaxed Problem is an *upper bound* for welfare improvements obtained by moving from an N-account Mechanism to the General Non-linear Mechansim. The fourth row of Table 1 returns to an N-account Mechanism with $N = 2$ accounts. Because we are interested in real-world analogs, in this row we study the special case where we require the planner to set up a completely liquid account (i.e., $\pi_1 = 0$) and a completely illiquid account (i.e., $\pi_2 = 1$).

Table 1 reveals that a simple $N = 1$-account system generates most of the obtainable

[27] This is slightly less than $\sigma = 0.25$ because of the truncation of the deep tails.

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

	Value of β				
	0.1	0.2	0.3	0.4	0.5
N-account Mechanism: $N = 1$	69.658	31.775	17.124	9.630	5.347
N-account Mechanism: $N = 2$	71.648	32.698	17.605	9.882	5.478
Relaxed Problem	71.674	32.748	17.659	9.929	5.511
N-account Mechanism: $N = 2$, $\pi_1 = 0$, $\pi_2 = 1$	71.633	32.648	17.482	9.671	5.196

	Value of β				
	0.6	0.7	0.8	0.9	1.0
N-account Mechanism: $N = 1$	2.794	1.283	0.446	0.067	0.012
N-account Mechanism: $N = 2$	2.860	1.314	0.458	0.070	0.012
Relaxed Problem	2.881	1.325	0.462	0.071	0.014
N-account Mechanism: $N = 2$, $\pi_1 = 0$, $\pi_2 = 1$	2.542	1.018	0.256	0.015	0.000

Table 1: *Welfare gains from four versions of the homogeneous-β model: an N-account Mechanism with $N = 1$; an N-account Mechanism with $N = 2$; the Relaxed Problem, which is the relaxed version of the General Non-linear Mechanism; an N-account Mechanism with $N = 2$ accounts that have exogenously set penalties, specifically $\pi_1 = 0$ and $\pi_2 = 1$, thereby implying that the first account is completely liquid and the second account is completely illiquid in period 1. In the columns we report welfare gains for* 10 *different values of β (namely* $0.1, 0.2, ..., 1.0$*). The welfare gain is calculated as the percentage increase in household wealth that would produce the same average welfare in the autarkic case. Welfare is calculated using the planner's welfare criterion (i.e., without present bias in the welfare objective).*

welfare gains. For example, for $\beta = 0.6$ (a natural value for a homogeneous calibration in light of current estimates in the empirical literature–see Cohen et al. 2020), this $N = 1$ system generates a social-welfare gain equal to 2.794% of the endowment (relative to the autarky reference case). An $N = 2$-account system generates a social-welfare gain equal to 2.860% of the endowment. The Relaxed Problem generates an upper bound on the welfare gains for the General Non-linear Mechanism. Because the Relaxed Problem generates a welfare gain that is 2.881% of the endowment, the welfare gains of extending beyond the N-account Mechanism with $N = 1$ accounts are quantitatively modest.

This analysis also reveals another important feature of the homogeneous-β 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 $N = 1$ accounts, which turns out to be nearly identical to $(1 - \beta)$, both of which are plotted in Figure 1. This near-Pigouvian result

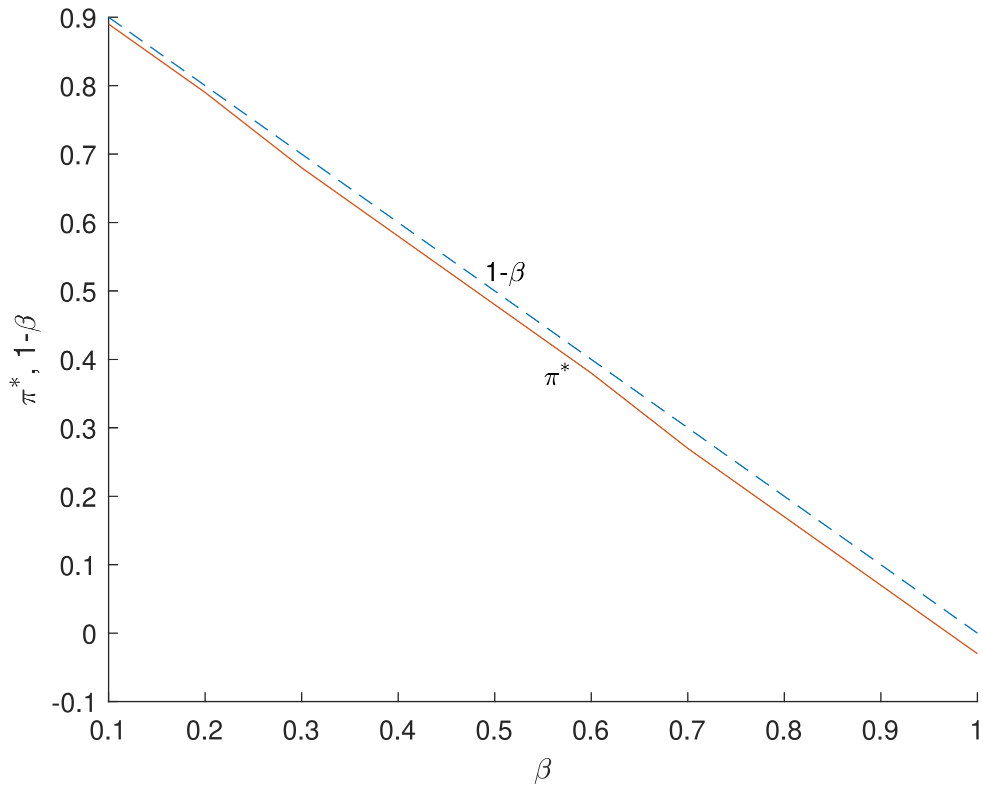

Figure 1: *The optimal penalty π^* and the notional Pigouvian tax $1 - \beta$ as a function of β in the case in which: (i) the population has homogeneous β; (ii) the planner is confined to the N-account Mechanism with $N = 1$ accounts, with penalty π. Note that π^* is always lower than $1 - \beta$. In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

echoes the exact Pigouvian correction that arises in the quasi-linear case (subsection 3.1).[29]

However, an *exact* Pigouvian correction (which *did* arise in the quasi-linear case) is not generally socially optimal because, with concave utility, 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 1-account system (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

[29] 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 $1 - \beta = 0.4$.

$(1-\beta)$ 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.[30]

4 Optimal Liquidity with Heterogeneous Present Bias and Inter-Household Transfers

In this section, we continue to allow inter-household transfers. We relax our assumption of homogeneous β and study a heterogeneous population of β-types. We continue to exploit the revelation principle and study mechanisms in which agents reveal their intertemporal preferences (between periods 1 and 2). Household utility $\theta\, u_1(c_1) + \beta\, u_2(c_2)$ is maximized iff the following expression is also maximized over consumption:

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

Roughly speaking, if we use a continuum of types, the revelation principle can be implemented using variable $\phi = \theta/\beta$.[31] Accordingly, we study mechanisms in which the agents each report ϕ and receive a consumption pair (c_1, c_2) that depends on their report of ϕ. The social planner's objective function (1), therefore takes the form

$$\int \left[E\left[\theta|\phi\right]\, u_1(c_1) + u_2(c_2)\right] dH(\phi).$$

[30] In fact, in our numerical simulations taste shocks are sufficiently large that there are economically significant welfare gains that would be available to a planner with *symmetric* information. However, in our framework, taste shocks are private and incentive compatibility limits the scope for efficient redistribution of resources in the General Non-linear Mechanism. We could extend our model to incorporate observable variation in resources/needs, and, in that case, the planner would engage in full redistribution in those categories.

[31] The primary assumption is that either θ or β or both have non-atomic distributions. It follows that ϕ has a non-atomic distribution. Given the preferences that we have assumed, individual choices will be monotonic, so there are only a countable number of values of ϕ where the set of optimal choices is non-unique. Because ϕ is non-atomic, this set has measure zero. By shifting from truth-telling in (θ, β) to truth-telling in ϕ, we reduce the feasible set of mechanisms, but we do not change the optimal social welfare.

Here, $E[\theta|\phi]$ is the conditional expectation of θ, given a household's value of ϕ. $H(.)$ is the CDF of ϕ.

This representation highlights the importance of the conditional expectation $E[\theta|\phi]$, which is the weight the planner assigns to period 1 utility of a household with value ϕ. Given heterogeneity in both θ and β, taste shock θ cannot be directly inferred from ϕ. For a given impatient choice (i.e., high revealed ϕ), the planner doesn't know whether the household is making that choice because of high θ (i.e., real need), or because of low β (more present bias). The planner wants to give higher period-1 consumption to a high-θ household, but not to low-β household.

As in the previous section, we begin with the quasi-linear case and then provide quantitative simulations.

4.1 Optimal policy with quasi-linear utility

With homogeneous β, a planner knows β even without observing ϕ. By contrast, in the heterogeneous-β case that we are now studying, the planner has to try to infer β and θ from ϕ. Because of this inference problem, exact Pigouvian taxation can not emerge. However, Proposition 2 shows that conditionally expected Pigouvian taxation emerges as the socially optimal mechanism in a quasi-linear economy.[32]

Proposition 2 *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* $\mathrm{E}[\,\theta\,|\,\phi\,]$ *is non-decreasing in* $\phi = \theta/\beta$. *Then the optimal allocation is characterized by*

$$\mathrm{E}[\,\theta\,|\,\phi\,]\, u_1'\,(c_1(\phi)) = 1,$$

[32]This result is connected to Diamond (1973), who shows that a uniform consumption tax on households that generate heterogeneous externalities targets the average externality. This result is also related to Farhi and Werning (2010) and to the independent work of Gerster and Kramm (2023).

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

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

The proof of Proposition 2 is in Appendix E. 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. Heterogeneity in β gives rise to a range of marginal taxes needed to implement the optimum. It is only at the extreme values of ϕ, $\underline{\phi}= \underline{\theta}/\overline{\beta}$, and $\overline{\phi} = \overline{\theta}/\underline{\beta}$, that the planner can *exactly* infer the values of β, respectively $\overline{\beta}$ and $\underline{\beta}$. Accordingly, at these extreme values for ϕ, the planner chooses the most extreme Pigouvian tax rates, respectively, $\pi = 1 - \overline{\beta}$ and $\pi = 1 - \underline{\beta}$.[33]

4.1.1 A case in which the support of β is wide relative to the support of θ

We now introduce a corollary that studies the case in which the distribution of β is uniform and the support of β is wide relative to the support of θ. This gives rise to a mechanism that features a convexly kinked budget set. In particular, in equilibrium households will pool at the kink. This budget set, which is not in general piecewise linear, nevertheless has a key similarity to a simple system of accounts: the opportunity cost of period-one consumption is discretely higher to the right of the pooling region than to the left.

Suppose that

$$\overline{\theta}/\overline{\beta} < \underline{\theta}/\underline{\beta}, \tag{8}$$

[33]The assumption that $E[\theta|\phi]$ increases in ϕ implies that $c_1(\phi)$ is increasing and hence Proposition 2, which is solved as usual by solving the relaxed problem and verifying monoticity ex post, in fact also the optimal general nonlinear. Intuitively, ϕ represents the household's relative valuation of date 1 utility while $E[\theta|\phi]$ represents the planner's average valuation of date 1 utility across households of type ϕ. The assumption that $E[\theta|\phi]$ is increasing implies these two valuations are everywhere positively related, or in other words that β does not fall too quickly as ϕ rises.

in other words that the support of β is wide relative to the support of θ. We call this the β-wide case. We emphasize that β-wide is a relative property. We then have the following corollary of Proposition 2.

Corollary 3 *Suppose that the assumptions of Proposition 2 are satisfied and: (i) the support of* (θ, β) *is* β*-wide and (ii)* β *is uniform. Then,* $E[\theta|\phi]$ *is constant for all* $\phi \in [\overline{\theta}/\overline{\beta}, \underline{\theta}/\underline{\beta}]$*. Consequently,*

1. *The optimal allocation pools all types* $\phi \in [\overline{\theta}/\overline{\beta}, \underline{\theta}/\underline{\beta}]$*.*

2. *There is a jump in the marginal penalty from just below the pooling region to just above the pooling region, that is,*

$$\lim_{\phi \uparrow \overline{\theta}/\overline{\beta}} \pi(\phi) < \lim_{\phi \downarrow \underline{\theta}/\underline{\beta}} \pi(\phi). \tag{9}$$

The key driver of Corollary 3 is that the conditional expectation $E[\theta|\phi]$ is constant in ϕ over an interval, implying the existence of a pooling region. We emphasize that Corollary 3 makes an assumption about the joint support of (θ, β) and an assumption about the density of β on its support. Corollary 3 places *no* restriction on the density of θ on its support. The proof of Corollary 3 is in Appendix E.

Proposition 2 and Corollary 3 are expressed in terms of truth-telling (second-best) optimal mechanisms. We now address the standard problem of reinterpreting these mechanisms in terms of a system of accounts – or more generally a budget set – that represents the institutional analog of the theoretical analysis that has been provided in this subsection.

In practice, the mechanism described in Corollary 3 would be implemented institutionally as a budget set, which is a set of (c_1, c_2) consumption bundles; each household picks a point in this budget set. The opportunity cost of period 1 consumption in terms of period 2 consumption is the slope of the frontier of the budget set. Corollary 3 implies that the budget set has a kink, at which a mass of agents pool. Moreover, Corollary 3 implies that the opportunity cost of period 1 consumption jumps discretely as households move from just below to just above the pooling point.

Finally, we note that a system of accounts is equivalent to a convex budget set with a piecewise linear frontier. Each transition between accounts represents a kink in the budget set. Corollary 3 illustrates that such kinks may also arise in special cases of general mechanisms (i.e., the β-wide case).

4.2 Optimal policy with strictly concave utility

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. As before, each simulation has a different assumption on the the class of mechanisms studied – N-account Mechanism or the Relaxed Problem. Within the class of N-account Mechanisms, we also vary the number of accounts and the scope that the planner has to set withdrawal penalties on those accounts. We maintain simulation assumptions **S1** and **S2** from the previous section. We assume that β is uniformly distributed on the interval $[0.2, 1]$, implying a mean value of 0.6.[34] We explore the robustness of these particular cases in Section 5.

Table 2 reports the welfare improvements (again using a money metric) that are obtained when the planner shifts from the autarky reference system to an N-account Mechanism or a Relaxed Problem. The first row of Table 2 reports the case of an $N = 1$ account system. The second row reports the case of an $N = 2$ account system. The third row reports the case of the Relaxed Problem (see the earlier discussion in Subsection 3.2 and the full derivation in Appendix D). The fourth row reports the case of an N-account Mechanism with $N = 2$ accounts where we require the planner to set up a completely liquid account ($\pi_1 = 0$) and a completely illiquid account ($\pi_2 = 1$). The fifth row reports the case of an N-account Mechanism with $N = 3$ accounts, where the planner has to set a completely liquid account

[34] See Laibson et al. (2021) and Lockwood (2020) for evidence of substantial variation in β as well as mean values of β that are close to 0.6.

N-account Mechanism: $N = 1$	3.569
N-account Mechanism: $N = 2$	6.136
Relaxed Problem	6.144
N-account Mechanism: $N = 2$, $\pi_1 = 0$, $\pi_2 = 1$	6.105
N-account Mechanism: $N = 3$, $\pi_1 = 0$, $\pi_3 = 1$	6.137

Table 2: *Welfare gains from five versions of the heterogeneous-β model (with β distributed uniformly between 0.2 and 1): an N-account Mechanism with $N = 1$ (row 1); an N-account Mechanism with $N = 2$ (row 2); the Relaxed Problem (row 3), which is the relaxed version of the General Non-linear Mechanism; an N-account Mechanism with $N = 2$ accounts that have exogenously set penalties, specifically $\pi_1 = 0$ and $\pi_2 = 1$ (row 4); and an N-account Mechanism with $N = 3$ accounts, where accounts 1 and 3 have exogenously set penalties, specifically $\pi_1 = 0$ and $\pi_3 = 1$, and account 2 has an endogenous penalty (row 5).*

($\pi_1 = 0$) and a completely illiquid account ($\pi_3 = 1$), but can choose the penalty for account 2 (π_2).

Table 2 reveals that an N-account Mechanism with $N = 1$ account no longer obtains most of the feasible welfare gains: one account with a flexible penalty generates a social-welfare gain of only 3.569% of the endowment, well below the upper bound of 6.144% obtained with the Relaxed Problem (row 3).

In contrast, an N-account Mechanism with $N = 2$ accounts gets very close to this upper bound: two flexible accounts generate a social-welfare gain equal to 6.136% of the endowment. For the $N = 2$ case, we find that one penalty is close to zero and the other penalty is close to one. Accordingly, an $N = 2$ account system with a completely liquid and a completely illiquid account also gets very close to the upper bound, at a welfare gain of 6.105% of the endowment. Finally, the $N = 3$ account system with one completely liquid, one partially liquid and one completely illiquid account 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 and a very simple $N = 2$ 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).

The flat conditional expectation, pooling region, and kink in the budget set documented in Corollary 3 provide one intuition for why a two-account system performs better than a single account in this case. We next offer a complementary framework to build intuition by studying comparative statics for the penalty of the partially illiquid account.

4.3 Comparative Statics for the Penalty of the Partially Illiquid Account

To gain further intuition for this result, we report a related set of analyses in Figure 2. Here, we study a 2-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 the five values 0.2, 0.4, 0.6, 0.8 and 1.0. 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.

For the most inconsistent households, with $\beta = 0.2$, money-metric welfare as perceived by the planner rises dramatically as the early-withdrawal penalty increases (Figure 2). The gain to these households from moving from fully liquid, $\pi_2 = 0$, to fully illiquid, $\pi_2 = 1$ is equivalent to an increase of about 30% in their wealth level from the planner's perspective.

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

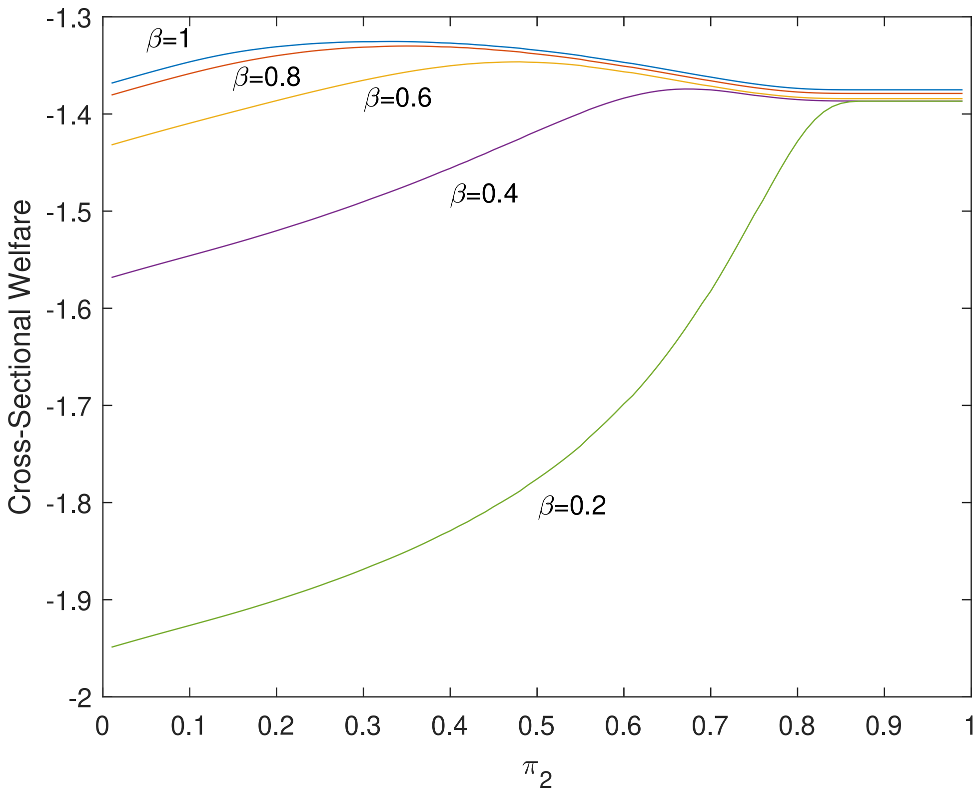

Figure 2: *The welfare of various β cross sections of the population as a function of π_2 in the case in which: (i) the population has heterogeneous β; (ii) the planner is confined to an N-account Mechanism with $N = 2$ accounts, with penalties π_1 and π_2 respectively; (iii) $\pi_1 = 0$ (i.e., the first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

they tend to make smaller early withdrawals and, therefore, pay fewer penalties than low-β households. These differential penalty payments are shown in Figure 3, which reports the gross penalties paid by households with different values of β (again integrating over θ). Penalties are hump-shaped, with lower-β households being willing on average to withdraw more and pay more at all penalty levels.

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 4 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.

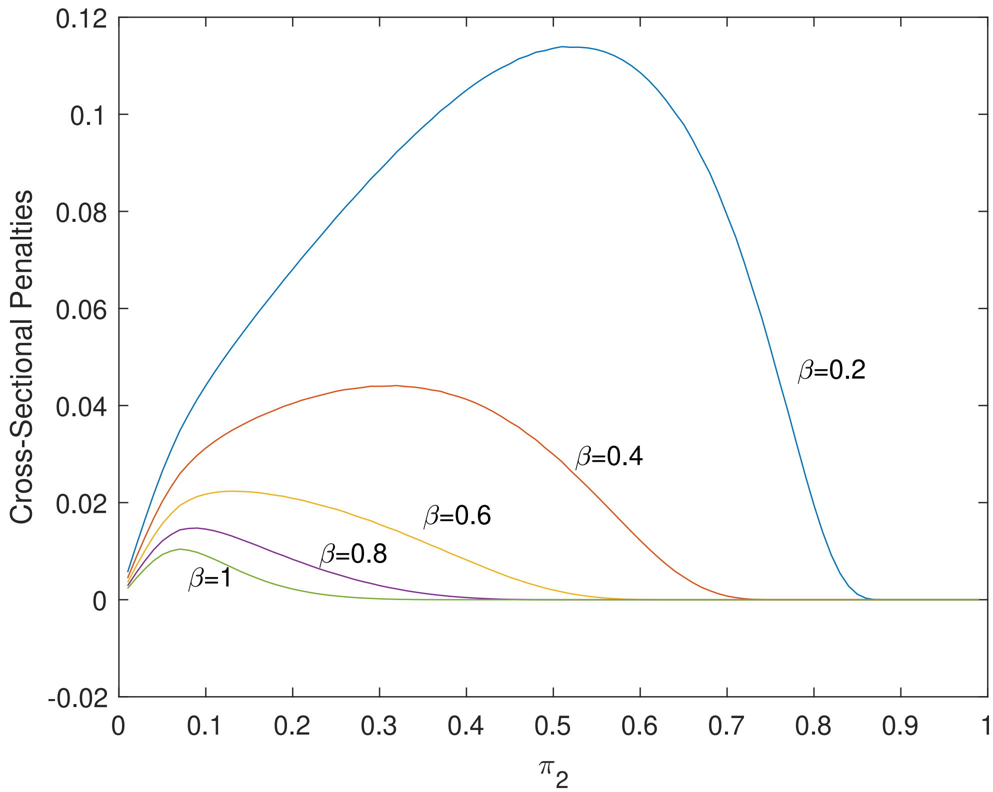

Figure 3: *The total penalties paid by various β cross sections of the population as a function of π_2 in the case in which: (i) the population has heterogeneous β; (ii) the planner is confined to an N-account Mechanism with $N = 2$ accounts, with penalties π_1 and π_2, respectively; (iii) $\pi_1 = 0$ (i.e., the first account is completely liquid); (iv) the account allocations are chosen to maximize the welfare of the population as a whole.*

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

The fifth row in Table 2 reports the welfare gains for an N-account Mechanism with $N = 3$ accounts. 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

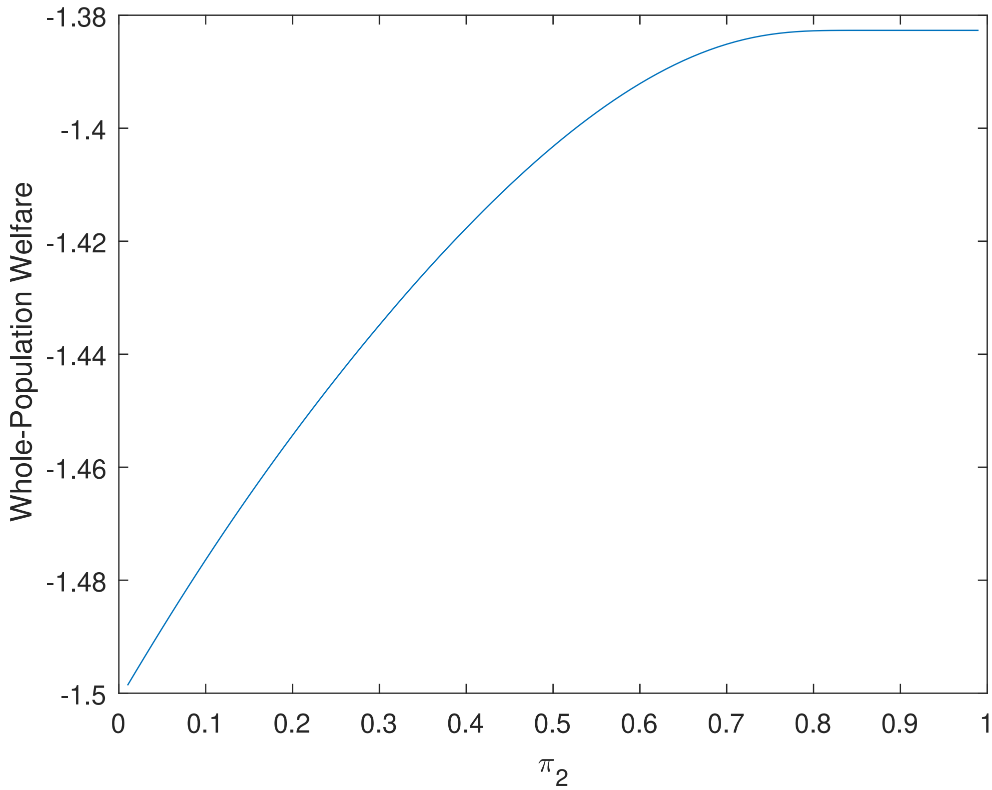

Figure 4: *The welfare of the population as a whole as a function of* π_2 *in the case in which: (i) the population has heterogeneous* β*; (ii) the planner is confined to an* N*-account mechanism with* $N = 2$ *accounts, with penalties* π_1 *and* π_2*, respectively; (iii)* $\pi_1 = 0$ *(i.e., the first 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.*

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 the optimized "middle" account to the constrained two-account system (row 4 in Table 2) slightly raises welfare, by $6.137\% - 6.105\% = 0.032\%$ of wealth.

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 $c_2/c_1 = 0.94$. Figure 5 shows that 76% of households fully draw down the partially illiquid account, while another 22% partially withdraw from it. Only 2% of households choose to withdraw nothing from the partially illiquid 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.[35] 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

[35] 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.

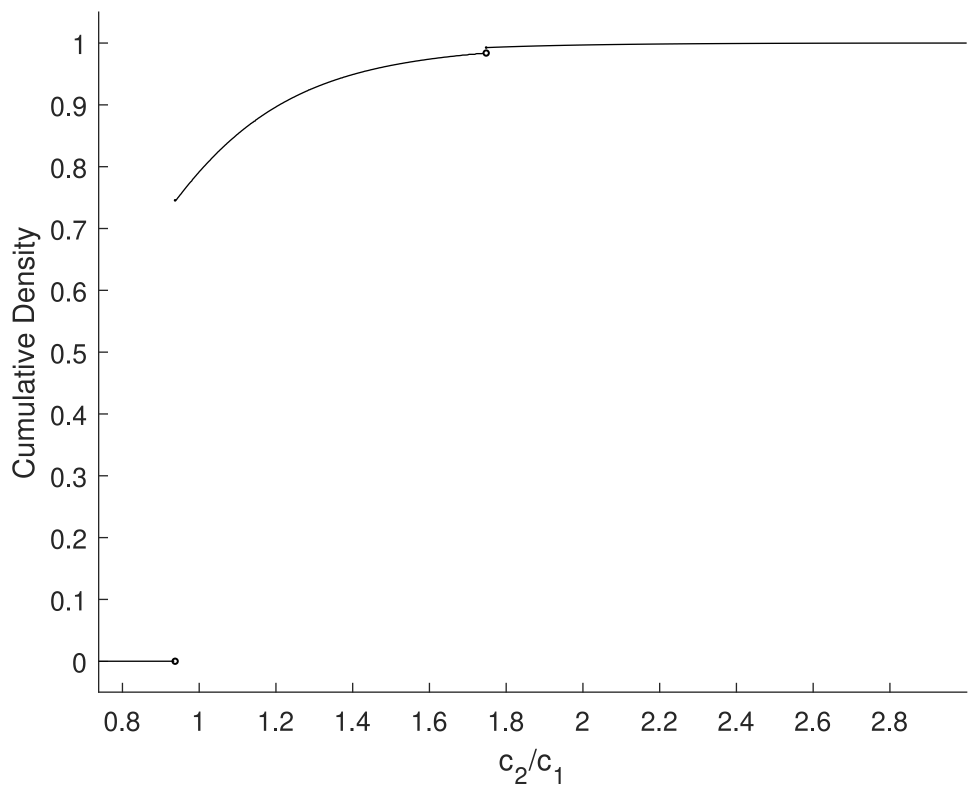

Figure 5: *The distribution function of the ratio c_2/c_1 of period-2 consumption to period-1 consumption in the population as a whole in the case in which: (i) the population has heterogeneous β; (ii) the planner is confined to an N-account Mechanism with $N = 3$ accounts, with penalties π_1, π_2 and 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.*

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 optimally chooses to put almost half of each household's resources into the completely illiquid account (47.4%), with 36.4% going to the completely liquid account and 16.2% going to the partially illiquid account. Accordingly, the completely illiquid account *alone* 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 assets 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 assets 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%.[36] Hence, very high rates of equilibrium leakage are consistent with *optimized* policy in an economy populated by agents with present-bias.

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). Relatedly, the fully liquid account plays a far more important role in practice than it does in our model. In addition, in the United States some withdrawals from retirement accounts are not penalized (e.g., education expenses, large unreimbursed health expenses, the purchase of a first home). To account for these factors, we report an illustrative calibration of the model where we exogenously fix the account balance allocations (rather than endogenously optimizing them)

[36]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.

to reflect the operation of the status quo system in the United States. We *exogenously* allocate 60% of lifetime resources to the liquid account, 10% of lifetime resources to the partially illiquid retirement account with a 0.10 early-withdrawal penalty to match U.S. system, 10% of lifetime resources to another partially illiquid retirement account with a 0.01 early-withdrawl 'penalty' to conceptually capture the fact that some retirement assets are accessible with only small logistical costs (i.e., non-penalized withdrawals), and 20% of assets to the completely illiquid account. With this calibration, we obtain an aggregate leakage rate (total leakage divided by total balances in the two partially illiquid retirement accounts) of 31%, which is within the range of historical leakage rates in the United States (see Argento, Bryant, and Sabelhaus 2015).

5 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 and transfers are allowed. 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. With respect to the second finding, the optimized penalty on the partially illiquid account ranges from 11% to 14% across our calibrated economies, similar to the penalties on 401(k)s and IRAs. With respect to the third finding, the equilibrium leakage rates remain very high, ranging from 84% to 99%.

The detailed results are reported in the three panels of Table 3, which report the welfare gain (relative to the autarky benchmark) 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 Relaxed Problem described in Appendix D. For case (ii), in addition to the welfare gain, we also report the penalty π_2 and the leakage rate.[37]

Table 3a varies the value of the coefficient of relative risk aversion (γ). In Table 3a, we study the cases $\gamma = 1/2$, $\gamma = 1$ (our benchmark, for comparison), and $\gamma = 2$. Table 3b varies the shape of the density of the taste shock θ, changing the variance of the normal distribution between $\sigma = 0.30$, $\sigma = 0.25$ (our benchmark, 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 $\beta = 0.6$. As shown in Section 4, results do not generalize to the degenerate case of homogeneous β (the last column of Table 3c). Gathering these results together, we infer that (at least partially unobservable) heterogeneity in β is necessary for a fully illiquid account to be optimal.

[37] Note that the upper bound on the welfare gain–provided by the Relaxed Problem–is economically close to the N-account Mechanisms that we study.

	Value of γ		
	0.5	1.0	2.0
N-account Mechanism: $N = 2$, $\pi_1 = 0$, $\pi_2 = 1$	8.851	6.105	3.261
N-account Mechanism: $N = 3$, $\pi_1 = 0$, $\pi_3 = 1$	8.919	6.137	3.274
— — — — Penalty π_2^*	0.13	0.13	0.11
— — — — Leakage Rate	0.89	0.90	0.99
Relaxed Problem	8.932	6.144	3.278

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

	Value of σ_θ		
	0.30	0.25	0.20
N-account Mechanism: $N = 2$, $\pi_1 = 0$, $\pi_2 = 1$	5.918	6.105	6.323
N-account Mechanism: $N = 3$, $\pi_1 = 0$, $\pi_3 = 1$	5.958	6.137	6.344
— — — — Penalty π_2^*	0.14	0.13	0.12
— — — — Leakage Rate	0.84	0.90	0.89
Relaxed Problem	5.966	6.144	6.349

(b) *Variation of the standard deviation* σ_θ *of the taste shock*

	Value of σ_β			
	$+\infty$	1.0	0.5	0.0
N-account Mechanism: $N = 2$, $\pi_1 = 0$, $\pi_2 = 1$	6.105	6.019	5.772	2.542
N-account Mechanism: $N = 3$, $\pi_1 = 0$, $\pi_3 = 1$	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
Relaxed Problem	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 sub-table: row 1 contains welfare gains from an N-account Mechanism with $N = 2$ accounts that have exogenously set penalties, specifically $\pi_1 = 0$ and $\pi_2 = 1$; row 2 contains welfare gains from an N-account Mechanism with $N = 3$ accounts, where accounts 1 and 3 have exogenously set penalties, specifically $\pi_1 = 0$ and $\pi_3 = 1$, and account 2 has an endogenous penalty; 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 from the Relaxed Problem. 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 Conclusions and Directions for Future Work

We focus our summary 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 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 2015).

Despite these superficial similarities, it is inappropriate to conclude that our findings demonstrate the social optimality of the U.S. system. Most importantly, our theoretical model includes key simplifications.[38] First, we assume a particular conceptual formulation of self-defeating behavior (present bias).[39] Second, we assume only two periods (e.g., working life and retirement).[40] We anticipate that generalization to many periods without labor income – i.e., decades of retirement – would engender optimal policy characterized by a stream of illiquid payments (instead of a single illiquid account); such a stream mirrors the annuity payments that characterize most defined benefit pension plans. Third, we assume a particular form of multiplicative taste shifter, θ.[41] Fourth, we assume that households are naive with respect to their present bias parameter, β. Fifth, we study a limited set of distributions of θ and β (and no correlation).[42] Sixth, our economy has a fixed endowment, Y, which is not endogenously responsive to the tax (and redistribution) system. Seventh, our models omits heterogeneity in income or endowments (e.g., Mirrlees 1971), which weakens the motive for redistribution. Our framework could be readily generalized to incorporate observable type heterogeneity, which would require the inclusion of labor income taxation. Incorporating unobservable heterogeneity would require a substantial change in our framework.

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,[43] raising the normative possibility that mandatory savings might be underutilized in the U.S.[44] However,

[38] In addition, the U.S. system contains some scope for tax arbitrage, which is not present in our model.

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

[40] 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 (2013) and Cao and Werning (2018).

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

[42] Research is only beginning on the distribution of present bias. For analysis of this issue, see Moser and Olea de Souza e Silva (2019), Lockwood (2020), and Cohen et al. (2020).

[43] See Stephens Jr. and Toohey (2018).

[44] In our model, mandatory savings are achieved through a funded system. Our model takes no position

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 to retirement.[45]

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 clear whether these results will continue to hold as future research enriches and expands our understanding of household behavior and optimal policy.

on the distinction between funded (e.g., the superannuation scheme in Australia) and unfunded (e.g., U.S. Social Security) mandatory savings systems.

[45]See Beshears et al. (2018) for a recent review of the literature on consumption dynamics at and through retirement.

References

Allcott, Hunt, Joshua Kim, Dmitry Taubinsky, and Jonathan Zinman. 2022. "Are High-Interest Loans Predatory? Theory and Evidence from Payday Lending." *The Review of Economic Studies* 89 (3):1041–1084.

Allcott, Hunt, Benjamin B. Lockwood, and Dmitry Taubinsky. 2019. "Regressive Sin Taxes, with an Application to the Optimal Soda Tax." *The Quarterly Journal of Economics* 134 (3):1557–1626.

Amador, Manuel, Iván Werning, and George-Marios Angeletos. 2006. "Commitment vs. flexibility." *Econometrica* 74 (2):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. 2015. "Early Withdrawals from Retirement Accounts During the Great Recession." *Contemporary Economic Policy* 33 (1):1–16.

Atkeson, Andrew and Robert E. Lucas Jr. 1992. "On efficient distribution with private information." *The Review of Economic Studies* 59 (3):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 (3):1067–1115.

Bernheim, B. Douglas and Antonio Rangel. 2004. "Addiction and cue-triggered decision processes." *American Economic Review* 94 (5):1558–1590.

Bernheim, B. Douglas and Dmitry Taubinsky. 2018. "Behavioral Public Economics." In *Handbook of Behavioral Economics: Applications and Foundations*, vol. 1, edited by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson. Elsevier, 381–516.

Beshears, John, James J. Choi, Christopher Harris, David Laibson, Brigitte C. Madrian, and Jung Sakong. 2020. "Which early withdrawal penalty attracts the most deposits to a commitment savings account?" *Journal of Public Economics* 183:104144.

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 (5):420–425.

Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian. 2018. "Behavioral household finance." In *Handbook of Behavioral Economics: Applications and Foundations 1*, vol. 1, edited by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson. Elsevier, 177–276.

Bond, Philip and Gustav Sigurdsson. 2018. "Commitment contracts." *The Review of Economic Studies* 85 (1):194–222.

Bower, Anthony G. 1993. "Procurement policy and contracting efficiency." *International Economic Review* 34 (4):873–901.

Camerer, Colin, Samuel Issacharoff, George Loewenstein, Ted O'Donoghue, and Matthew Rabin. 2003. "Regulation for conservatives: behavioral economics and the case for 'asymmetric paternalism'." *University of Pennsylvania Law Review* 151 (3):1211–1254.

Cao, Dan and Iván Werning. 2018. "Saving and dissaving with hyperbolic discounting." *Econometrica* 86 (3):805–857.

Chu, Leon Yang and David E. M. Sappington. 2007. "Simple cost-sharing contracts." *American Economic Review* 97 (1):419–428.

Cohen, Jonathan, Keith Marzilli Ericson, David Laibson, and John Myles White. 2020. "Measuring time preferences." *Journal of Economic Literature* 58 (2):299–347.

DellaVigna, Stefano and Ulrike Malmendier. 2004. "Contract design and self-control: Theory and evidence." *The Quarterly Journal of Economics* 119 (2):353–402.

———. 2006. "Paying not to go to the gym." *American Economic Review* 96 (3):694–719.

DellaVigna, Stefano and M. Daniele Paserman. 2005. "Job search and impatience." *Journal of Labor Economics* 23 (3):527–588.

Diamond, Peter A. 1973. "Consumption externalities and imperfect corrective pricing." *The Bell Journal of Economics and Management Science* 4 (2):526–538.

Ericson, Keith Marzilli and David Laibson. 2019. "Intertemporal choice." In *Handbook of Behavioral Economics: Applications and Foundations 1*, vol. 2, edited by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson. Elsevier, 1–67.

Farhi, Emmanuel and Xavier Gabaix. 2020. "Optimal taxation with behavioral agents." *American Economic Review* 110 (1):298–336.

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

Fudenberg, Drew and David K. Levine. 2006. "A dual-self model of impulse control." *American Economic Review* 96 (5):1449–1476.

Galperti, Simone. 2015. "Commitment, flexibility, and optimal screening of time inconsistency." *Econometrica* 83 (4):1425–1465.

Gasmi, Farid, Jean-Jacques Laffont, and William W. Sharkey. 1999. "Empirical evaluation of regulatory regimes in local telecommunications markets." *Journal of Economics & Management Strategy* 8 (1):61–93.

Gerster, Andreas and Michael Kramm. 2023. "Optimal Internality Taxation of Product Attributes."

Giné, 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 (4):213–235.

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

Gruber, Jonathan and Botond Köszegi. 2001. "Is addiction "rational"? Theory and evidence." *The Quarterly Journal of Economics* 116 (4):1261–1303.

———. 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 (6):1403–1435.

Halac, Marina and Pierre Yared. 2014. "Fiscal rules and discretion under persistent shocks." *Econometrica* 82 (5):1557–1614.

Harris, Christopher and David Laibson. 2013. "Instantaneous gratification." *The Quarterly Journal of Economics* 128 (1):205–248.

Heidhues, Paul and Botond Kőszegi. 2010. "Exploiting naivete about self-control in the credit market." *American Economic Review* 100 (5):2279–2303.

———. 2018. "Behavioral industrial organization." In *Handbook of Behavioral Economics: Applications and Foundations 1*, vol. 1, edited by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson. Elsevier, 517–612.

Hewitt Associates. 2009. "The Erosion of Retirement Security from Cash-outs: Analysis and Recommendations." http://www.aon.com/human-capital-consulting/ thought-leadership/retirement/reports-pubs retirement cash outs.jsp. Accessed: April 18, 2011.

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

Laibson, David, Sean Chanwook Lee, Peter Maxted, Andrea Repetto, and Jeremy Tobacman. 2021. "Estimating discount functions with consumption choices over the lifecycle." Tech. rep., revised version of NBER Working Paper 13314.

Laibson, David, Andrea Repetto, and Jeremy Tobacman. 2003. *"A Debt Puzzle"*. In Knowledge, Information, and Expectations in Modern Economics: In Honor of Edmund S. Phelps. Eds. Philippe Aghion, Roman Frydman, Joseph Stiglitz, and Michael Woodford, Princeton: Princeton University Press, 228–266.

Laibson, David I., Andrea Repetto, and Jeremy Tobacman. 1998. "Self-control and saving for retirement." *Brookings Papers on Economic Activity* 1998 (1):91–196.

Lockwood, Benjamin B. 2020. "Optimal income taxation with present bias." *American Economic Journal: Economic Policy* 12 (4):298–327.

Loewenstein, George and Ted O'Donoghue. 2004. "Animal spirits: Affective and deliberative processes in economic behavior." *Available at SSRN 539843* .

Maxted, Peter. 2022. "Present Bias Unconstrained: Consumption, Welfare, and the Present-Bias Dilemma." Tech. rep., Mimeo.

McAfee, R. Preston. 2002. "Coarse matching." *Econometrica* 70 (5):2025–2034.

Meier, Stephan and Charles Sprenger. 2010. "Present-biased preferences and credit card borrowing." *American Economic Journal: Applied Economics* 2 (1):193–210.

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

Moser, Christian and Pedro Olea de Souza e Silva. 2019. "Optimal paternalistic savings policies." *Columbia Business School Research Paper* (17-51).

O'Donoghue, Ted and Matthew Rabin. 1999a. "Doing it now or later." *American Economic Review* 89 (1):103–124.

———. 1999b. "Incentives for procrastinators." *The Quarterly Journal of Economics* 114 (3):769–816.

———. 2006. "Optimal sin taxes." *Journal of Public Economics* 90 (10-11):1825–1849.

Paluszynski, Radoslaw and Pei Cheng Yu. 2019. "Optimal taxation with risky human capital and retirement savings." In *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, vol. 112. JSTOR, 1–59.

Pavoni, Nicola and Hakki Yazici. 2016. "Optimal Life-Cycle Capital Taxation Under Self-Control Problems." *The Economic Journal* 127 (602):1188–1216.

Phelps, Edmund S. and Robert A. Pollak. 1968. "On second-best national saving and game-equilibrium growth." *The Review of Economic Studies* 35 (2):185–199.

Reichelstein, Stefan. 1992. "Constructing incentive schemes for government contracts: An application of agency theory." *Accounting Review* 67 (4):712–731.

Rogerson, William P. 2003. "Simple menus of contracts in cost-based procurement and regulation." *American Economic Review* 93 (3):919–926.

Sappington, David E. M. and Dennis L. Weisman. 1996. *Designing incentive regulation for the telecommunications industry.* Washington, DC: American Enterprise Institute Press.

Shapiro, Jesse M. 2005. "Is there a daily discount rate? Evidence from the food stamp nutrition cycle." *Journal of Public Economics* 89 (2-3):303–325.

Stephens Jr., Melvin and Desmond Toohey. 2018. "Changes in nutrient intake at retirement." Tech. rep., National Bureau of Economic Research.

Strotz, R. H. 1955. "Myopia and Inconsistency in Dynamic Utility Maximization." *The Review of Economic Studies* 23 (3):165–180. URL https://doi.org/10.2307/2295722.

Sulka, Tomasz. 2022. "Exploitative contracting in a life cycle savings model." *Available at SSRN 3507413* .

Thaler, Richard H. and Hersh M. Shefrin. 1981. "An economic theory of self-control." *Journal of Political Economy* 89 (2):392–406.

Yu, Pei Cheng. 2021. "Optimal retirement policies with present-biased agents." *Journal of the European Economic Association* 19 (4):2085–2130.

Online Appendices

A Optimal Liquidity with Homogeneous Present Bias and No Inter-household Transfers

In this section, we consider a first deviation from the (autarky) 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

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

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:

$$c_1 + \frac{c_2}{R} \leq Y \text{ for each household.} \tag{10}$$

To simplify notation, we henceforth we set $\delta = 1$, $R = 1$ and $Y = 1$.[46]

We now formulate a version of a proposition by AWA (2006).

We begin by denoting the support of the taste shifter θ by $\Theta = [\,\underline{\theta}, \overline{\theta}\,]$, where $0 < \underline{\theta} < \overline{\theta} < \infty$. We denote the distribution function of θ by $F : (0, \infty) \to [0, 1]$; we denote the density function of θ by $F' : (0, \infty) \to [0, \infty)$; and, following AWA (2006), we define a function $\Gamma : (0, \infty) \to \mathbb{R}$ by the formula

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

[46] This involves no loss of generality because the utility function can be rescaled.

Next, we define the "pooling type" θ_1 to be the minimum $\theta \in (0, \overline{\theta})$ such that

$$\tfrac{1}{\overline{\theta}-t} \int_t^{\overline{\theta}} \Gamma(s) ds \geq 1 \text{ for all } t \in [\theta, \overline{\theta}).$$

Notice that θ_1 is well-defined. Indeed, $\theta_1 > 0$ and – if we denote by Θ_1 the set of all $\theta \in (0, \overline{\theta})$ such that $\frac{1}{\overline{\theta}-t} \int_t^{\overline{\theta}} \Gamma(s)\, ds \geq 1$ for all $t \in [\theta, \overline{\theta})$ – then Θ_1 is the non-empty half-open interval $[\theta_1, \overline{\theta})$. However, it is entirely possible that θ_1 is a "hypothetical" type, in the sense that $\theta_1 < \underline{\theta}$.[47]

Our candidate for an optimal consumption allocation is then obtained by requiring that: (i) all types in the "separating interval" $\Theta_S = \{\theta \mid \theta \in \Theta,\ \theta < \theta_1\}$ choose freely from the unconstrained budget line, namely the set of all (c_1, c_2) such that $c_1 \geq 0$, $c_2 \geq 0$ and $c_1 + c_2 = 1$; and (ii) all types in the "pooling interval" $\Theta_P = \{\theta \mid \theta \in \Theta,\ \theta \geq \theta_1\}$ receive the allocation that the (possibly hypothetical) type θ_1 would choose freely from the unconstrained budget line. Notice that Θ_S may be empty, but that Θ_P never is.

If this construction is to work, then we need to ensure that all the allocated consumption bundles lie in the interior of the unconstrained budget line. If $\theta_1 > \underline{\theta}$, then this will be the case if and only if: (i) the most patient of the relevant types, namely $\underline{\theta}$, would choose $c_1 > 0$ from the unconstrained budget line; and (ii) the least patient of the relevant types, namely θ_1, would choose $c_2 > 0$ from the unconstrained budget line. If $\theta_1 \leq \underline{\theta}$, then the only relevant type is the pooling type θ_1, and we need only require that this type chooses both $c_1 > 0$ and $c_2 > 0$ from the unconstrained budget line.[48]

Finally, we need to ensure that the Lagrange multiplier used in the sufficiency argument is non-negative. To that end, we assume that Γ is non-decreasing on the separating interval $\Theta_S = [\underline{\theta}, \theta_1)$.[49] Notice that, if $\theta_1 \leq \underline{\theta}$, then Θ_S is empty; so in that case this assumption places no restriction on Γ.

We now enumerate all of our assumptions.

A1 u_1, u_2 are twice continuously differentiable, with $u_1', u_2' > 0$ and $u_1'', u_2'' < 0$.

[47]It is helpful to compare our definition of θ_1 with AWA's (2006) definition of θ_p. AWA define θ_p to be the minimum value of $\theta \in [\underline{\theta}, \overline{\theta})$ such that $\int_t^{\overline{\theta}} (1 - \Gamma(s))\, ds \leq 0$ for all $t \in [\theta, \overline{\theta})$. Hence θ_p and θ_1 are related by the formula $\theta_p = \max\{\theta_1, \underline{\theta}\}$. Hence AWA's Proposition 3 holds when $\theta_p > \underline{\theta}$, in which case, $\theta_p = \theta_1$. AWA's Proposition 3 does not, however, hold when $\theta_p = \underline{\theta}$. To see why, consider the following counterexample. Suppose that $\overline{\theta} - \underline{\theta}$ is small and that $1 - \beta$ is large. Then offering different consumption bundles to different θ is not a priority for the planner, but preventing overconsumption is. So the planner will want to choose a pooling type strictly less than $\underline{\theta}$.

[48]A simple sufficient condition ensuring that all the allocated consumption bundles lie in the interior of the unconstrained budget line is therefore that $u_1'(0+) = u_2'(0+) = +\infty$.

[49]It is helpful to compare our Assumption A4 with AWA's (2006) Assumption A. AWA assume that Γ is non-decreasing on the interval $[\underline{\theta}, \theta_p]$.

A2 $u_1'(0+) = u_2'(0+) = \infty$.

A3 F' is a function of bounded variation.[50]

A4 Γ is non-decreasing on the separating interval $\Theta_S = [\,\underline{\theta}, \theta_1)$.

A5 $0 < \beta < 1$.

Proposition 4 *(Cf. Proposition 3 of AWA (2006).) Suppose that β is the same for all households. Suppose further that inter-household transfers are not possible. Then welfare is maximized by dividing the endowment between two accounts: a completely liquid account (that can be used in both period 1 and period 2) and a completely illiquid account (that can be used only in period 2). In particular, types in the separating interval Θ_S – which consists of those $\theta \in \Theta$ such that $\theta < \theta_1$, and which will be empty if $\theta_1 \leq \underline{\theta}$ – choose c_1 strictly less than the balance of the liquid account; and types in the pooling interval Θ_P – which consists of those $\theta \in \Theta$ such that $\theta \geq \theta_1$, and which is never empty – set c_1 equal to the balance of the liquid account (and therefore set c_2 equal to the balance of the completely illiquid account).*[51]

In other words, in the case of homogeneous β, no inter-household transfers and a weak restriction on the distribution function of the taste shifter θ, the socially optimal allocation is achieved with an N-account Mechanism with $N = 2$: one account that is completely liquid, and a second account that is completely illiquid in period 1 and completely liquid in period 2. Additional accounts (with intermediate levels of liquidity) do not have any value.

This proposition embeds two cases: in one case ($\theta_1 > \underline{\theta}$), some types are separated and some types are pooled; and in the other case ($\theta_1 \leq \underline{\theta}$), all agents are pooled. We emphasize that, in the second case, it is entirely possible that $\theta_1 < \underline{\theta}$. In other words, the pooling type θ_1 is a hypothetical type that is not a member of the population Θ. Either way, all types $\theta \in \Theta$ with $\theta \geq \theta_1$ pool on the choice that type θ_1 would make from the unconstrained budget line. The key difference between our analysis and that of AWA (2006) is that their analysis covers the case $\theta_1 > \underline{\theta}$, whereas our analysis holds for all values of θ_1.[52]

[50] Intuitively speaking, F' is a function of bounded variation iff there exists a bounded Borel measure F'' on $(0, \infty)$ such that F' is the distribution function of F''. For example, if F'' assigns mass 1 to the point 1 and mass -1 to the point 2 (and assigns no mass to any other point) then F' will be the density of the uniform distribution on $[1, 2]$. More generally: (i) the truncation to the interval $[\,\underline{\theta}, \overline{\theta}\,]$ of the densities of most named distributions are functions of bounded variation; and (ii) any step function, the support of which is contained in $[\,\underline{\theta}, \overline{\theta}\,]$, is a function of bounded variation. See Appendices B.3 and B.4 for a detailed discussion of functions of bounded variation.

[51] In particular, no money burning arises in equilibrium. See Ambrus and Egorov (2013) for cases (that do not satisfy our assumptions) in which money burning arises.

[52] There is another important difference between our analysis and AWA's. The original AWA proof shows that the two-account system is optimal in the class of *continuous* incentive-compatible consumption alloca-

In summary, Proposition 4 implies that no gain in welfare is achieved by increasing the number of accounts beyond $N = 2$ in the family of N-account Mechanisms (equations 2-6). But the proposition relies on two strong assumptions – homogeneous β and no inter-household transfers. We next analyze the model in the case in which the latter assumption does not hold.

B Proof of Proposition 4

B.1 Formulation of the Proposition

In the main text we assumed that the income Y of a household was 1 and that the total mass $F(\overline{\theta})$ of households was 1. This was done in order to reduce notation. In this appendix we will work with general Y and general $F(\overline{\theta})$, since it is easier to follow the derivations in the general case.

The first step in formulating Proposition 4 is then to define θ_1 in this more general setting. Recalling that the function Γ is given by the formula

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

we define θ_1 to be the minimum $\theta \in (0, \overline{\theta})$ such that

$$\frac{1}{\overline{\theta}-t}\int_t^{\overline{\theta}} \Gamma(s)ds \geq F(\overline{\theta}) \text{ for all } t \in [\theta, \overline{\theta}).$$

for all $t \in [\theta, \overline{\theta})$. Assumptions A1-A5 are then assumed to hold exactly as stated in the main text. Finally, we restate Proposition 4 for the reader's convenience.

4

Proposition 0 *(Cf. Proposition 3 of AWA (2006).) Suppose that β is the same for all households. Suppose further that inter-household transfers are not possible. Then welfare is maximized by dividing the endowment between two accounts: a completely liquid account (that can be used in both period 1 and period 2) and a completely illiquid account (that can be used only in period 2). In particular types in the separating interval Θ_S – which consists of those $\theta \in \Theta$ such that $\theta < \theta_1$, and which will be empty if $\theta_1 \leq \underline{\theta}$ – choose c_1 strictly less*

tions, whereas our proof shows that the two-account system is optimal in the class of *all* incentive-compatible consumption allocations. This is potentially important because many incentive-compatible consumption allocations are in fact discontinuous. For example, suppose that there is a type $\theta_2 \in (\underline{\theta}, \overline{\theta})$ and two consumption bundles $\underline{c}$ and $\overline{c}$ such that all types in $[\underline{\theta}, \theta_2)$ choose $\underline{c}$ and all types in $(\theta_2, \overline{\theta}]$ choose $\overline{c}$. Then there is a jump in the allocation at θ_2.

than the balance of the liquid account; and types in the pooling interval Θ_P – which consists of those $\theta \in \Theta$ such that $\theta \geq \theta_1$, and which is never empty – set c_1 equal to the balance of the liquid account (and therefore set c_2 equal to the balance of the completely illiquid account).

Our proposition generalizes AWA's analysis in two respects. First, AWA's analysis covers the case $\theta_1 > \underline{\theta}$, whereas our analysis holds for all values of θ_1. Second, AWA's analysis shows that the two-account system is optimal in the class of *continuous* incentive-compatible consumption allocations, whereas our analysis shows that the two-account system is optimal in the class of *all* incentive-compatible consumption allocations. The first point could be expressed by saying that AWA's analysis covers the partial-separation case, whereas our analysis covers both the pooling and the partial-separation case. The second point is important, because many incentive-compatible consumption allocations – including some of the simplest possible incentive-compatible consumption allocations – are discontinuous.

B.2 A Candidate Utility Allocation

Our strategy of proof is to construct a candidate utility allocation and a candidate Lagrange multiplier, and then show that the utility allocation maximises the Lagrangian when violations of the resource constraint are penalized using the Lagrange multiplier.

We begin by constructing a candidate *consumption* allocation. This is obtained by requiring that: (i) all types θ in the separating interval $\Theta_S = \{\theta \mid \theta \in \Theta,\ \theta < \theta_1\} = [\underline{\theta}, \theta_1)$ choose freely from the unconstrained budget line, namely the set of all (c_1, c_2) such that $c_1 \geq 0$, $c_2 \geq 0$ and $c_1 + c_2 = Y$; and (ii) all types θ in the pooling interval $\Theta_P = \{\theta \mid \theta \in \Theta,\ \theta \geq \theta_1\} = [\max\{\underline{\theta}, \theta_1\}, \overline{\theta}]$ receive the allocation that the (possibly hypothetical) type θ_1 would choose freely from the unconstrained budget line.[53]

We transform the candidate consumption allocation $(c_1, c_2) : \Theta \to (0\,\infty)$ into a candidate *utility* allocation $(r_1, r_2) : \Theta \to \mathbb{R}$ by setting $r_1(\theta) = u_1(c_1(\theta))$ and $r_2(\theta) = u_2(c_2(\theta))$. We would like to show that the utility allocation (r_1, r_2) is optimal among *all* economically meaningful utility allocations (v_1, v_2).

This sets up a mathematical hurdle. For, while (r_1, r_2) itself is fairly regular (it is a continuously differentiable function of θ with a kink at θ_1), the alternative utility allocations (v_1, v_2) may not even be continuous. We will get over this hurdle by using the one regularity

[53] The consumption allocation will be interior if and only if

$$\frac{u_2'(Y)}{u_1'(0+)} < \frac{\min\{\theta_1, \underline{\theta}\}}{\beta} \leq \frac{\theta_1}{\beta} < \frac{u_2'(0+)}{u_1'(Y)}.$$

Assumption A2 obviously implies this condition.

property that incentive-compatible utility allocations do possess: they are monotonic. Hence they are functions of bounded variation.

B.3 Functions of Bounded Variation on Θ

There are a number of competing definitions of a function of bounded variation. According to one elementary definition, a function $f : \Theta \to \mathbb{R}$ is of bounded variation iff it is the difference of two bounded and non-decreasing functions $f_+, f_- : \Theta \to \mathbb{R}$. The most serious drawback with this definition for our purposes is that the functions defined in this way do not form a function space. This definition cannot therefore be used in a Lagrangian analysis. A second drawback of the definition is that it does not capture the behaviour of a function of bounded variation at the endpoints of Θ. We shall therefore adopt a definition that leads directly to a usable function space, and which ties down the behaviour of a function at the endpoints of Θ.

The intuitive idea is to say that f is a function of bounded variation on Θ iff it is the distribution function of a bounded Borel measure on Θ plus a constant of integration. More precisely, we begin from a constant of integration, denoted suggestively by $f_L(\underline{\theta})$, and a bounded Borel measure on Θ, denoted suggestively by f'. We then define the left-hand limits f_L of f by

$$f_L(\theta) = f_L(\underline{\theta}) + f'([\underline{\theta}, \theta))$$

for all $\theta \in \Theta$ (including $\underline{\theta}$) and the right-hand limits f_R of f by

$$f_R(\theta) = f_L(\underline{\theta}) + f'([\underline{\theta}, \theta])$$

for all $\theta \in \Theta$ (including $\overline{\theta}$). And we endow the set of functions obtained in this way with the norm

$$\|f\|_{\mathcal{BV}} = |f_L(\underline{\theta})| + \|f'\|_{TV},$$

where $\|\cdot\|_{TV}$ is the total-variation norm on bounded Borel measures on Θ.

This definition has at least three advantages: it is concrete; it builds on familiar ideas like distribution functions and the total-variation norm; and it brings out the subtleties implicit in the concept of a function of bounded variation. One subtlety is the fact that a "function" of bounded variation is not a function in the narrow sense of that word: it is only well defined where $f_L = f_R$, and there may be a countable set of points at which this is not the case. (These points are precisely the atoms of the bounded Borel measure f'. As such, they may include the endpoints $\underline{\theta}$ and $\overline{\theta}$.) A second subtlety is the fact that a function of bounded variation has limits from both the left and right at *all* points of Θ, including a limit from the left at $\underline{\theta}$ and a limit from the right at $\overline{\theta}$. (This makes perfect sense if one views functions of bounded variation on Θ as restrictions to Θ of functions of bounded variation on $(0,\infty)$.) In view of these subtleties, one cannot simply adopt a convention that all functions of bounded variation are (say) right continuous.

B.4 Functions of Bounded Variation on $(0,\infty)$

The discussion of the previous section applies *mutatis mutandis* to functions of bounded variation on $(0,\infty)$. The main differences are that: (i) we do not need to consider behaviour at the endpoints of the interval $(0,\infty)$; and (ii) it is preferable to specify the constant of integration at an interior point. Rather than work through this material in general, we shall simply discuss the special case of F'.

We note first that – according to Assumption A2 – the support of F' (as a function) is contained in Θ. It follows, first, that $F'_L(\,\underline{\theta}\,) = F'_R(\,\overline{\theta}\,) = 0$. It follows, second, that the support of F'' (as a measure) is contained in Θ. In other words, $|F''|\,((0,\underline{\theta})) = |F''|\,((\underline{\theta},\infty)) = 0$.

Now, because $|F''|\,((0,\underline{\theta})) = 0$, we can suppress the constant of integration in the formulae for F' in terms of F''. More explicitly, we have

$$\begin{aligned} F'_L(\theta) &= F''((0,\theta)), \\ F'_R(\theta) &= F''((0,\theta]) \end{aligned}$$

for all $\theta > 0$. It then follows that

$$\begin{aligned} 0 &= F'_R(\,\overline{\theta}\,) = F''((\,0,\overline{\theta}\,]) = F''((\,0,\underline{\theta}\,)) + F''([\,\underline{\theta},\overline{\theta}\,]) \\ &= F'_L(\,\underline{\theta}\,) + F''([\,\underline{\theta},\overline{\theta}\,]) = F''([\,\underline{\theta},\overline{\theta}\,]). \end{aligned}$$

In other words, F'' assigns total mass 0 to Θ.

B.5 The Lagrangian

Denote by $\mathcal{BV}(\Theta,\mathbb{R})$ the Banach space of functions of bounded variation on Θ with the norm $\|\cdot\|_{\mathcal{BV}}$, and by

$$\mathcal{O}_t = \mathcal{BV}(\Theta,(u_t(0+),u_t(\infty-)))$$

the subset of $\mathcal{BV}(\Theta, \mathbb{R})$ consisting of functions taking values in the interior of the range of u_t. (Recall that u_t is the utility function for date t.) Denote by Ω the set of utility allocations

$$v = (v_1, v_2) \in \mathcal{O}_1 \times \mathcal{O}_2$$

such that

$$\theta\, v_1' + \beta\, v_2' = 0 \tag{ICL}$$

and

$$v_1' \geq 0. \tag{ICM}$$

(The idea here is to split the incentive-compatibility condition into the linear part IC<u>L</u> and the monotonic part IC<u>M</u>.) In other words, let Ω be the set of incentive-compatible utility allocations. Define the objective function

$$M : \mathcal{BV}(\Theta, \mathbb{R})^2 \to \mathbb{R}$$

by the formula

$$M(v) = \int (\theta\, v_1 + v_2)\, F'\, \ell(d\theta),$$

where ℓ is Lebesgue measure, and define the budget operator

$$N : \mathcal{O}_1 \times \mathcal{O}_2 \to \mathcal{BV}(\Theta, \mathbb{R})$$

by the formula

$$(N(v))(\theta) = Y - C_1(v_1(\theta)) - C_2(v_2(\theta)).$$

Then the planner's problem is to maximize M over the the set of all utility allocations $v \in \Omega$ such that $N(v) \geq 0$.

Remark 5 *We use the notation $\ell(d\theta)$ rather than $d\theta$ in the formula for M in order to be consistent with the notation for integration elsewhere in this appendix.*

Since N takes values in $\mathcal{BV}(\Theta, \mathbb{R})$, a Lagrange multiplier is a continuous linear functional on $\mathcal{BV}(\Theta, \mathbb{R})$. Denote the space of all continuous linear functionals on $\mathcal{BV}(\Theta, \mathbb{R})$ by $\mathcal{BV}(\Theta, \mathbb{R})^*$. Then the Lagrangian is the mapping

$$L : \Omega \times \mathcal{BV}(\Theta, \mathbb{R})^* \to \mathbb{R}$$

given by the formula

$$L(v; \lambda) = M(v) + \langle N(v), \lambda \rangle\,,$$

where $\langle N(v), \lambda \rangle$ denotes the real number obtained when the continuous linear functional $\lambda \in \mathcal{BV}(\Theta, \mathbb{R})^*$ is evaluated at the point $N(v) \in \mathcal{BV}(\Theta, \mathbb{R})$.

Remark 6 *Notice that both M and N are defined on open sets containing Ω, and not just on Ω itself.*

Remark 7 *M is well defined since v_1 and v_2 are well defined except at a countable number of points.*

Remark 8 *N is well defined since $u_t(0+) < \min v_t \leq \max v_t < u_t(\infty-)$ and C_t is continuously differentiable on $(u_t(0+), u_t(\infty-))$. Hence*

$$\|(C_t \circ v_t)'\|_{TV} \leq K \; \|v_t'\|_{TV} ,$$

where

$$K = \max\{C_t'(w) \mid w \in [\, \min v_t, \max v_t \,]\} .$$

Remark 9 *According the the Riesz representation theorem, the dual $\mathcal{C}(\Theta, \mathbb{R})^*$ of the space $\mathcal{C}(\Theta, \mathbb{R})$ of continuous functions on Θ can be represented by the space $\mathcal{M}(\Theta, \mathbb{R})$ of bounded Borel measures on Θ. Unfortunately, there does not seem to be a correspondingly tractable representation for the dual $\mathcal{BV}(\Theta, \mathbb{R})^*$ of the space $\mathcal{BV}(\Theta, \mathbb{R})$ of functions of bounded variation on Θ. This might be an obstacle to analyzing necessary conditions, where we would not have any control over the Lagrange multiplier. It is less of a problem when it comes to analyzing sufficiency conditions, where we are free to choose the Lagrange multiplier.*

B.6 A Space of Lagrange Multipliers

One can associate continuous linear functionals in $\mathcal{BV}(\Theta, \mathbb{R})^*$ with bounded Borel measures in $\mathcal{M}(\Theta, \mathbb{R})$ as follows. Suppose that we are given $\Lambda \in \mathcal{M}(\Theta, \mathbb{R})$. Then we can construct $\lambda_R \in \mathcal{BV}(\Theta, \mathbb{R})^*$ by means of the formula

$$\langle f, \lambda_R \rangle = \int f_R(\theta) \, \Lambda(d\theta),$$

where f_R is the right-continuous version of f. In this way we obtain a closed linear subspace of $\mathcal{BV}(\Theta, \mathbb{R})^*$. It turns out that this subspace is big enough for our purposes.

Remark 10 *By the same token, we can construct $\lambda_L \in \mathcal{BV}(\Theta, \mathbb{R})^*$ by means of the formula*

$$\langle f, \lambda_L \rangle = \int f_L(\theta) \, \Lambda(d\theta),$$

where f_L is the left-continuous version of f.

Remark 11 *Notice that $\lambda_L \neq \lambda_R$ and, while both λ_L and λ_R seem quite natural, neither seems to have a claim to being canonical.*

Remark 12 *We use the notation $\Lambda(d\theta)$ in the definition of $\langle f, \lambda_R \rangle$ and $\langle f, \lambda_L \rangle$ in order to emphasize that the integral in question is the Lebesgue integral of a measurable function with respect to the measure Λ. (The notation $d\Lambda(\theta)$ might be taken to suggest that the integral in question was the Riemann-Stieltjes integral of a continuous function with respect to the function of bounded variation Λ.)*

B.7 The Directional Derivative of the Lagrangian

Let us fix $\Lambda \in \mathcal{M}(\Theta, \mathbb{R})$ and consider $L(\,\cdot\,; \lambda_R)$. If our candidate allocation $r \in \Omega$ maximizes $L(\,\cdot\,; \lambda_R)$ then, for all $v \in \Omega$, the directional derivative $\nabla_s L(r; \lambda_R)$ of $L(\,\cdot\,; \lambda_R)$ at r in the direction $s = v - r$ must be non-positive. Conversely if, for all $v \in \Omega$, $\nabla_s L(r; \lambda_R)$ is non-positive, then $r \in \Omega$ maximizes $L(\,\cdot\,; \lambda_R)$. The purpose of the present section is to derive a formula for $\nabla_s L(r; \lambda_R)$. This formula will then be used to guide our eventual choice of Λ.

In view of our choice of λ_R, we have

$$L(v; \lambda_R) = \int (\theta\, v_{1R} + v_{2R})\, F'\, \ell(d\theta) + \int (Y - C_1(v_{1R}) - C_2(v_{2R}))\, \Lambda(d\theta).$$

Hence

$$\nabla_s L(r; \lambda_R) = \int (\theta\, s_{1R} + s_{2R})\, F'\, \ell(d\theta) - \int (C_1'(r_{1R})\, s_{1R} + C_2'(r_{2R})\, s_{2R})\, \Lambda(d\theta).$$

Now, because F is continuous, the standard formula for integration by parts shows that

$$\int s_{2R}\, F'\, \ell(d\theta) = [\, s_2\, F\,]_{\underline{\theta}-}^{\overline{\theta}+} - \int F\, s_2'(d\theta),$$

where:

- $[\, s_2\, F\,]_{\underline{\theta}-}^{\overline{\theta}+}$ denotes the difference between the right-hand limit of $s_2\, F$ at $\overline{\theta}$ and the left-hand limit of $s_2\, F$ at $\underline{\theta}$;
- $\int F\, s_2'(d\theta)$ denotes the integral of F with respect to the measure s_2'.

Furthermore, it follows from incentive compatibility that $\theta\, s_1' + \beta\, s_2' = 0$. Hence

$$\int F\, s_2'(d\theta) = -\int F\, \tfrac{\theta}{\beta}\, s_1'(d\theta) = -\tfrac{1}{\beta}\, [\, s_1\, (\theta\, F)\,]_{\underline{\theta}-}^{\overline{\theta}+} + \tfrac{1}{\beta} \int s_{1R}\, (\theta\, F)'\, \ell(d\theta),$$

(integrating by parts again, and using the fact that F is continuous). Hence the first integral in the directional derivative

$$\begin{aligned}\int (\theta\, s_{1R} + s_{2R})\, F'\, \ell(d\theta) &= \int \theta\, s_{1R}\, F'\, \ell(d\theta) + \int s_{2R}\, F'\, \ell(d\theta)\\ &= \int \theta\, s_{1R}\, F'\, \ell(d\theta) + [\, s_2\, F\,]_{\underline{\theta}-}^{\overline{\theta}+} + \tfrac{1}{\beta}\, [\, s_1\, (\theta\, F)\,]_{\underline{\theta}-}^{\overline{\theta}+} - \tfrac{1}{\beta} \int s_{1R}\, (\theta\, F)'\, \ell(d\theta)\\ &= \left(\tfrac{\overline{\theta}}{\beta}\, s_{1R}(\,\overline{\theta}\,) + s_{2R}(\,\overline{\theta}\,)\right) F(\,\overline{\theta}\,) - \tfrac{1}{\beta} \int s_{1R}\, ((1-\beta)\, \theta\, F' + F)\, \ell(d\theta)\end{aligned}$$

(where we have used the fact that $F(\,\underline{\theta}\,) = 0$).

Next, G be the distribution function of the measure $C_2'(r_{2R})\, \Lambda$. I.e. let G be the unique element of $\mathcal{BV}(\Theta, \mathbb{R})$ such that $G' = C_2'(r_{2R})\, \Lambda$ and $G_L(\,\underline{\theta}\,) = 0$. Then

$$\begin{aligned}\int C_2'(r_{2R})\, s_{2R}\, \Lambda(d\theta) &= \int s_{2R}\, G'(d\theta)\\ &= [\, s_2\, G\,]_{\underline{\theta}-}^{\overline{\theta}+} - \int G\, s_2'(d\theta) + \sum_{\theta \in [\,\underline{\theta}, \overline{\theta}\,]} \Delta s_2\, \Delta G,\end{aligned}$$

where Δs_2 and ΔG denote the jumps in s_2 and G at θ (if any). Furthermore, it follows from incentive compatibility that $\theta\, s_1' + \beta\, s_2' = 0$. In particular, $\theta\, \Delta s_1 + \beta\, \Delta s_2 = 0$. Hence

$$\begin{aligned}\int G\, s_2'(d\theta) &= -\int G\, \tfrac{\theta}{\beta}\, s_1'(d\theta)\\ &= -\tfrac{1}{\beta}\, [\, s_1\, (\theta\, G)\,]_{\underline{\theta}-}^{\overline{\theta}+} + \tfrac{1}{\beta} \int s_{1R}\, (\theta\, G)'\, (d\theta) - \tfrac{1}{\beta} \sum_{\theta \in [\,\underline{\theta}, \overline{\theta}\,]} \Delta s_1\, \Delta(\theta\, G)\\ &= -\tfrac{1}{\beta}\, [\, s_1\, (\theta\, G)\,]_{\underline{\theta}-}^{\overline{\theta}+} + \tfrac{1}{\beta} \int s_{1R}\, (\theta\, G)'\, (d\theta) - \tfrac{1}{\beta} \sum_{\theta \in [\,\underline{\theta}, \overline{\theta}\,]} \Delta s_1\, \theta\, \Delta G\end{aligned}$$

(integrating by parts again and using the fact that $\Delta(\theta\, G) = \theta\, \Delta(G)$), and

$$\sum_{\theta \in [\,\underline{\theta}, \overline{\theta}\,]} \Delta s_2\, \Delta G = -\tfrac{1}{\beta} \sum_{\theta \in [\,\underline{\theta}, \overline{\theta}\,]} \theta\, \Delta s_1\, \Delta G.$$

Overall,

$$\int C_1'(r_{1R})\, s_{1R}\, \Lambda(d\theta) = \int \frac{C_1'(r_{1R})}{C_2'(r_{2R})}\, s_{1R}\, G'(d\theta)$$

and

$$\begin{aligned}
\int C_2'(r_{2R})\, s_{2R}\, \Lambda(d\theta) &= [\, s_2\, G\,]_{\underline{\theta}-}^{\overline{\theta}+} + \tfrac{1}{\beta}\, [\, s_1\, (\theta\, G)\,]_{\underline{\theta}-}^{\overline{\theta}+} - \tfrac{1}{\beta} \int s_{1R}\ (\theta\, G)'\, (d\theta) \\
&= s_{2R}(\,\overline{\theta}\,)\, G_R(\,\overline{\theta}\,) + \tfrac{1}{\beta}\, s_{1R}(\,\overline{\theta}\,)\, \overline{\theta}\, G_R(\,\overline{\theta}\,) - \tfrac{1}{\beta} \int s_{1R}\, (\theta\, G'(d\theta) + G\, \ell(d\theta)) \\
&= \left(\tfrac{\overline{\theta}}{\beta}\, s_{1R}(\,\overline{\theta}\,) + s_{2R}(\,\overline{\theta}\,)\right) G_R(\,\overline{\theta}\,) - \tfrac{1}{\beta} \int s_{1R}\, (\theta\, G'(d\theta) + G\, \ell(d\theta))
\end{aligned}$$

(where we have used the facts that $G_L(\,\underline{\theta}\,) = 0$ and $(\theta\, G)'\, (d\theta) = \theta\, G'(d\theta) + G\, \ell(d\theta)$).

Finally, putting all of this information together, we have

$$\begin{aligned}
\nabla_s L(r; \lambda_R) &= \left(\tfrac{\overline{\theta}}{\beta}\, s_{1R}(\,\overline{\theta}\,) +\ s_{2R}(\,\overline{\theta}\,)\right) F(\,\overline{\theta}\,) - \tfrac{1}{\beta} \int s_{1R}\, ((1-\beta)\, \theta\, F' + F)\, \ell(d\theta) \\
&\quad - \int \frac{C_1'(r_{1R})}{C_2'(r_{2R})}\, s_{1R}\, G'(d\theta) \\
&\quad - \left(\tfrac{\overline{\theta}}{\beta}\, s_{1R}(\,\overline{\theta}\,) + s_{2R}(\,\overline{\theta}\,)\right) G_R(\,\overline{\theta}\,) + \tfrac{1}{\beta} \int s_{1R}\, (\theta\, G'(d\theta) + G\, \ell(d\theta)) \\
&= \left(\tfrac{\overline{\theta}}{\beta}\, s_{1R}(\,\overline{\theta}\,) + s_{2R}(\,\overline{\theta}\,)\right) \left(F(\,\overline{\theta}\,) - G_R(\,\overline{\theta}\,)\right) \\
&\quad + \tfrac{1}{\beta} \int s_{1R}\, (G - (1-\beta)\, \theta\, F' - F)\, \ell(d\theta) \\
&\quad + \tfrac{1}{\beta} \int s_{1R} \left(\theta - \beta\, \frac{C_1'(r_{1R})}{C_2'(r_{2R})}\right) G'(d\theta).
\end{aligned}$$

B.8 A Candidate Lagrange Multiplier

We are now in a position to motivate our choice of Lagrange multiplier Λ. We shall do this in two steps. First, we motivate our choice of G. Second, we show how to translate our choice of G into a choice of Λ.

In choosing G, the broad aim is to ensure that $\nabla_s L(r; \lambda_R) \le 0$. However, given that we have only limited control over s, it will be helpful to make as many of the terms in the formula for $\nabla_s L(r; \lambda_R)$ vanish as possible.

Recall that

$$\begin{aligned}
\nabla_s L(r; \lambda_R) &= \left(\tfrac{\overline{\theta}}{\beta}\, s_{1R}(\,\overline{\theta}\,) + s_{2R}(\,\overline{\theta}\,)\right) \left(F(\,\overline{\theta}\,) - G_R(\,\overline{\theta}\,)\right) \\
&\quad + \tfrac{1}{\beta} \int s_{1R}\, (G - \Gamma)\, \ell(d\theta) \\
&\quad + \tfrac{1}{\beta} \int s_{1R} \left(\theta - \beta\, \frac{C_1'(r_{1R})}{C_2'(r_{2R})}\right) G'(d\theta),
\end{aligned}$$

where $\Gamma = (1-\beta)\,\theta\,F' - F$. We can therefore make a start by requiring that

$$G_R(\,\overline{\theta}\,) = F(\,\overline{\theta}\,).$$

This will ensure that the first term vanishes.

Remark 13 *At this point we have specified both $G_L(\,\underline{\theta}\,)$ and $G_R(\,\overline{\theta}\,)$. It remains to specify G in the interior of Θ.*

Next, suppose that $\theta_1 > \underline{\theta}$. Then

$$\frac{C_1'(r_{1R})}{C_2'(r_{2R})} = \left\{ \begin{array}{ll} \frac{\theta}{\beta} & \text{for } \theta \in \Theta_S = [\,\underline{\theta}, \theta_1\,) \\ \frac{\theta_1}{\beta} & \text{for } \theta \in \Theta_P = [\,\theta_1, \overline{\theta}\,] \end{array} \right\}.$$

Hence the expression for $\nabla_s L(r;\lambda_R)$ simplifies to

$$\tfrac{1}{\beta}\int_{\Theta_S\cup\Theta_P} s_{1R}\,(G-\Gamma)\,\ell(d\theta) + \tfrac{1}{\beta}\int_{\Theta_P} s_{1R}\,(\theta-\theta_1)\,G'(d\theta).$$

Suppose further that we follow the suggestion of AWA (2006), and put $G = \Gamma$ on Θ_S, where Γ is the function defined in Section B.1 above. Then the contribution to $\nabla_s L(r;\lambda_R)$ from the separating interval Θ_S vanishes altogether, and all that is left is the contribution

$$\tfrac{1}{\beta}\int_{\Theta_P} s_{1R}\,(G-\Gamma)\,\ell(d\theta) + \tfrac{1}{\beta}\int_{\Theta_P} s_{1R}\,(\theta-\theta_1)\,G'(d\theta)$$

to $\nabla_s L(r;\lambda_R)$ from the pooling interval Θ_P. Suppose finally that we follow the suggestion of AWA (2006), and put $G = F(\,\overline{\theta}\,)$ on $(\,\theta_1, \overline{\theta}\,)$. Then the measure G' will have an atom of size $F(\,\overline{\theta}\,) - \Gamma_L(\theta_1)$ at θ_1, and it will vanish on $(\,\theta_1, \overline{\theta}\,]$. Since the term $\theta - \theta_1$ multiplying $G'(d\theta)$ vanishes at θ_1, the second integral itself vanishes, and the first integral reduces to

$$\tfrac{1}{\beta}\int_{\Theta_P} s_{1R}\,\big(F(\,\overline{\theta}\,) - \Gamma\big)\,\ell(d\theta).$$

Next, suppose that $\theta_1 \le \underline{\theta}$. In this case, the expression for $\nabla_s L(r;\lambda_R)$ simplifies to

$$\tfrac{1}{\beta}\int s_{1R}\,(G-\Gamma)\,\ell(d\theta) + \tfrac{1}{\beta}\int s_{1R}\,(\theta-\theta_1)\,G'(d\theta).$$

Suppose further that we follow the suggestion of AWA (2006), and put $G = F(\,\overline{\theta}\,)$ on the whole of $(\,\underline{\theta}, \overline{\theta}\,)$. Then the measure G' will have an atom of size $F(\,\overline{\theta}\,)$ at $\underline{\theta}$, and it will vanish on $(\,\underline{\theta},\overline{\theta}\,]$. Hence the expression for $\nabla_s L(r;\lambda_R)$ becomes

$$\tfrac{1}{\beta}\int s_{1R}\left(F(\,\overline{\theta}\,)-\Gamma\right)\ell(d\theta)+\tfrac{1}{\beta}\,s_{1R}(\,\underline{\theta}\,)\,(\,\underline{\theta}-\theta_1)\,F(\,\overline{\theta}\,).$$

In other words, compared with the case $\theta_1>\underline{\theta}$, there is an extra term arising from the atom of G' at $\underline{\theta}$.

Finally, we obtain the Lagrange multiplier Λ itself from the formula

$$\Lambda=\frac{1}{C_2'(r_{2R})}\,G'.$$

B.9 Non-Negativity of the Lagrange Multiplier

Since $C_2'(r_{2R})>0$, $\Lambda\geq 0$ iff $G'\geq 0$. We will show that $G'\geq 0$. Suppose first that $\theta_1>\underline{\theta}$. Then we have

$$\begin{aligned} G_L(\,\underline{\theta}\,) &= 0 \\ G &= \Gamma \text{ on } (\,\underline{\theta},\theta_1\,) \\ G &= F(\,\overline{\theta}\,) \text{ on } (\,\theta_1,\overline{\theta}\,) \\ G_R(\,\overline{\theta}\,) &= F(\,\overline{\theta}\,) \end{aligned}$$

Now, it follows from the formula for Γ that

$$G_R(\,\underline{\theta}\,)=\Gamma_R(\,\underline{\theta}\,)=(1-\beta)\,\theta\,F_R'(\,\underline{\theta}\,)+F(\,\underline{\theta}\,)=(1-\beta)\,\underline{\theta}\,F_R'(\,\underline{\theta}\,)\geq 0.$$

And $G_L(\,\underline{\theta}\,)=0$ by construction. Hence

$$\Delta G(\,\underline{\theta}\,)=G_R(\,\underline{\theta}\,)-G_L(\,\underline{\theta}\,)\geq 0.$$

Next, it follows from Assumption A4 that Γ is non-decreasing on $(\,\underline{\theta},\theta_1)$. Hence $G'=\Gamma'\geq 0$ there. Third, we have

$$\Delta G(\,\theta_1\,)=F(\,\overline{\theta}\,)-\Gamma_L(\,\theta_1\,).$$

But if it were the case that $\Gamma_L(\,\theta_1\,)>F(\,\overline{\theta}\,)$ then there would be an open interval $(\,\theta_1-\varepsilon,\theta_1\,)$ on which $\Gamma>F(\,\overline{\theta}\,)$. This would contradict the definition of θ_1 as the minimum $\theta\in(\,0,\overline{\theta}\,)$ such that $\frac{1}{\overline{\theta}-t}\int_t^{\overline{\theta}}\Gamma(s)ds\geq F(\,\overline{\theta}\,)$ for all $t\in[\,\theta,\overline{\theta}\,)$. Hence $\Gamma_L(\,\theta_1\,)\leq F(\,\overline{\theta}\,)$ and $\Delta G(\,\theta_1\,)\geq 0$. Fourth, we have $G'=0$ on $(\,\theta_1,\overline{\theta}\,)$. Finally, we obviously have $\Delta G(\,\overline{\theta}\,)=0$.

Suppose now that $\theta_1 \leq \underline{\theta}$. Then we have

$$\begin{aligned} G_L(\underline{\theta}\,) &= 0 \\ G &= F(\,\overline{\theta}\,) \text{ on } (\,\underline{\theta}, \overline{\theta}\,) \\ G_R(\,\overline{\theta}\,) &= F(\,\overline{\theta}\,). \end{aligned}$$

So it is obvious that $G' \geq 0$ on the whole of $[\,\underline{\theta}, \overline{\theta}\,]$.

B.10 Non-Positivity of the Directional Derivative

Suppose that $\theta_1 > \underline{\theta}$. Then, in the light of the discussion in Section B.8, we have

$$\nabla_s L(r; \lambda_R) = \tfrac{1}{\beta} \int_{\Theta_P} s_{1R} \left(F(\,\overline{\theta}\,) - \Gamma \right) \ell(d\theta).$$

Define $H : (0, \infty) \to \mathbb{R}$ by the formula

$$H(\theta) = \int_{\theta}^{\overline{\theta}} \left(\Gamma - F(\,\overline{\theta}\,) \right) \ell(d\theta).$$

Then

$$\begin{aligned} \int_{\Theta_P} s_{1R} \left(F(\,\overline{\theta}\,) - \Gamma \right) \ell(d\theta) &= \int_{[\,\theta_1, \overline{\theta}\,]} s_{1R} \left(F(\,\overline{\theta}\,) - \Gamma \right) \ell(d\theta) \\ &= \int_{[\,\theta_1, \overline{\theta}\,]} s_{1R}\, H'\, \ell(d\theta) \\ &= [\, s_1\, H\,]_{\theta_1 -}^{\overline{\theta}+} - \int_{[\,\theta_1, \overline{\theta}\,]} H\, s_1'(d\theta) \end{aligned}$$

(integrating by parts and using the fact that H is continuous). Moreover

$$[\, s_1\, H\,]_{\theta_1 -}^{\overline{\theta}+} = s_{1R}(\,\overline{\theta}\,)\, H(\,\overline{\theta}\,) - s_{1L}(\,\theta_1\,)\, H(\,\theta_1\,)$$

and

$$\int_{[\,\theta_1, \overline{\theta}\,]} H\, s_1'(d\theta) = H(\,\theta_1\,)\, \Delta s_1(\,\theta_1\,) + \int_{(\,\theta_1, \overline{\theta}\,)} H\, s_1'(d\theta) + H(\,\overline{\theta}\,)\, \Delta s_1(\,\overline{\theta}\,).$$

Hence, overall,

$$\begin{aligned}\nabla_s L(r;\lambda_R) &= -H(\,\theta_1\,)\, s_{1R}(\,\theta_1\,) - \int_{(\,\theta_1,\overline{\theta}\,)} H\, s_1'(d\theta) + H(\,\overline{\theta}\,)\, s_{1L}(\,\overline{\theta}\,) \\ &= -\int_{(\,\theta_1,\overline{\theta}\,)} H\, s_1'(d\theta)\end{aligned}$$

(since $H(\,\overline{\theta}\,) = 0$ by construction and $H(\,\theta_1\,) = 0$ by definition of θ_1). Now $v_1' \geq 0$ on the whole of Θ, since v_1 is non-decreasing, and $r_1' = 0$ on $(\,\theta_1, \overline{\theta}\,)$, since r_1 is constant there. Hence $s_1' \geq 0$ on $(\,\theta_1, \overline{\theta}\,)$. On the other hand, for all $\theta \in [\,\theta_1, \overline{\theta}\,]$, we have

$$H(\theta) = \int_\theta^{\overline{\theta}} \big(\Gamma - F(\,\overline{\theta}\,)\big)\, \ell(d\theta) = (\,\overline{\theta} - \theta\,) \left(\tfrac{1}{\overline{\theta}-\theta} \int_\theta^{\overline{\theta}} \Gamma\, \ell(d\theta) - F(\,\overline{\theta}\,) \right) \geq 0,$$

by definition of θ_1. Hence $\nabla_s L(r;\lambda_R) \leq 0$, as required.

Remark 14 *Notice that s_1 is the difference of the two non-decreasing functions v_1 and r_1. Hence there is no general reason why s_1 should be non-decreasing. The situation is saved by the fact that r_1 is constant on $(\,\theta_1, \overline{\theta}\,)$.*

Suppose now that $\theta_1 \leq \underline{\theta}$. Then, in the light of the discussion in Section B.8, we have

$$\nabla_s L(r;\lambda_R) = \tfrac{1}{\beta}\, s_{1R}(\,\underline{\theta}\,)\,(\,\underline{\theta} - \theta_1\,)\, F(\,\overline{\theta}\,) + \tfrac{1}{\beta} \int s_{1R} \big(F(\,\overline{\theta}\,) - \Gamma\big)\, \ell(d\theta).$$

Now, arguing as in the case $\theta_1 > \underline{\theta}$, we have

$$\begin{aligned}\int s_{1R} \big(F(\,\overline{\theta}\,) - \Gamma\big)\, \ell(d\theta) &= \int_{[\,\underline{\theta},\overline{\theta}\,]} s_{1R} \big(F(\,\overline{\theta}\,) - \Gamma\big)\, \ell(d\theta) \\ &= \int_{[\,\underline{\theta},\overline{\theta}\,]} s_{1R}\, H'\, \ell(d\theta) \\ &= [\, s_1\, H\,]_{\underline{\theta}-}^{\overline{\theta}+} - \int_{[\,\underline{\theta},\overline{\theta}\,]} H\, s_1'(d\theta) \\ &= -H(\,\underline{\theta}\,)\, s_{1R}(\,\underline{\theta}\,) - \int_{(\,\underline{\theta},\overline{\theta}\,)} H\, s_1'(d\theta) + H(\,\overline{\theta}\,)\, s_{1L}(\,\overline{\theta}\,) \\ &= -H(\,\underline{\theta}\,)\, s_{1R}(\,\underline{\theta}\,) - \int_{(\,\underline{\theta},\overline{\theta}\,)} H\, s_1'(d\theta)\end{aligned}$$

(since $H(\,\overline{\theta}\,) = 0$ by construction). Hence, overall, we have

$$\beta\, \nabla_s L(r;\lambda_R) = \big((\,\underline{\theta} - \theta_1\,)\, F(\,\overline{\theta}\,) - H(\,\underline{\theta}\,)\big)\, s_{1R}(\,\underline{\theta}\,) - \int_{(\,\underline{\theta},\overline{\theta}\,)} H\, s_1'(d\theta).$$

But

$$(\,\underline{\theta}-\theta_1\,)\,F(\,\overline{\theta}\,)-H(\,\underline{\theta}\,)=\int_{\theta_1}^{\underline{\theta}}F(\,\overline{\theta}\,)\,\ell(d\theta)-\int_{\underline{\theta}}^{\overline{\theta}}(\Gamma-F(\,\overline{\theta}\,))\,\ell(d\theta)$$

(by definition of H)

$$=-\int_{\theta_1}^{\underline{\theta}}\big(\Gamma-F(\,\overline{\theta}\,)\big)\,\ell(d\theta)-\int_{\underline{\theta}}^{\overline{\theta}}\big(\Gamma-F(\,\overline{\theta}\,)\big)\,\ell(d\theta)$$

(since $\Gamma=0$ on $[\,\theta_1,\underline{\theta}\,))$

$$=-H(\,\theta_1\,)$$

(by definition of H again)

$$=0.$$

(by definition of θ_1). Hence

$$\beta\,\nabla_s L(r;\lambda_R)=-\int_{(\,\underline{\theta},\overline{\theta}\,)}H\,s_1'(d\theta).$$

Hence, arguing as in the case $\theta_1>\underline{\theta}$, $\nabla_s L(r;\lambda_R)\le 0$.

C Proof of Proposition 15

We now study the case in which the government can make inter-household transfers. Specifically, we now replace *household-by-household* budget balance (Equation 7) with *overall* budget balance (Equation 6). With overall budget balance, we will show that a combination of a perfectly liquid and a perfectly illiquid account is not sufficient to maximize social surplus. We continue to make assumptions A1-A5. To these assumptions we add:

A6 F' is bounded away from 0 on $(\,\underline{\theta},\overline{\theta}\,)$.[54]

Proposition 15 *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 possible (in the interior case, with partial separation), we can use an incentive compatible mechanism to redistribute c_1 away from low$-\theta$ types (i.e., households with low marginal utility, ceteris paribus).

[54] In particular, both the right-hand limit $F_R'(\,\underline{\theta}\,)$ of F' at $\underline{\theta}$ and the left-hand limit $F_L'(\,\overline{\theta}\,)$ of F' at $\overline{\theta}$ are strictly positive.

C.1 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 $\theta_2 \in (0, \infty)$ such that: if $\theta < \theta_2$, then she consumes less than the balance x_{liquid} in her liquid account: and, if $\theta \geq \theta_2$, then she consumes the whole of x_{liquid}. The cutoff θ_2 need not lie in $[\underline{\theta}, \overline{\theta}]$. It could be that $\theta_2 < \underline{\theta}$, 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 $\theta_2 > \overline{\theta}$, 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 $c_1, c_2 : \Theta \to (0, \infty)$ and associated utility allocations $r_1, r_2 : \Theta \to \mathbb{R}$, 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 $r = (r_1, r_2)$ will be a smooth function of θ for $\theta < \theta_2$, have a kink at θ_2, and be constant for $\theta > \theta_2$. 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 $v_1, v_2 : \Theta \to \mathbb{R}$, 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

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

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

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

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

$$\theta\, v_1' + \beta\, v_2' = 0 \tag{ICL}$$

and a monotonic part

$$v_2' \le 0. \tag{ICM}$$

Remark 16 *The two conditions (ICL) and (ICM) are simply the differential counterpart of the usual integral representation of incentive compatibility in a mechanism-design problem.*

C.2 The Case $\theta_2 \in (\underline{\theta}, \overline{\theta})$

Consider first the case in which x_{liquid} and x_{illiquid} are such that $\theta_2 \in (\underline{\theta}, \overline{\theta})$. In this case, the second step is to parameterize candidate solutions $v = (v_1, v_2)$ to the planner's problem in terms of boundary values $v_1(\overline{\theta})$, $v_2(\overline{\theta})$ and continuous functions $v_{1L}' : [\underline{\theta}, \theta_2] \to \mathbb{R}$, $v_{1R}' : [\theta_2, \overline{\theta}] \to \mathbb{R}$. More precisely, we can put:

1. $v_1(\theta) = v_1(\overline{\theta}) - \int_\theta^{\overline{\theta}} v_{1R}'(t)\, dt$ for $\theta \in [\theta_2, \overline{\theta}]$;
2. $v_1(\theta) = v_1(\theta_2) - \int_\theta^{\theta_2} v_{1L}'(t)\, dt$ for $\theta \in [\underline{\theta}, \theta_2]$;
3. $v_{2R}'(\theta) = -\frac{\theta}{\beta} v_{1R}'(\theta)$ for $\theta \in [\theta_2, \overline{\theta}]$;
4. $v_{2L}'(\theta) = -\frac{\theta}{\beta} v_{1L}'(\theta)$ for $\theta \in [\underline{\theta}, \theta_2]$;
5. $v_2(\theta) = v_2(\overline{\theta}) - \int_\theta^{\overline{\theta}} v_{2R}'(t)\, dt$ for $\theta \in [\theta_2, \overline{\theta}]$;
6. $v_2(\theta) = v_2(\theta_2) - \int_\theta^{\theta_2} v_{2L}'(t)\, dt$ for $\theta \in [\underline{\theta}, \theta_2]$.

In other words: v_1 is the continuous function with continuous derivative v_{1L}' on $[\underline{\theta}, \theta_2)$, continuous derivative v_{1R}' on $(\theta_2, \overline{\theta}]$ and value $v_1(\overline{\theta})$ at $\overline{\theta}$; and v_2 is the continuous function with continuous derivative v_{2L}' on $[\underline{\theta}, \theta_2)$, continuous derivative v_{2R}' on $(\theta_2, \overline{\theta}]$ and value $v_2(\overline{\theta})$ at $\overline{\theta}$.

Remark 17 *Notice that the two-account system described in Proposition 15 gives rise to a utility allocation $r = (r_1, r_2)$ satisfying conditions 1-6. Moreover – as we shall see below – in order to show that r is not optimal, it suffices to consider variations in this same class. We simply do not need to consider variations in which (say) θ_2 changes or $v = (v_1, v_2)$ can be discontinuous.*

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

$$\begin{aligned} L(v_1(\overline{\theta}), v_2(\overline{\theta}), v'_{1L}, v'_{1R}, \lambda, \zeta_L, \zeta_R) &= \int (\theta\, v_1(\theta) + v_2(\theta))\, dF(\theta) \\ &\quad + \lambda \int (Y - C_1(v_1(\theta)) - C_2(v_2(\theta)))\, dF(\theta) \\ &\quad - \int_{[\underline{\theta}, \theta_2]} v'_{2L}(\theta)\, d\zeta_L(\theta) \\ &\quad - \int_{[\theta_2, \overline{\theta}]} v'_{2R}(\theta)\, d\zeta_R(\theta), \end{aligned} \tag{11}$$

where:

1. the arguments of L are the parameters $v_1(\overline{\theta})$, $v_2(\overline{\theta})$, 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 $[\underline{\theta}, \theta_2]$ (namely the multiplier associated with the non-positivity constraint on v'_{2L});

4. ζ_R is a finite non-negative Borel measure on $[\theta_2, \overline{\theta}]$ (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 $v_1(\overline{\theta})$, $v_2(\overline{\theta})$, v'_{1L} and v'_{1R} as explained above.

Remark 18 *The Langrangian 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 $(r_1(\overline{\theta}), r_2(\overline{\theta}), r'_{1L}, r'_{1R})$ with the reference utility allocation (r_1, r_2) and parameters $(v_1(\overline{\theta}), v_2(\overline{\theta}), v'_{1L}, v'_{1R})$ with the alternative utility allocation (v_1, v_2) in the obvious way, and take the derivative of the Langrangian at the parameter values $(r_1(\overline{\theta}), r_2(\overline{\theta}), r'_{1L}, r'_{1R})$ in the direction $(s_1(\overline{\theta}), s_2(\overline{\theta}), s'_{1L}, s'_{1R})$, where $s = v - r$. 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 Langrangian are linear in the underlying parameters $(v_1(\overline{\theta}), v_2(\overline{\theta}), v'_{1L}, v'_{1R})$. Hence we can simply take the derivative of the RHS at the point $(r_1, r_2, r'_{2L}, r'_{2R})$ in the direction $(s_1, s_2, s'_{2L}, s'_{2R})$ and only then substitute for $(s_1, s_2, s'_{2L}, s'_{2R})$ in terms of $(s_1(\overline{\theta}), s_2(\overline{\theta}), s'_{1L}, s'_{1R})$.

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

$$\begin{aligned} 0 &= \int (\theta\, s_1 + s_2)\, dF - \lambda \int (C'_1(r_1)\, s_1 + C'_2(r_2)\, s_2)\, dF \\ &\quad - \int_{[\underline{\theta}, \theta_2]} s'_{2L}(\theta)\, d\zeta_L(\theta) - \int_{[\theta_2, \overline{\theta}]} s'_{2R}(\theta)\, d\zeta_R(\theta) \end{aligned} \tag{12}$$

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,

$$0 = \int_{[\underline{\theta}, \theta_2]} r'_{2L}(\theta)\, d\zeta_L(\theta), \tag{13}$$

$$0 = \int_{[\theta_2, \overline{\theta}]} r'_{2R}(\theta)\, d\zeta_R(\theta). \tag{14}$$

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

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

$$\begin{aligned} \int \theta\, s_1\, dF(\theta) &= \big[(\theta\, F - \overline{F}\,)\, s_1\big]_{\underline{\theta}-}^{\overline{\theta}} - \int (\theta\, F - \overline{F}\,)\, s'_1\, d\theta \\ &= \big(\overline{\theta}\, F(\overline{\theta}\,) - \overline{F}(\overline{\theta}\,)\big)\, s_1(\overline{\theta}\,) - \int (\theta\, F - \overline{F}\,)\, s'_1\, d\theta \\ &= \big(\overline{\theta}\, F(\overline{\theta}\,) - \overline{F}(\overline{\theta}\,)\big)\, s_1(\overline{\theta}\,) \\ &\quad - \int_{[\underline{\theta}, \theta_2]} (\theta\, F - \overline{F}\,)\, s'_{1L}\, d\theta - \int_{[\theta_2, \overline{\theta}]} (\theta\, F - \overline{F}\,)\, s'_{1R}\, d\theta, \end{aligned}$$

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

$$\begin{aligned}
\int s_2\, dF(\theta) &= [F\, s_2]_{\underline{\theta}-}^{\overline{\theta}} - \int F\, s_2'\, d\theta \\
&= F(\,\overline{\theta}\,)\, s_2(\,\overline{\theta}\,) - \int F\, s_2'\, d\theta \\
&= F(\,\overline{\theta}\,)\, s_2(\,\overline{\theta}\,) - \int_{[\,\underline{\theta},\theta_2\,]} F\, s_{2L}'\, d\theta - \int_{\left[\,\theta_2,\overline{\theta}\,\right]} F\, s_{2R}'\, d\theta \\
&= F(\,\overline{\theta}\,)\, s_2(\,\overline{\theta}\,) + \int_{[\,\underline{\theta},\theta_2\,]} F\, \tfrac{\theta}{\beta}\, s_{1L}'\, d\theta + \int_{\left[\,\theta_2,\overline{\theta}\,\right]} F\, \tfrac{\theta}{\beta}\, s_{1R}'\, d\theta.
\end{aligned}$$

Next, putting $\Lambda_1(\theta) = \int_{[\underline{\theta},\theta]} C_1'(r_1(t))\, dF(t)$, we have

$$\begin{aligned}
-\lambda \int C_1'(r_1)\, s_1\, dF &= -\int s_1\, \lambda\, \Lambda_1'\, d\theta \\
&= -\left[s_1\, \lambda\, \Lambda_1\right]_{\underline{\theta}-}^{\overline{\theta}} + \int \lambda\, \Lambda_1\, s_1'\, d\theta \\
&= -s_1(\,\overline{\theta}\,)\, \lambda\, \Lambda_1(\,\overline{\theta}\,) + \int \lambda\, \Lambda_1\, s_1'\, d\theta \\
&= -s_1(\,\overline{\theta}\,)\, \lambda\, \Lambda_1(\,\overline{\theta}\,) \\
&\qquad + \int_{[\,\underline{\theta},\theta_2\,]} \lambda\, \Lambda_1\, s_{1L}'\, d\theta + \int_{\left[\,\theta_2,\overline{\theta}\,\right]} \lambda\, \Lambda_1\, s_{1R}'\, d\theta.
\end{aligned}$$

Similarly, putting $\Lambda_2(\theta) = \int_{[\underline{\theta},\theta]} C_2'(r_2(t))\, dF(t)$,

$$\begin{aligned}
-\lambda \int C_2'(r_2)\, s_2\, dF &= -\int s_2\, \lambda\, \Lambda_2'\, d\theta \\
&= -\left[s_2\, \lambda\, \Lambda_2\right]_{\underline{\theta}-}^{\overline{\theta}} + \int \lambda\, \Lambda_2\, s_2'\, d\theta \\
&= -s_2(\,\overline{\theta}\,)\, \lambda\, \Lambda_2(\,\overline{\theta}\,) + \int \lambda\, \Lambda_2\, s_2'\, d\theta \\
&= -s_2(\,\overline{\theta}\,)\, \lambda\, \Lambda_2(\,\overline{\theta}\,) \\
&\qquad + \int_{[\,\underline{\theta},\theta_2\,]} \lambda\, \Lambda_2\, s_{2L}'\, d\theta + \int_{\left[\,\theta_2,\overline{\theta}\,\right]} \lambda\, \Lambda_2\, s_{2R}'\, d\theta \\
&= -s_2(\,\overline{\theta}\,)\, \lambda\, \Lambda_2(\,\overline{\theta}\,) \\
&\qquad - \int_{[\,\underline{\theta},\theta_2\,]} \lambda\, \Lambda_2\, \tfrac{\theta}{\beta}\, s_{1L}'\, d\theta - \int_{\left[\,\theta_2,\overline{\theta}\,\right]} \lambda\, \Lambda_2\, \tfrac{\theta}{\beta}\, s_{1R}'\, d\theta.
\end{aligned}$$

Finally, we have

$$-\int_{[\underline{\theta},\theta_2]} s_{2L}'(\theta)\, d\zeta_L(\theta) = \int_{[\underline{\theta},\theta_2]} \tfrac{\theta}{\beta}\, s_{1L}'(\theta)\, d\zeta_L(\theta)$$

and

$$-\int_{[\theta_2,\overline{\theta}]} s'_{2R}(\theta)\, d\zeta_R(\theta) = \int_{[\theta_2,\overline{\theta}]} \tfrac{\theta}{\beta}\, s'_{1R}(\theta)\, d\zeta_R(\theta).$$

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

$$\begin{aligned} 0 &= \overline{\theta}\, F(\overline{\theta}) - \overline{F}(\overline{\theta}) - \lambda\, \Lambda_1(\overline{\theta}), && (15)\\ 0 &= F(\overline{\theta}) - \lambda\, \Lambda_2(\overline{\theta}), && (16)\\ 0 &= -(\theta\, F - \overline{F})\, d\theta + \tfrac{\theta}{\beta}\, F\, d\theta + \lambda\, \Lambda_1\, d\theta - \tfrac{\theta}{\beta}\, \lambda\, \Lambda_2\, d\theta + \tfrac{\theta}{\beta}\, d\zeta_L, && (17)\\ 0 &= -(\theta\, F - \overline{F})\, d\theta + \tfrac{\theta}{\beta}\, F\, d\theta + \lambda\, \Lambda_1\, d\theta - \tfrac{\theta}{\beta}\, \lambda\, \Lambda_2\, d\theta + \tfrac{\theta}{\beta}\, d\zeta_R. && (18) \end{aligned}$$

Now, we certainly have $r'_{2L} < 0$ on $[\underline{\theta}, \theta_2]$. (This is because, if $\theta < \theta_2$, then self 1 consumes less than x_{liquid}. Hence $r'_{1L} > 0$ and $r'_{2L} < 0$.) It therefore follows from constraint qualification (namely (13)) that $\zeta_L = 0$. Equation (17) therefore implies that

$$\lambda\, (\theta\, \Lambda_2 - \beta\, \Lambda_1) = \theta\, F - \beta\, (\theta\, F - \overline{F}) = (1-\beta)\, \theta\, F + \beta\, \overline{F} = \overline{\Gamma} \qquad (19)$$

almost everywhere on $[\underline{\theta}, \theta_2]$, where $\Gamma = (1-\beta)\, \theta\, F' + F$ and $\overline{\Gamma}(\theta) = \int_{[\underline{\theta},\theta]} \Gamma(t)\, dt$. Furthermore, since F' is of bounded variation,

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

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

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

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

$$\frac{\overline{\Gamma}}{\theta - \underline{\theta}} \to (1-\beta)\, \underline{\theta}\, F'(\underline{\theta}+) > 0$$

as $\theta \downarrow \underline{\theta}$. Passing to the limit in equation (19), we therefore obtain

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

But all three terms on the RHS are strictly positive. Indeed: $\beta < 1$; $\underline{\theta} > 0$; and F' is bounded away from 0 on $(\underline{\theta}, \overline{\theta}\,)$. We have therefore reached a contradiction. This establishes that we cannot have $\theta_2 \in (\underline{\theta}, \overline{\theta}\,)$.

C.3 The Case $\theta_2 \in [\,\overline{\theta}, \infty\,)$

Consider now the case in which x_{liquid} and x_{illiquid} are such that $\theta_2 \in [\,\overline{\theta}, \infty\,)$. In this case, we can derive equations (15, 16 and 17) exactly as in Section C.2 above. In particular, we can still derive equation (17). We can therefore derive a contradiction by essentially the same argument.

C.4 The Case $\theta_2 \in (\,0, \underline{\theta}\,]$

Consider now the case in which x_{liquid} and x_{illiquid} are such that $\theta_2 \in (\,0, \underline{\theta}\,]$. In this case, we can still derive equations (15, 16 and 18). However, we can no longer derive equation (17). We therefore need new arguments. The first point to note is that, since $\theta_2 \leq \underline{\theta}$, all types $\theta \in [\,\underline{\theta}, \overline{\theta}\,]$ choose the point that a hypothetical θ_2 type would choose from the ambient budget set. We therefore have

$$\Lambda_1(\,\overline{\theta}\,) \;=\; \int_{[\underline{\theta},\overline{\theta}]} C_1'(r_1(t))\,dF(t) = F(\,\overline{\theta}\,)\,C_1'(r_1(\theta_2)), \tag{20}$$

$$\Lambda_2(\,\overline{\theta}\,) \;=\; \int_{[\underline{\theta},\overline{\theta}]} C_2'(r_2(t))\,dF(t) = F(\,\overline{\theta}\,)\,C_2'(r_2(\theta_2)). \tag{21}$$

Furthermore, since the θ_2 type chooses freely from the ambient budget set, we have

$$\frac{C_1'(r_1(\theta_2))}{\theta_2} = \frac{C_2'(r_2(\theta_2))}{\beta}.$$

Using (15) and (16), we therefore obtain

$$\frac{\overline{\theta}\,F(\,\overline{\theta}\,) - \overline{F}(\,\overline{\theta}\,)}{F(\,\overline{\theta}\,)} = \frac{\Lambda_1(\,\overline{\theta}\,)}{\Lambda_2(\,\overline{\theta}\,)} = \frac{C_1'(r_1(\theta_2))}{C_2'(r_2(\theta_2))} = \frac{\theta_2}{\beta}. \tag{22}$$

Hence

$$\begin{aligned}(\,\overline{\theta}-\theta_2\,)\,F(\,\overline{\theta}\,) &= \overline{\theta}\,F(\,\overline{\theta}\,)-\beta\left(\,\overline{\theta}\,F(\,\overline{\theta}\,)-\overline{F}(\,\overline{\theta}\,)\,\right) \\ &= (1-\beta)\,\overline{\theta}\,F(\,\overline{\theta}\,)+\beta\,\overline{F}(\,\overline{\theta}\,) \\ &= \overline{\Gamma}(\,\overline{\theta}\,), \end{aligned} \tag{23}$$

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

Remark 19 *Bearing in mind that $\theta_2 \leq \underline{\theta}$, so that $\overline{\Gamma}(\theta_2) = 0$, this equation can also be written*

$$(\overline{\theta}-\theta_2)\,F(\,\overline{\theta}\,) = \overline{\Gamma}(\,\overline{\theta}\,) - \overline{\Gamma}(\theta_2)$$

or

$$\frac{1}{\overline{\theta}-\theta_2}\int_{[\theta_2,\overline{\theta}]}\Gamma(t)\,dt = F(\,\overline{\theta}\,). \tag{24}$$

The significance of this observation is that θ_1 satisfies equation (24) too. So, while the necessary conditions that we have used here do not quite imply that $\theta_2 = \theta_1$, they do highlight a close relationship between the two. The intuitive reason for this relationship is clear. If $\theta_2 \leq \underline{\theta}$ then all types make the same choice. In particular, there are no interpersonal transfers. Since this outcome is – by hypothesis – the optimum in the class of outcomes with or without transfers, then a fortiori *it is the optimum in the class of outcomes without transfers.*

However, we have not yet used equation (18). It follows from this equation that

$$d\zeta_R = \tfrac{\beta}{\theta}\,(\,\theta\,F-\overline{F}\,)\,d\theta - F\,d\theta + \lambda\left(\Lambda_2 - \tfrac{\beta}{\theta}\,\Lambda_1\right)d\theta.$$

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

$$\zeta_R' = \tfrac{\beta}{\theta}\,(\,\theta\,F-\overline{F}\,) - F + \lambda\left(\Lambda_2 - \tfrac{\beta}{\theta}\,\Lambda_1\right).$$

Furthermore:

$$\begin{aligned}\Lambda_1(\theta) &= \int_{[\underline{\theta},\theta]} C_1'(r_1(t))\,dF(t) = F(\theta)\,C_1'(r_1(\theta_2)) = \frac{F(\theta)}{F(\,\overline{\theta}\,)}\Lambda_1(\,\overline{\theta}\,) \\ &= \frac{F(\theta)}{F(\,\overline{\theta}\,)}\,\frac{\theta_2}{\beta}\,\Lambda_2(\,\overline{\theta}\,) = \frac{F(\theta)}{F(\,\overline{\theta}\,)}\,\frac{\theta_2}{\beta}\,\frac{F(\,\overline{\theta}\,)}{\lambda} = \frac{\theta_2}{\beta}\,\frac{F(\theta)}{\lambda}\end{aligned}$$

(where the last line follows from (22) and (16)); and

$$\begin{aligned}\Lambda_2(\theta) &= \int_{[\underline{\theta},\theta]} C_2'(r_2(t))\,dF(t) = F(\theta)\,C_2'(r_2(\theta_2)) = \frac{F(\theta)}{F(\overline{\theta}\,)}\Lambda_2(\,\overline{\theta}\,) \\ &= \frac{F(\theta)}{F(\,\overline{\theta}\,)}\,\frac{F(\,\overline{\theta}\,)}{\lambda} = \frac{F(\theta)}{\lambda}\end{aligned}$$

(where the last line follows from (16)). Hence

$$\lambda\,(\theta\,\Lambda_2 - \beta\,\Lambda_1) = (\theta - \theta_2)\,F(\theta)$$

and

$$\begin{aligned}\theta\,\zeta_R' &= \beta\,(\,\theta\,F - \overline{F}\,) - \theta\,F + (\theta - \theta_2)\,F \\ &= (\theta - \theta_2)\,F(\theta) - \overline{\Gamma}.\end{aligned}$$

Now, $F(\underline{\theta}) = \overline{\Gamma}(\underline{\theta}) = 0$. Hence $\underline{\theta}\,\zeta_R'(\underline{\theta}) = 0$. Furthermore, we must have $\theta\,\zeta_R' \geq 0$ on $\big(\underline{\theta}, \overline{\theta}\big)$. Hence

$$\frac{\theta\,\zeta_R'(\theta) - \underline{\theta}\,\zeta_R'(\underline{\theta})}{\theta - \underline{\theta}} \geq 0.$$

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

$$(\theta\,\zeta_R')'(\underline{\theta}+) = (\beta\,\underline{\theta} - \theta_2)\,F'(\underline{\theta}+) \geq 0.$$

Since $F'(\underline{\theta}+) > 0$, it follows that

$$\theta_2 \leq \beta\,\underline{\theta}. \tag{25}$$

Similarly, (23) implies that $(\overline{\theta} - \theta_2)\,F(\,\overline{\theta}\,) - \overline{\Gamma}(\,\overline{\theta}\,) = 0$. Hence $\overline{\theta}\,\zeta_R'(\,\overline{\theta}\,) = 0$. Hence

$$\frac{\overline{\theta}\,\zeta_R'(\,\overline{\theta}\,) - \theta\,\zeta_R'(\theta)}{\overline{\theta} - \theta} \leq 0.$$

Letting $\theta \to \overline{\theta}-$, we therefore obtain

$$(\theta\,\zeta_R')'(\overline{\theta}-) = (\beta\,\overline{\theta} - \theta_2)\,F'(\overline{\theta}-) \leq 0.$$

Since $F'(\overline{\theta}-) > 0$, it follows that

$$\theta_2 \geq \beta\,\overline{\theta}. \tag{26}$$

But inequalities (25) and (26) are inconsistent with one another, so we have a contradiction.

Remark 20 *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*

$$\begin{aligned} (\theta\, \zeta_R')' &= (\theta - \theta_2)\, F' + F - \Gamma \\ &= (\theta - \theta_2)\, F' - (1-\beta)\, \theta\, F' \\ &= (\beta\, \theta - \theta_2)\, F'. \end{aligned}$$

Moreover we will have the boundary conditions $\underline{\theta}\, \zeta_R'(\underline{\theta}) = 0$ *and* $\overline{\theta}\, \zeta_R'(\overline{\theta}) = 0$*. It follows that* $\theta_2 \in \left(\beta\, \underline{\theta}, \beta\, \overline{\theta}\right)$ *and* $\theta\, \zeta_R' < 0$ *on* $\left(\underline{\theta}, \overline{\theta}\right)$*. 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* $(\underline{\theta}, \overline{\theta})$*.) In other words, it is always desirable to allow some flexibility to the decision maker to respond to the information contained in* θ*.*

D Differential Equations that Provide an Upper Bound for Welfare in the General Non-Linear Mechanism

Here we study the case of an economy populated by households with heterogeneous values of β. The case of homogeneous β is a simpler variant of the case studied in this section.

D.1 The General Non-Linear Problem

In the General Non-linear Mechanism, the planner chooses a budget set

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

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

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

subject to the resource constraint

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

and the incentive-compatibility constraint

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

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: $\Theta = [\,\underline{\theta}, \overline{\theta}\,]$; $B = [\,\underline{\beta}, \overline{\beta}\,]$; $0 < \underline{\theta} < \overline{\theta} < \infty$; $0 < \underline{\beta} < \overline{\beta} < \infty$; f is continuous and bounded away from 0 on Θ; g is continuous and bounded away from 0 on B.

Remark 21 *For example: f might take the form*

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

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 $[\,\underline{\theta}, \overline{\theta}\,]$; and g might take the form

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

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

D.2 Transforming the Problem

The first step in solving this problem is to note that

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

iff

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

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 $\Phi = [\,\underline{\phi}, \overline{\phi}\,]$ where $\underline{\phi} = \underline{\theta}\,/\,\overline{\beta}$ and $\overline{\phi} = \overline{\theta}\,/\,\underline{\beta}$, then the planner can work with consumption allocations $c_1, c_2 : \Phi \to (0,\infty)$ instead of with consumption allocations $c_1, c_2 : \Theta \times B \to (0,\infty)$.

The second step is to note that we can work with utility allocations $v_1, v_2 : \Phi \to \mathbb{R}$ instead of with consumption allocations $c_1, c_2 : \Phi \to (0, \infty)$. 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 $v_1, v_2 : \Phi \to \mathbb{R}$ 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,

$$0 = \phi\, v_1'(\phi) + v_2'(\phi) \tag{ICL}$$

and a monotonic part,

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

Remark 22 *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 (ICL) or (ICM), 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

$$h(\phi) = \int \beta\, f(\beta\,\phi)\, g(\beta)\, d\beta \tag{27}$$

and

$$j(\beta \mid \phi) = \frac{\beta\, f(\beta\,\phi)\, g(\beta)}{h(\phi)}. \tag{28}$$

We can also introduce the conditional expectation of β, namely

$$b(\phi) = \int \beta\, j(\beta \mid \phi)\, d\beta. \tag{29}$$

Remark 23 *The limits of integration in the definition of h (namely (27)) 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* $\min\{\,\overline{\beta}\,,\,\overline{\theta}\,/\,\phi\}$. *In particular, the support of the conditional distribution of β varies with ϕ:*

1. *For $\phi \in \left[\,\underline{\phi}, \min\{\,\underline{\theta}\,/\,\underline{\beta}\,,\,\overline{\theta}\,/\,\overline{\beta}\,\}\right]$, the support of β is $\left[\,\underline{\theta}\,/\,\phi\,,\,\overline{\beta}\,\right]$. In other words: the range of β types that is consistent with ϕ is increasing in ϕ, and this range always includes $\overline{\beta}$. 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 $\phi \in \left[\max\{\,\underline{\theta}\,/\,\underline{\beta}\,,\,\overline{\theta}\,/\,\overline{\beta}\,\}\,,\overline{\phi}\,\right]$, the support of β is $\left[\,\underline{\beta}\,,\,\overline{\theta}\,/\,\phi\,\right]$. 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 $\underline{\theta}\,/\,\underline{\beta} < \overline{\theta}\,/\,\overline{\beta}$ then, for $\phi \in \left[\min\{\,\underline{\theta}\,/\,\underline{\beta}\,,\,\overline{\theta}\,/\,\overline{\beta}\,\}\,,\max\{\,\underline{\theta}\,/\,\underline{\beta}\,,\,\overline{\theta}\,/\,\overline{\beta}\,\}\right]$, the support of β is $[\,\underline{\beta}\,,\,\overline{\beta}\,]$. 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 $\underline{\theta}\,/\,\underline{\beta} > \overline{\theta}\,/\,\overline{\beta}$ then, for $\phi \in \left[\min\{\,\underline{\theta}\,/\,\underline{\beta}\,,\,\overline{\theta}\,/\,\overline{\beta}\,\}\,,\max\{\,\underline{\theta}\,/\,\underline{\beta}\,,\,\overline{\theta}\,/\,\overline{\beta}\,\}\right]$, the support of β is $\left[\,\underline{\theta}\,/\,\phi\,,\,\overline{\theta}\,/\,\phi\,\right]$. 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

$$\int \Big(b(\phi)\,\phi\,v_1(\phi) + v_2(\phi)\Big)\,h(\phi)\,d\phi \tag{W}$$

and

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

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

The Langrangian for the one-dimensional problem can be written

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

where the Lagrange multipliers on the resource constraint, the incentive-compatibility constraint (ICL) and the incentive-compatibility constraint (ICM) take the form $\lambda \in \mathbb{R}$, $\mu : \Phi \to \mathbb{R}$ and $\nu : \Phi \to \mathbb{R}$.

D.3 The First-Order Conditions

In order to derive first-order conditions from this Langrangian, we must first eliminate v_1' and v_2'. We can do this by integrating by parts. Taking the third term of the Langrangian, 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

$$
\begin{aligned}
-\int ((\phi\,v_1)' + v_2')\,\mu\,h\,d\phi &= -\left[((\phi\,v_1) + v_2)\,\mu\,h\right]_{\underline{\phi}}^{\overline{\phi}} + \int ((\phi\,v_1) + v_2)\,(\mu\,h)'\,d\phi \\
&= \int ((\phi\,v_1) + v_2)\,(\mu\,h)'\,d\phi
\end{aligned}
$$

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

$$
\begin{aligned}
-\int v_2'\,\nu\,h\,d\phi &= -\left[v_2\,\nu\,h\right]_{\underline{\phi}}^{\overline{\phi}} + \int v_2\,(\nu\,h)'\,d\phi \\
&= \int v_2\,(\nu\,h)'\,d\phi.
\end{aligned}
$$

The Langrangian can therefore be written

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

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

$$0 = \Big(b\,\phi + \mu - \lambda\, C_1'(v_1) \Big)\, h + \phi\,(\mu\, h)'$$

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

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

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

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

D.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

$$0 = \Big(b\,\phi + \mu - \lambda\, C_1'(v_1) \Big)\, h + \phi\,(\mu\, h)', \tag{30}$$

$$0 = \left(1 - \lambda\, \frac{1}{R}\, C_2'(v_2) \right) h + (\mu\, h)', \tag{31}$$

$$0 = \phi\, v_1' + v_2' \tag{32}$$

and the integral equation

$$0 = \int \left(Y - C_1(v_1) - \frac{1}{R}\, C_2(v_2) \right) h \, d\phi. \tag{33}$$

The first step is to make v_1 and v_2 the subjects of equations (30) and (31). Putting $U_1 = (C_1')^{-1}$ and $U_2 = (C_2')^{-1}$, we obtain

$$v_1 = U_1\Big(\frac{a_1}{\lambda}\Big), \tag{34}$$
$$v_2 = U_2\Big(\frac{a_2}{\lambda}\Big), \tag{35}$$

where

$$a_1 = b\,\phi + \mu + \frac{\phi\,(\mu\, h)'}{h}, \tag{36}$$
$$a_2 = R\left(1 + \frac{(\mu\, h)'}{h}\right). \tag{37}$$

D.5 Solving (30-32) where b and h are Smooth

Consider the equations (30-32) in the open region $\mathring{\Phi} = \Phi \smallsetminus \{\, \underline{\phi}\,, \underline{\theta}\,/\,\underline{\beta}\,, \overline{\theta}\,/\,\overline{\beta}\,, \overline{\phi}\,\}$. In this region, both b and h are smooth. Hence we may differentiate (34,35) to obtain

$$v_1' = U_1'\Big(\frac{a_1}{\lambda}\Big)\, \frac{a_1'}{\lambda}, \tag{38}$$
$$v_2' = U_2'\Big(\frac{a_2}{\lambda}\Big)\, \frac{a_2'}{\lambda} \tag{39}$$

and, substituting (38,39) in (32),

$$0 = \phi\, U_1'\Big(\frac{a_1}{\lambda}\Big)\, \frac{a_1'}{\lambda} + U_2'\Big(\frac{a_2}{\lambda}\Big)\, \frac{a_2'}{\lambda}.$$

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

$$0 = \left(\phi^2\, U_1'\,(a_1) + R\, U_2'\,(a_2)\right) h^2\,\mu''$$
$$+ \left(\phi\,(\phi\, h' + 2\, h)\, U_1'\,(a_1) + h'\, R\, U_2'\,(a_2)\right) h\,\mu'$$
$$+ \left(\phi\left(h\,(\phi\, h'' + h') - \phi\, h'^2\right) U_1'\,(a_1) + \left(h\, h'' - h'^2\right) R\, U_2'\,(a_2)\right) \mu$$
$$+\phi\,(\phi\, b' + b)\, U_1'\,(a_1)\ h^2. \tag{40}$$

In other words, in the region $\mathring{\Phi}$, equations (30-32) reduce to a second-order ordinary differential equation for μ.

D.6 Solving (30-32) where b and h have Kinks

Now consider the equations (30-32) at the points $\phi_1 = \underline{\theta}\,/\,\underline{\beta}$ and $\phi_2 = \overline{\theta}\,/\,\overline{\beta}$, where both b and h have kinks. We cannot differentiate (34,35) 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

$$0 = \Delta\mu(\phi_i) = \mu(\phi_i+) - \mu(\phi_i-) \tag{41}$$

and the incentive condition

$$0 = \phi_i\,\left(U_1(a_1(\phi_i+)) - U_1(a_1(\phi_i-))\right) + \left(U_2(a_2(\phi_i+)) - U_2(a_2(\phi_i-))\right). \tag{42}$$

D.7 Solving (30-32) at the Endpoints

Assuming for concreteness that $\phi_1 < \phi_2$, we now have the second-order ordinary differential equation (40) in the three open intervals $(\underline{\phi}, \phi_1)$, (ϕ_1, ϕ_2) and $(\phi_2, \overline{\phi})$. Moreover, we have two boundary conditions at each of ϕ_1 and ϕ_2. (Cf. (41) and (42).) The obvious way of completing the equation would therefore be to require that μ take on appropriate values at the boundaries $\underline{\phi}$ and $\overline{\phi}$. However, h decays linearly to 0 at both $\underline{\phi}$ and $\overline{\phi}$. Moreover, inspection of (40) shows that:

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

Hence μ will not take on boundary values at $\underline{\phi}$ and $\overline{\phi}$ in the usual way.[55] 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 $\overline{\phi}$. In particular, it is bounded. Hence the relevant solution of the equation is the one that is bounded near $\underline{\phi}$ and $\overline{\phi}$.[56]

D.8 Solving for λ

As we have seen, we can find μ by solving the second-order o.d.e. (40) 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 $\overline{\phi}$. 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 (33).

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

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

[55] Intuitively speaking, the dynamics of ϕ move away from the endpoints $\underline{\phi}$ and $\overline{\phi}$.

[56] Since the inhomogeneous term is of order h^2 near $\underline{\phi}$ and $\overline{\phi}$, the solution can in fact be expected to decay quadratically to 0 at both $\underline{\phi}$ and $\overline{\phi}$. In particular, we would expect that it would satisfy $\mu(\underline{\phi}) = \mu'(\underline{\phi}) = 0$ and $\mu(\overline{\phi}) = \mu'(\overline{\phi}) = 0$. 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.

Hence, substituting in (33),

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

or

$$\lambda^{\frac{1}{\gamma}}=\frac{\int\left(a_1^{\frac{1}{\gamma}}+\frac{1}{R}a_2^{\frac{1}{\gamma}}\right)h\,d\phi}{\int Y\,h\,d\phi}. \tag{43}$$

Bearing in mind that a_1 and a_2 are given in terms of μ by equations (36) and (37), this gives us a formula for λ in terms of μ.

D.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 (34) and (35); 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 $c_2 = C_2(v_2)$.

D.10 Numerical implementation

We generate a numerical solution (using Matlab's bvp4c function[57]) for the second-order differential equation for μ (equation 40) with the boundary conditions described in section D.7 of this appendix. In order to calculate welfare, we solve the second-order differential equation simultaneously with two other first-order differential equations. Our procedure to obtain such system of o.d.e.'s is explained below.

Notice that the numerator of $\lambda^{\frac{1}{\gamma}}$, given by (43), is a definite integral. Its value can be accurately obtained by adding an appropriate expression to the system of differential equations. Let:

$$\begin{aligned}Num\lambda(\phi)&=\int_{\underline{\phi}}^{\phi}\left(a_1(x)^{\frac{1}{\gamma}}+\frac{1}{R}a_2(x)^{\frac{1}{\gamma}}\right)h(x)dx\\ \frac{\partial Num\lambda(\phi)}{\partial\phi}&=\left(a_1(\phi)^{\frac{1}{\gamma}}+\frac{1}{R}a_2(\phi)^{\frac{1}{\gamma}}\right)h(\phi)\end{aligned} \tag{44}$$

Going by these definitions, we are interested in calculating $Num\lambda(\overline{\phi})$, which is exactly the terminal condition that one obtains when solving the o.d.e. given by (44). The boundary

[57]See https://www.mathworks.com/help/matlab/ref/bvp4c.html

conditions for $Num\lambda(\phi)$ are straightforward to obtain and are given by:

$$
\begin{aligned}
&Num\lambda(\underline{\phi}) = 0 \\
&Num\lambda(\phi_i+) - Num\lambda(\phi_i-) = 0
\end{aligned}
$$

Next, optimized welfare (from the planner's perspective) is given by:

$$W^{opt} = \int (b(\phi)\phi v_1(\phi) + v_2(\phi))h(\phi)d\phi$$

but it cannot be simultaneously calculated with a similar procedure as the previous one, because v_1 and v_2 are given in terms of λ and λ is only obtained after solving the system of o.d.e.'s. To go past this problem, consider the following affine transformation of W^{opt}, where we have plugged in (34), (35), U_1, and U_2 into the definition of W^{opt} and λ has been factorized out of the RHS:

$$
\begin{aligned}
W^{opt} \cdot \lambda^{\frac{1-\gamma}{\gamma}} + \frac{\lambda^{\frac{1-\gamma}{\gamma}}}{(1-\gamma)} \left(\int \phi b(\phi) h(\phi) d\phi + 1 \right) = \\
\underbrace{\int \left(b(\phi)\phi \frac{(a_1(\phi))^{\frac{1-\gamma}{\gamma}}}{(1-\gamma)} + \frac{(a_2(\phi))^{\frac{1-\gamma}{\gamma}}}{(1-\gamma)} \right) h(\phi) d\phi}_{\equiv \hat{W}}
\end{aligned}
\tag{45}
$$

We can now solve for $\hat{W}$ just like we did for $Num\lambda$, by adding its corresponding o.d.e. to the system and solving them all simultaneously using Matlab's bvp4c. Finally, we can use (45) to recover W^{opt}.

D.11 The Case with Homogeneous Present Bias

The preceding derivations and numerical implementation correspond to the Relaxed Problem with heterogeneous present bias. The problem with homogeneous present bias is a special case of the previous one and its solution procedure differs in the following aspects.

Analytically, the derivation of the solution only differs in Section D.3, where one cannot use the result that $h(\bar{\phi}) = h(\underline{\phi}) = 0$. Solving the Relaxed Problem without using that result leads to the exact same second-order differential equation for μ (40). This occurs because the new first-order conditions of the problem directly imply that $\mu(\underline{\phi}) = \mu(\bar{\phi}) = 0$. Replacing this information in the remaining FOCs leads to the same set of equations as in the heterogeneous present bias problem.

Notice as well that the discussion in Section D.6 does not apply to the case with homogeneous present bias. This happens because each of the open intervals $(\underline{\phi}, \phi_1)$ and $(\phi_2, \bar{\phi})$ collapses to a single point as $Var(\beta) \to 0$. In this sense, the homogeneous present bias case can be thought of as a limiting case of the heterogeneous present bias case and (40) can be solved in a single interval in Φ. On the contrary, the heterogeneous present bias problem had to be solved in up to three open intervals. Hence, switching from solving the homogeneous to the heterogeneous present bias case entails switching from solving a regular boundary value problem to a multipoint boundary value problem. This increases the complexity of the programming required to obtain a numerical solution with Matlab's bvp4c function. Hence, and for ease of exposition, we present two separate pieces of code in the replication materials: one for the homogeneous present bias case and another code for the heterogeneous one.

E Analysis of the Quasi-Linear Limit Case

E.1 Proof of Proposition 1

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 internality generated by over-consumption. 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.

E.2 Proof of Proposition 2

In Subsections 3.1 and 4.1, 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

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

subject to

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

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

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.

We now turn to proving Proposition 2.

E.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 problem (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 $\frac{d}{d\phi}\left(\phi v_1(\phi) + v_2(\phi)\right) = v_1(\phi)$, and we obtain integral incentive compatibility by integrating:

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

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

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

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

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

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 $v_2(\overline{\phi})$in terms of v_1:

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

E.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 $\mathrm{E}[\,\theta\,|\,\phi\,]\, v_1(\phi) + v_2(\phi)$. Therefore, the planner objective function is:

$$\int \Big(\mathrm{E}[\,\theta\,|\,\phi\,]\, v_1(\phi) + v_2(\phi)\Big)\, dH(\phi).$$

Substituting in the characterization of v_2 above, we get:

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

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.

From here, we solve the relaxed problem, not subject to monotonicity. The relaxed problem is simply given by

$$\max_{v_1} \left\{ \int \Big(\mathrm{E}[\,\theta\,|\,\phi\,]\, v_1(\phi) - u_1^{-1}\left(v_1(\phi)\right)\Big)\, dH(\phi) + Y \right\}$$

and so has a solution given by the first order condition for optimal allocation

$$\mathrm{E}[\,\theta\,|\,\phi\,]\, u_1'\left(c_1(\phi)\right) = 1.$$

From here, all that remains is to verify that this allocation satisfies monotonicity. Monotonicity arises provided that $\mathrm{E}[\,\theta\,|\,\phi\,]$ is non-decreasing. Hence, provided $\mathrm{E}[\,\theta\,|\,\phi\,]$ is non-decreasing, we have characterized the optimal allocation.

E.2.3 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:

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

E.2.4 Homogeneous β

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

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

That is, we simply have a Pigouvian tax. This gives another proof of Proposition 1.

E.2.5 Heterogeneous β

If β is heterogeneous *and* the regularity condition of Proposition 2 is satisfied, then as mentioned before we have:

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

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 $1 - \overline{\beta}$ 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 $\overline{\phi}$. This suggests a large degree of flexibility over initial withdrawals, and much tighter restrictions on flexibility for households withdrawing a lot.

E.2.6 Corollary 3

The joint density of (ϕ, θ) takes the form $\theta\phi^{-2}f(\theta)g(\theta\phi^{-1})$ (up to normalization to integrate to one). Thus if we are in the beta wide case, we can write

$$E[\theta|\phi] = \begin{cases} \frac{\int_{\underline{\theta}}^{\overline{\beta}\phi} \theta^2\phi^{-2}f(\theta)g(\theta\phi^{-1})d\theta}{\int_{\underline{\theta}}^{\overline{\beta}\phi} \theta\phi^{-2}f(\theta)g(\theta\phi^{-1})d\theta}, & \phi < \frac{\overline{\theta}}{\overline{\beta}} \\ \\ \frac{\int_{\underline{\theta}}^{\overline{\theta}} \theta^2\phi^{-2}f(\theta)g(\theta\phi^{-1})d\theta}{\int_{\underline{\theta}}^{\overline{\theta}} \theta\phi^{-2}f(\theta)g(\theta\phi^{-1})d\theta}, & \frac{\overline{\theta}}{\overline{\beta}} \leq \phi \leq \frac{\underline{\theta}}{\underline{\beta}} \\ \\ \frac{\int_{\underline{\beta}\phi}^{\overline{\theta}} \theta^2\phi^{-2}f(\theta)g(\theta\phi^{-1})d\theta}{\int_{\underline{\beta}\phi}^{\overline{\theta}} \theta\phi^{-2}f(\theta)g(\theta\phi^{-1})d\theta}, & \frac{\underline{\theta}}{\underline{\beta}} < \phi \end{cases}$$

If β is uniformly distributed, this simplifies to

$$E[\theta|\phi] = \begin{cases} \frac{\int_{\underline{\theta}}^{\overline{\beta}\phi} \theta^2 f(\theta)d\theta}{\int_{\underline{\theta}}^{\overline{\beta}\phi} \theta f(\theta)d\theta}, & \phi < \frac{\overline{\theta}}{\overline{\beta}} \\ \\ \frac{\int_{\underline{\theta}}^{\overline{\theta}} \theta^2 f(\theta)d\theta}{\int_{\underline{\theta}}^{\overline{\theta}} \theta f(\theta)d\theta}, & \frac{\overline{\theta}}{\overline{\beta}} \leq \phi \leq \frac{\underline{\theta}}{\underline{\beta}} \\ \\ \frac{\int_{\underline{\beta}\phi}^{\overline{\theta}} \theta^2 f(\theta)d\theta}{\int_{\underline{\beta}\phi}^{\overline{\theta}} \theta f(\theta)d\theta}, & \frac{\underline{\theta}}{\underline{\beta}} < \phi \end{cases}$$

It follows that $E[\theta|\phi]$ is constant over the middle interval, giving rise to a pooling region (part 1 of the result). Observe then that

$$E[\beta|\phi] = E[\frac{\theta}{\phi}|\phi] = \frac{1}{\phi}E[\theta|\phi]$$

which therefore decreases over the middle region, confirming part 2 of the result.