Optimal Illiquidity*

John Beshears † James J. Choi‡ Christopher Clayton§
Christopher Harris¶ David Laibson‖ Brigitte C. Madrian∗∗

May 2024

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, $\beta$, 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 TIAA Institute and Wharton School’s Pension Research Council/Boettner Center, the Eric M. Mindich Fund for Research on the Foundations of Human Behavior, the National Institute on Aging (P30AG034532), the Pershing Square Fund for Research on the Foundations of Human Behavior, and the Social Security Administration (#RRC08098400-09). The views expressed in this paper are not endorsed by the Social Security Administration or other funders. The authors received insightful comments from Toni Whited (editor), anonymous referees, 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 and NBER. Email: jbeshears@hbs.edu
‡Choi: Yale and NBER. Email: james.choi@yale.edu
§Clayton: Yale and NBER. Email: christopher.clayton@yale.edu
¶Harris: Cambridge. Email: cjharris@econ.cam.ac.uk
‖Laibson (Corresponding Author): Harvard and NBER. Email: dlaibson@harvard.edu
∗∗Madrian: BYU and NBER. 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). Moreover, 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/IRA retirement accounts of U.S. households under age 55, $0.40 simultaneously flows out of the DC/IRA system for the same age group (Argento, Bryant, and Sabelhaus, 2015). It is not clear whether allowing such leakage is consistent with overall social welfare maximization, because leakage creates potentially valuable short-run liquidity and depletes future retirement consumption, although most media coverage bemoans leakage.

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 account-specific penalties for pre-retirement withdrawals: i.e., with $N$ accounts, each account has an early withdrawal penalty of $\pi_n$ for $n = 1, 2, ..., N$. We also derive the welfare obtainable from a fully general account system: i.e., an account with general non-linear penalties for pre-retirement withdrawals. We show that a very simple two-account system with one completely liquid account and one completely illiquid account approximately attains the social welfare that would be obtained by a very complex fully general account system with non-linear penalties.

1 Rollovers to an IRA are not counted as leakage for this calculation because rollovers are just a transfer within the DC system – from a 401(k) to an IRA. Loans in the 401(k) system are also not counted as outflows because, on dollar-weighted basis, the overwhelming majority of 401(k) loans are repaid.

2 About half of leakage on a dollar-weighted basis occurs in categories that that avoid the 10% penalty.

We study preferences that include both normative taste shocks and non-normative self-control problems due to present bias. Specifically, we incorporate 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 studies households with homogeneous present bias \(1 - \beta\) and heterogeneous taste shocks \(\theta\), with each household’s \(\theta\) being private information. AWA does not allow for policies that admit inter-household transfers, and finds that mandatory minimum savings is optimal. Specifically, AWA find that the socially optimal system consists of two accounts, where one account is completely liquid and the other is completely illiquid until period 2.\(^4\)

In the current paper, we revisit AWA’s setting and make two changes that reflect important economic considerations faced by real-world policymakers. First, we allow the government to choose policies that generate inter-household transfers. In other words, households are not isolated economically: withdrawal penalties paid by one household can be distributed back to all households (i.e., tax and transfer). Second, we allow for unobserved heterogeneity in present bias: i.e., \(\beta\) varies across households and the social planner knows the distribution of \(\beta\), but each household’s \(\beta\) is private information to that household. We show that, absent heterogeneity of \(\beta\), a simple Pigouvian tax is approximately optimal (i.e., a one-account system with an early withdrawal tax of \(1 - \beta\) is approximately optimal). However, with sufficient heterogeneity of \(\beta\), which we view as the empirical benchmark, the optimal account system is well-approximated by a two-account system, with one completely liquid account and one completely illiquid account (echoing AWA’s social optimum, but arising from different economic forces).

The central contribution of the current paper is therefore to show that a simple two-account system that mirrors actual institutions in almost all middle- and high-income economies

\(^4\)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).
achieves social welfare that is very close to the welfare that would be achieved by a fully general (non-linear) retirement saving system. 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 savings regimes. We refer to these as mechanisms, adopting the language of the mechanism design literature, which is the analytic framework that we use to derive the upper bound for social welfare.

The first class is the class of mechanisms with $N$ accounts and account-specific linear penalties, in which households are free to draw down the $N$ accounts in whatever order they prefer. The class of $N$-Account Systems is restrictive (from the point of view of the planner) in two ways. First, an $N$-Account System has a finite number of accounts, each of which has a linear penalty for early (i.e., period 1) withdrawals. Second, an $N$-Account System 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 middle- and high-income countries, this first class of mechanisms is the focus of our numerical analysis. We denote the planner social welfare achieved under an $N$-Account System as $W^N$.

The second class of mechanisms consists of general non-linear mechanisms.\(^5\) 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 piece-wise linear budget sets (which were assumed in the first class of mechanisms).\(^6\)

The third class of mechanisms consists of general non-linear mechanisms satisfying the standard Local Incentive Compatibility Constraint but omitting the Monotonicity constraint (both constraints are necessary and sufficient for Global Incentive Compatibility). As is

\(^5\)Technically, this solution satisfies the standard Global Incentive Compatibility constraint, re-expressed as usual as a Local Incentive Compatibility constraint and a Monotonicity constraint.

\(^6\)Our first class of mechanisms – $N$ accounts with a flexible draw-down rule – is mathematically equivalent to assuming a piece-wise linear, convex budget set. Piece-wise 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.
standard, we refer to this solution, which drops the Monotonicity constraint, 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. This upper bound is useful because the Relaxed Problem is simpler to solve than the General Mechanism due to the absence of the Monotonicity constraint. Our numerical solution for the Relaxed Problem makes use of the ordinary differential equation (ODE) that characterizes the planner’s solution to the Relaxed Problem (see Appendix C). We denote the planner welfare achieved under this Relaxed Problem as $W^R$.

Because of the progressive relaxation of constraints, the three classes of mechanisms generate ranked levels of welfare (where welfare is evaluated from the perspective of the social planner). To provide a welfare benchmark (a baseline for comparison), we contrast all three mechanisms to welfare achieved in an autarkic economy in which all households are given a single liquid account (with a household storage technology) and inter-household transfers are not possible. We denote planner welfare under this autarkic benchmark by $W^A$.\footnote{Because we assume naivete, the autarky case does not generate self-imposed illiquidity.} We then have

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

In summary, we study the autarkic benchmark and three interventionist policy regimes: $N$-Account System, General Mechanism, and a Relaxed Problem (the last policy regime serves to generate an upper bound on welfare). Our analysis finds \emph{de minimus} welfare improvement when moving from an $N \leq 2$ Account System to the Relaxed Problem. Accordingly, the welfare improvement from the $N \leq 2$ Account System to the General Mechanism must also be \emph{de minimis} (because $W^G - W^N$ is necessarily weakly smaller than $W^R - W^N$). In other words, if $N \leq 2$ makes welfare under an $N$-Account System close to welfare under the Relaxed Problem, then $N \leq 2$ must also make welfare under that $N$-Account System close
to welfare under the General Mechanism.

In Section 3, we study the homogeneous-\(\beta\) case. We show that a single account, \(N = 1\), suffices to achieve a close welfare approximation to the General Mechanism. In particular, the 1-account system that approximates the theoretically optimal welfare consists of a partially illiquid account with (Pigouvian) early-withdrawal penalty \(\pi \approx 1 - \beta\).

In Section 4, we study the heterogeneous-\(\beta\) case. We show that two accounts, \(N = 2\), suffice to achieve a close welfare approximation to the General Mechanism. In particular, an \(N = 2\) Account system with a completely liquid account and a completely illiquid account closely approximates the welfare attainable under the General Mechanism. 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.\(^8\) The completely illiquid savings account generates large welfare gains for households with relatively low \(\beta\) values. These welfare gains swamp the welfare losses of the high-\(\beta\) households (who are made only slightly worse off by being forced to shift some of their wealth from completely liquid accounts to completely illiquid accounts; high-\(\beta\) households do lose some flexibility, but this welfare loss is small in all of our calibrated cases).

As implied by the welfare bounds discussed above, extending the \(N = 2\) Account System by adding a third account (on top of the completely liquid and completely illiquid accounts) delivers minimal welfare gains. The planner-optimized penalty on this third account is approximately 13%. Hence, the partially illiquid accounts, which are derived in this paper from first principles, look remarkably like a 401(k) account in the U.S., which has a 10% early withdrawal penalty. Moreover, the derived partially illiquid account generates 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.,

---

\(^8\)In our model the planner collects all of the resources from households and then deposits funds into accounts. This is equivalent 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).
The partially illiquid account generates modest welfare costs to households with low \( \beta < 1 \). The partially illiquid account generates modest welfare costs to households with low \( \beta \) values (who end up paying most of the early-withdrawal penalties) and generates offsetting (slightly greater) welfare benefits to households with relatively high \( \beta \) values (who benefit from a fiscal externality and flexibility). Hence, in our benchmark model, leakage is not a social problem. However, we go on to show that leakage is not benign when society relies heavily on 401(k)-style accounts for retirement wealth formation. In our benchmark model, leakage from partially illiquid accounts is not a social problem because 401(k)-style accounts have small balances, with the overwhelming majority of retirement wealth being held in completely illiquid accounts (e.g., illiquid DC systems and/or illiquid DB/Social Security systems).

We also use our stylized two-period model with \( N = 3 \) accounts, to generate an empirically calibrated model for the existing account system in the U.S. economy (Section 4.5). Specifically, our calibration values are chosen to match the empirically observed quantities of completely illiquid wealth, partially illiquid wealth, and completely liquid wealth for a typical household. Our stylized calibration generates a leakage rate of 40\%, consistent with historical empirical estimates. Although we find that the existing accounts in the U.S. system are approximately optimal with respect to their different liquidity properties, our stylized model implies that the allocations to those accounts are not. Specifically, our model implies that a materially higher allocation should be made to the completely illiquid account (including Social Security claims): 48\% more wealth accumulated during working life in the completely illiquid account relative to the magnitude of wealth accumulation in completely illiquid accounts under existing U.S. policy.

There is a growing literature that studies how present bias effects retirement savings and how governments should optimally respond.\textsuperscript{9} Our model is related to the independent and

\textsuperscript{9}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).
contemporaneous work of Moser and Olea de Souza e Silva (2019), who study an environment with unobservable earnings ability, unobservable $\beta$, 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.\footnote{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.} We contribute to the literature that identifies settings in which very simple mechanisms provide good welfare approximations to arbitrarily complex, optimal mechanisms.\footnote{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).}

Finally, a large literature studies how firms attempt to exploit agents with present bias.\footnote{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. (2020).} By contrast, our paper studies how a benevolent social 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. The online appendix contain proofs and extensions, including a method for calculating welfare for the Relaxed Problem (Appendix C).

2 Model

We study a two-period model of consumption for a continuum of households, with idiosyncratic taste shock $\theta$ and idiosyncratic present bias $\beta$. 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 $\theta$ and $\beta$ 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 System 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 System 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 Systems 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 $\theta$ is a stochastic taste shifter,$^{13}$ $u_t : (0, \infty) \to \mathbb{R}$ is the period-$t$ utility function, $c_t$ is period-$t$ consumption, $\beta$ is the present-bias discount factor, and $\delta$ is the standard discount factor.$^{14}$ 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 naivete 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 naivete is critical for the results of this paper because it eliminates the motive for self-commitment as well as the opportunity for government/planner screening in a hypothetical ‘pre-period’.$^{15}$

We assume that taste shifters, $\theta$, and present bias, $\beta$, are private information of each household in the economy. The social planner knows the aggregate distribution of $(\theta, \beta)$ across households. We denote the distribution function of $\theta$ by $F(\cdot)$ and of $\beta$ by $G(\cdot)$. In our analysis, we assume that $\theta$ and $\beta$ are independent.$^{16}$

### 2.3 Preferences of the social planner

The social planner and the household (with taste shifter $\theta$) 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

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

---

$^{13}$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 $\theta$ is a taste shifter. This case is beyond the scope of the current paper.

$^{14}$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.

$^{15}$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).

$^{16}$We can generalize this framework to allow for a joint distribution (see Appendix D.4).
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 social objective:

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

(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 $\theta$ and $\beta$ 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 $\theta$ values to agents with high $\theta$ values. The first motive is an inter-temporal reallocation (within a household) and the second motive is an inter-personal redistribution.

### 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 $\pi_n$, which applies only to withdrawals in period 1 (i.e., an early-withdrawal penalty).\(^{18}\) If a consumer withdraws $\omega$ dollars from account $n$ in period 1, the consumer actually receives $(1 - \pi_n)\omega$ dollars.\(^{19}\) A completely liquid account has $\pi_n = 0$, a partially liquid account has $0 < \pi_n < 1$, and a com-

\(^{17}\)We can generalize this framework to incorporate Pareto weights (see Appendix D.4).

\(^{18}\)WLOG, there are no withdrawal penalties in period 2.

\(^{19}\)The framework admits negative penalties for period 1 consumption (i.e., subsidies).
pletely 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 System

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 $\theta$ and $\beta$ maximizes

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

where consumption is given by

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

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

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, $\theta$ and $\beta$, is the same as integrating over consumers. In period 0, the planner
faces the problem

$$\max_{(x_n)^N_{n=1}, (\pi_n)^N_{n=1}} \int \int \left( \theta u_1(c_1(\theta, \beta)) + \delta u_2(c_2(\theta, \beta)) \right) dF(\theta) dG(\beta)$$  \hspace{1cm} (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:

$$\int \int \left( c_1(\theta, \beta) + \frac{c_2(\theta, \beta)}{R} \right) dF(\theta) dG(\beta) \leq Y.$$  \hspace{1cm} (6)$$

In other words, the planner chooses the account allocation vector, $(x_n)^N_{n=1}$, and the penalty vector, $(\pi_n)^N_{n=1}$, 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).\footnote{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 $x = (1 + m) z$.}

We choose to use an *involuntary* framing in our model presentation because it is without loss of generality and notionally 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).\footnote{Some of these systems are funded, some are unfunded, and some are hybrid. The key unifying feature (for the purposes of our model) is that they are involuntary.}

The $N$-account System 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.

20  
21
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),

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

(7)

ruling out inter-household transfers. As in AWA, households have homogeneous $\beta$.

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-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.\textsuperscript{22}

We now embed the AWA result in the conceptual framework described in the introduction: i.e., the three classes of welfare-ranked mechanisms.\textsuperscript{23} 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 System and the

\textsuperscript{22}See Ambrus and Egorov (2013) for cases (that do not satisfy our assumptions) in which money burning arises.

\textsuperscript{23}Because we directly know the General Non-linear Mechanism, we do not need to discuss the welfare bound provided by the Relaxed Problem.
3 Optimal Liquidity with Homogeneous Present Bias and Inter-Household Transfers

We now study the case in which present bias $\beta$ 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 B 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-$\theta$ 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.\footnote{This involves no loss of generality because utility functions can be rescaled.} We now turn to studying socially optimal 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 System and the General Non-linear Mechanism are identical.\footnote{Because we directly know the General Non-linear Mechanism, we do not need to discuss the welfare bound provided by the Relaxed Problem.} 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.\footnote{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.}

**Proposition 1** Suppose that all households have the same value of $\beta$. 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 D.

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 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 System 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

\[
\int (\theta v_1(\theta) + v_2(\theta)) f(\theta) \, d\theta \quad \text{(Planner Objective)}
\]

subject to the resource constraint

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

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,

\[
0 = \theta v'_1(\theta) + \beta v'_2(\theta) \quad \text{(Local IC)}
\]

and a monotonic part,

\[
0 \leq -v'_2(\theta). \quad \text{(Monotonicity)}
\]

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.\(^{27}\) We characterize the solution to the Relaxed Problem using a system of differential equations (see Appendix C). For the homogeneous-\(\beta\)

\(^{27}\)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-\(\beta\) case and close for the heterogeneous-\(\beta\) case.
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 the optimal $N$-account System 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 System with $N = 1$ account and the Relaxed Problem, implying that the $N$-account System 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 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 to integrate to unit mass.

Our simulations use a coefficient of relative risk aversion of one (Assumption S1), a magnitude that is consistent with structural econometric estimates from lifecycle savings models (Gourinchas and Parker, 2002; Laibson et al., Forthcoming).

Due to truncation of the tails, assumption S2 implies that a one standard deviation taste shock will now induce marginal utility in period 1 to vary by $\pm 24.2\%^{28}$ holding consumption constant. We view this as a plausible magnitude given the myriad uninsurable shocks that buffet households during working life, but we are not aware of structural estimates of this parameter; this is an important domain for future research. To gain intuition for a quantitative interpretation of this effect, under our benchmark assumption of $\ln$ utility, a $\varepsilon\%$ increase in marginal utility due to a taste shock is equal in impact on marginal utility to a

\footnote{This is slightly less than $\sigma = 0.25$ because of the truncation of the deep tails.}
z% fall in net resources available for regular consumption.\textsuperscript{29} Our parametric assumptions in S2 have three conceptual motivations. First, we use a unimodal density that is bell-shaped to reflect an underlying assumption that our taste shock $\theta$ aggregates many separate shocks to produce variation in spending needs. (Sources of variation include categories like medical spending, support for parents and other family members experiencing financial hardship, educational expenses for children, local cost of living, stochastic durable depreciation like the need for a new roof, spending norms established by peers, etc.) Second, we truncate the tails to generate a strictly positive density on a bounded support\textsuperscript{30} and to avoid a lifetime utility function with a negative weight on $u_1$. Third, we center the distribution of shocks (i.e., $\mu = 1$) so that marginal utility during working life (period 1) is on average the same as marginal utility in retirement (period 2), holding consumption fixed. This centering assumption implies that there are counterbalanced reasons for marginal utility to be relatively high in retirement (e.g., ample time to travel and elevated medical expenses, including uninsured long-term care) and reasons for marginal utility to be relatively low in retirement (e.g., health is a complement with certain consumption activities, like travel). In Section 5, we evaluate the robustness of our paper's findings with respect to variation in all of our parametric assumptions.

Finally, the combination of S1 and S2 pins down the utility function; hence, the assumption $\delta R = 1$ is no longer made without loss of generality (because the utility function, once specified, can’t be adjusted to offset the value of $\delta R$). Structural econometric models of lifecycle behavior with present bias estimate $\delta$ values close to one when period lengths are a year (e.g., Laibson et al. (Forthcoming) study a present-biased model with annual periods and estimate $\delta = 0.99$). The risk-free, short-run (inflation-adjusted, annual) gross interest rate tends to be close to 1.01.\textsuperscript{31} Hence, the assumption $\delta R = 1$ is consistent with

\textsuperscript{29}For example, multiplying ln utility by $z$ has the same effect on marginal utility of subtracting $z\%$ from the consumption bundle because those funds are needed for a separable expense, like a home repair.

\textsuperscript{30}For technical reasons, we need $\theta$ to have a strictly positive density on its support (including the boundaries) to make our general mechanism design problem well-behaved.

\textsuperscript{31}For example, Aswath Damodaran reports an average real return on 3-month Treasury Bills of 0.32% from 1928 to 2023.
Table 1: Welfare gains from four versions of the homogeneous-β model: an N-account System with N = 1; an N-account System with N = 2; the Relaxed Problem, which is the relaxed version of the General Non-linear Mechanism; an N-account System with N = 2 accounts that have exogenously set penalties, specifically π₁ = 0 and π₂ = 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).

<table>
<thead>
<tr>
<th>Value of β</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-account System: N = 1</td>
<td>69.658</td>
<td>31.775</td>
<td>17.124</td>
<td>9.630</td>
<td>5.347</td>
</tr>
<tr>
<td>N-account System: N = 2</td>
<td>71.648</td>
<td>32.698</td>
<td>17.605</td>
<td>9.882</td>
<td>5.478</td>
</tr>
<tr>
<td>Relaxed Problem</td>
<td>71.674</td>
<td>32.748</td>
<td>17.659</td>
<td>9.928</td>
<td>5.511</td>
</tr>
<tr>
<td>N-account System: N = 2, π₁ = 0, π₂ = 1</td>
<td>71.633</td>
<td>32.648</td>
<td>17.482</td>
<td>9.671</td>
<td>5.196</td>
</tr>
<tr>
<td>Value of β</td>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
<td>0.9</td>
<td>1.0</td>
</tr>
<tr>
<td>N-account System: N = 1</td>
<td>2.794</td>
<td>1.283</td>
<td>0.446</td>
<td>0.067</td>
<td>0.012</td>
</tr>
<tr>
<td>N-account System: N = 2</td>
<td>2.860</td>
<td>1.314</td>
<td>0.458</td>
<td>0.070</td>
<td>0.012</td>
</tr>
<tr>
<td>Relaxed Problem</td>
<td>2.881</td>
<td>1.325</td>
<td>0.462</td>
<td>0.070</td>
<td>0.011</td>
</tr>
<tr>
<td>N-account System: N = 2, π₁ = 0, π₂ = 1</td>
<td>2.542</td>
<td>1.018</td>
<td>0.256</td>
<td>0.015</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Note the the first period of our model is “working life”, so if δR = 1 for one year, than a modelling period comprised of τ years of working life would still have δR = 1.

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).
The first two rows of Table 1 report welfare analyses for $N$-account Systems with, respectively $N = 1$ and $N = 2$ accounts. In $N$-account Systems, the planner sets the penalty-level for each account, $\pi_n$, 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 System to the General Non-linear Mechanism. The fourth row of Table 1 returns to an $N$-account System 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 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 System with $N = 1$ accounts are quantitatively modest.

This analysis also reveals another important feature of the homogeneous-$\beta$ 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 echoes the exact Pigouvian correction that arises in the quasi-linear case (subsection 3.1).\textsuperscript{34}

However, an exact Pigouvian correction (which did arise in the quasi-linear case) is not

\textsuperscript{34}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 $1 - \beta = 0.4$.  

21
Figure 1: The optimal penalty $\pi^*$ and the notional Pigouvian tax $1 - \beta$ as a function of $\beta$ in the case in which: (i) the population has homogeneous $\beta$; (ii) the planner is confined to the $N$-account System with $N = 1$ accounts, with penalty $\pi$. Note that $\pi^*$ is always lower than $1 - \beta$. In particular, $\pi^*$ is negative at $\beta = 1$. This is due to the redistributive motive of the planner: she wishes to redistribute from types with low $\theta$ to types with high $\theta$.

generally socially optimal because, with concave utility, the planner would like to reallocate resources from low-$\theta$ types to high-$\theta$ types. This redistributive motive is reflected in the fact that the socially optimal penalties in the 1-account system (for any given value of $\beta$) 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 $\theta$ values. To transfer resources to these households, the planner lowers the socially optimal penalty below the $(1 - \beta)$ benchmark value. However, as one can see in Figure 1, this downward adjustment is small in magnitude. Accordingly, the Pigouvian correction is the dominant force in these simulations.\footnote{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 \textit{symmetric} information. However, in our...}
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 $\beta$ and study a heterogeneous population of $\beta$-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 $\phi = \theta/\beta$. Accordingly, we study mechanisms in which the agents each report $\phi$ and receive a consumption pair $(c_1, c_2)$ that depends on their report of $\phi$. The social planner’s objective function (1), therefore takes the form

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

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

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 $\phi$. Given heterogeneity in both $\theta$ and $\beta$, taste shock $\theta$ cannot be directly inferred from $\phi$. For a given 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.

The primary assumption is that either $\theta$ or $\beta$ or both have non-atomic distributions. It follows that $\phi$ 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 $\phi$ where the set of optimal choices is non-unique. Because $\phi$ is non-atomic, this set has measure zero. By shifting from truth-telling in $(\theta, \beta)$ to truth-telling in $\phi$, we reduce the feasible set of mechanisms, but we do not change the optimal social welfare.
impatient choice (i.e., high revealed $\phi$), the planner doesn’t know whether the household is making that choice because of high $\theta$ (i.e., real need), or because of low $\beta$ (more present bias). The planner wants to give higher period-1 consumption to a high-$\theta$ household, but not to a low-$\beta$ 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 $\beta$, a planner knows $\beta$ even without observing $\phi$. By contrast, in the heterogeneous-$\beta$ case that we are now studying, the planner has to try to infer $\beta$ and $\theta$ from $\phi$. 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.\footnote{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).}

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

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

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

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

The proof of Proposition 2 is in Appendix D. 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 $\phi$. Heterogeneity in $\beta$ gives rise to a range of marginal taxes needed to implement the optimum. It is only at the extreme values of $\phi$, $\bar{\phi} = \frac{\theta}{\beta}$, and $\underline{\phi} = \frac{\overline{\theta}}{\overline{\beta}}$, that the planner can exactly infer the values of $\beta$, respectively $\bar{\beta}$ and $\underline{\beta}$. Accordingly, at these extreme values for $\phi$, the planner chooses the most extreme Pigouvian tax rates, respectively, $\pi = 1 - \bar{\beta}$ and $\pi = 1 - \underline{\beta}$.38

4.1.1 A case in which the support of $\beta$ is wide relative to the support of $\theta$

We now introduce a corollary that studies the case in which the distribution of $\beta$ is uniform and the support of $\beta$ is wide relative to the support of $\theta$. 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
\[ \frac{\bar{\theta}}{\overline{\beta}} < \frac{\theta}{\underline{\beta}}, \] (8)
in other words that the support of $\beta$ is wide relative to the support of $\theta$. We call this the $\beta$-wide case. We emphasize that $\beta$-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 $(\theta, \beta)$ is $\beta$-wide and (ii) $\beta$ is uniform. Then, $E[\theta|\phi]$ is constant for all $\phi \in [\bar{\theta}/\overline{\beta}, \theta/\underline{\beta}]$. Consequently,

38The assumption that $E[\theta|\phi]$ increases in $\phi$ implies that $c_1(\phi)$ is increasing and hence Proposition 2, which is solved as usual by solving the Relaxed Problem and verifying monotonicity ex post, is in fact the optimal General Nonlinear Mechanism. Intuitively, $\phi$ 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 $\phi$. The assumption that $E[\theta|\phi]$ is increasing implies these two valuations are everywhere positively related, or in other words that $\beta$ does not fall too quickly as $\phi$ rises.
1. The optimal allocation pools all types $\phi \in [\theta/\beta, \theta/\beta]$.

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

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

(9)

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

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 $\beta$-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 class of mechanisms studied – \( N \)-account System or the Relaxed Problem. Within the class of \( N \)-account Systems, 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. Existing evidence suggests \( \beta \) has substantial variation with a mean close to 0.6 (Lockwood, 2020; Laibson et al., Forthcoming), which we capture by assuming that \( \beta \) is uniformly distributed on the interval \([0, 1] \). We show in Section 4.5 that when we impose an account system whose penalties and account allocations are in line with the existing U.S. system, our simulations are able to match the historical leakage rate of approximately 40% (see Argento, Bryant, and Sabelhaus (2015)). 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 System 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 C). The fourth row reports the case of an \( N \)-account System 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 System with \( N = 3 \) accounts, where the planner has to set a completely liquid account (\( \pi_1 = 0 \)) and a completely illiquid account (\( \pi_3 = 1 \)), but can choose the penalty for account 2 (\( \pi_2 \)).

Table 2 reveals that an \( N \)-account System 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 6.145% obtained with
<table>
<thead>
<tr>
<th>System Description</th>
<th>Welfare Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$-account System: $N = 1$</td>
<td>3.569</td>
</tr>
<tr>
<td>$N$-account System: $N = 2$</td>
<td>6.136</td>
</tr>
<tr>
<td>Relaxed Problem</td>
<td>6.145</td>
</tr>
<tr>
<td>$N$-account System: $N = 2$, $\pi_1 = 0$, $\pi_2 = 1$</td>
<td>6.105</td>
</tr>
<tr>
<td>$N$-account System: $N = 3$, $\pi_1 = 0$, $\pi_3 = 1$</td>
<td>6.137</td>
</tr>
</tbody>
</table>

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

In contrast, an $N$-account System 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., \( \pi_2 \) varies). As we vary the penalty \( \pi_2 \) 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 \( \pi_2 \), and the vertical axis shows the average welfare of the cross sections of the population obtained by fixing \( \beta \) 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 completely liquid, \( \pi_2 = 0 \), to completely illiquid, \( \pi_2 = 1 \) is equivalent to an increase of about 30% in their wealth level from the planner’s perspective.

Households with other \( \beta \) values experience increasing and then decreasing welfare as \( \pi_2 \) increases from 0 to 1. However, conditional on \( \beta \), all households experience a rise in expected welfare as \( \pi_2 \) rises from zero. For low-\( \beta \) households, this rise occurs because higher penalties prevent low-\( \beta \) households from overconsuming in period 1. For high-\( \beta \) households, this rise occurs because higher penalties generate larger cross-subsidies from low-\( \beta \) households to high-\( \beta \) 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-\( \beta \) households are net recipients of cross-subsidies because they tend to make smaller early withdrawals and, therefore, pay fewer penalties than low-\( \beta \) households. These differential penalty payments are shown in Figure 3, which reports the gross penalties paid by households with different values of \( \beta \) (again integrating over \( \theta \)). Penalties are hump-shaped, with lower-\( \beta \) households being willing on average to withdraw
Figure 2: The welfare of various $\beta$ cross sections of the population as a function of $\pi_2$ in the case in which: (i) the population has heterogeneous $\beta$; (ii) the planner is confined to an $N$-account System with $N = 2$ accounts, with penalties $\pi_1$ and $\pi_2$ 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 $\pi_2 = 1$. 
Figure 3: The total penalties paid by various $\beta$ cross sections of the population as a function of $\pi_2$ in the case in which: (i) the population has heterogeneous $\beta$; (ii) the planner is confined to an $N$-account System with $N = 2$ accounts, with penalties $\pi_1$ and $\pi_2$, 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.

more and pay more at all penalty levels.

Unlike the welfare of low-$\beta$ households, which rises monotonically as $\pi_2$ rises, the welfare of high-$\beta$ households eventually peaks and thereafter falls with $\pi_2$. This single-peaked property arises because, while initial rises in $\pi_2$ simply result in greater revenue from the early-withdrawal penalties paid by low-$\beta$ households, later rises tend to eliminate early withdrawals altogether. Hence the cross-subsidy to high-$\beta$ households first rises and then falls. By the time $\pi_2$ reaches 1, the cross-subsidy has been completely eliminated, and high-$\beta$ households are now facing a binding constraint (if they have a sufficiently high $\theta$ value) that limits their ability to adjust consumption in period 1, so high-$\beta$ households are slightly worse off on average than they were when $\pi_2$ was 0. On a money-metric basis, the $\beta = 1$ house-
holds 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).\(^{39}\)

Figure 4 shows the welfare of the population as a whole as a function of the early-withdrawal penalty \( \pi_2 \). It confirms that—as one would expect—the enormous welfare gains for low-\( \beta \) households swamp the modest welfare losses for high-\( \beta \) 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.

**Optimal \( \pi_2 \) in the Quasilinear Case.** To further develop intuition, we map the analysis of this section to the quasilinear case of Section 4.1. That is, we study a 2-account system in which one account is completely liquid, the second account has a penalty \( \pi_2 \), and the planner optimizes over the penalty \( \pi_2 \) and the account allocations. In Appendix D.3, we show that the planner’s optimal penalty in the quasilinear case satisfies

\[
E \left[ \left( E[1 - \beta |\phi] - \pi_2 \right) \cdot \frac{\partial c_1(\phi)}{\partial \pi_2} \mid \phi \geq \phi_2 \right] = 0, \tag{10}
\]

where \( \phi_2 \) is the lowest \( \phi \) type that withdraws from the second account in equilibrium. Equation (10) reveals two reasons why the optimal penalty on the second account tends to be large. The first is that the penalty targets a conditional average Pigouvian correction among types that withdraw from the second account, that is the Pigouvian correction is among

\(^{39}\)In Appendix E, we show that this asymmetry in welfare impact from introducing a completely illiquid account is a general property. In particular, starting from autarky, we show that it is always possible to introduce a Pareto efficient completely illiquid account with a positive balance (with weak efficiency for \( \beta = 1 \) types). Welfare gains are larger for lower \( \beta \) types, and increasing the account balance beyond the Pareto efficient level generates second order losses for \( \beta = 1 \) types and first order gains for \( \beta < 1 \) types.
Figure 4: The welfare of the population as a whole as a function of $\pi_2$ in the case in which: (i) the population has heterogeneous $\beta$; (ii) the planner is confined to an $N$-account System with $N = 2$ accounts, with penalties $\pi_1$ and $\pi_2$, 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 $\pi_2^*$ is approximately 85%; (iii) $\pi_2^*$ 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.
types with $\phi \geq \phi_2$. Since higher $\phi = \frac{\theta}{\beta}$ types tend to have low values of $\beta$, this engenders larger penalties on the second account.

The second reason is that the Pigouvian correction is weighted by the responsiveness of type-$\phi$ consumption to variation in the penalty, $\frac{\partial c_1(\phi)}{\partial \pi_2}$. Under constant relative risk aversion, $u_1(c_1) = \frac{c_1^{1-\gamma} - 1}{1-\gamma}$, the consumption response is $\frac{\partial c_1(\phi)}{\partial \pi_2} = -\frac{1}{\gamma(1-\pi_2)}c_1(\phi)$, implying that higher-$\phi$ types, who have higher date 1 consumption levels, also decrease their date 1 consumption more in response to an increase in the early withdrawal penalty $\pi_2$. In summary, because households with relatively high consumption sensitivity to $\pi_2$, tend to be households with relatively low $\beta$ values, the optimal penalty is further raised because the weighting term in the Pigouvian correction effectively puts greater weight on low-$\beta$ households.

Equation (10) also sheds light on Figures 2 and 3. Regarding equation (10), the penalty on the second account targets types that withdraw from it, that is with high $\phi$ (low $\beta$). These types are particularly dynamically inconsistent. These types also tend to have large consumption responses to increases in penalties, which can be seen from the drops in penalties paid by low-$\beta$ types when penalties are increased in the high-penalty region (Figure 3). Putting these observations together, this explains why welfare increases sharply for low-$\beta$ types, who are benefitting from a large correction (consumption response) to a severe overconsumption problem (arising from low $\beta$ values). In contrast, high $\beta$ types tend to have low $\phi$ and do not withdraw from the second account. As a result, variation in the second account penalty has no direct effect on them (equation 10), but changes in penalties paid affects their wealth level. Thus their welfare starts to fall as the penalty on the second account rises far enough and transfers fall. In the quasilinear model, these transfers are a wash from a social welfare perspective, focusing the planner’s objective only on correcting overconsumption from present bias. This helps to understand the result of Figure 4, that high penalties (if not necessarily complete illiquidity) on the second account are desirable.
4.4 A three-account system that has many of the institutional features of the U.S. retirement savings system

The fifth row in Table 2 reports the welfare gains for an $N$-account System 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 ($\pi_3 = 1$). We think of this third account—the illiquid account—as Social Security or a defined-benefit pension. The planner picks the penalty on the “middle” account ($0 < \pi_2 < 1$) and the values of the three endowments ($x_1$, $x_2$, and $x_3$) to optimize social welfare (while satisfying the budget constraint). The “middle” account turns out to have an optimal penalty of $\pi_2 = 0.13$, which is close to the actual penalty associated with a 401(k) or IRA account, namely 0.10. Adding 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 $c_2/c_1$. 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. When a third account is added, it looks and performs somewhat like a U.S. 401(k) plan: the third account has an optimized...
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 $\beta$; (ii) the planner is confined to an $N$-account System with $N = 3$ accounts, with penalties $\pi_1$, $\pi_2$ and $\pi_3$, 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 $\pi_2$ 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 $\pi_2$ on the entire balance of the second account.
penalty of 0.13 and generates a very high rate of leakage in equilibrium. This high leakage rate (90%) is even higher than the empirical leakage rate (40%) observed in the U.S. system.

One explanation for this difference in leakage rates 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 $\beta$ values and lower $\theta$ values in the first place. In this sense, one can’t directly compare the leakage rate in the model (which is the aftermath of universal forced savings in a DC system) and the leakage rate in the US economy (which is the aftermath of voluntary savings in a DC system). Differential selection makes this an apples to oranges comparison.

4.5 Comparison with a Calibrated Model of the US System

Another important factor that explains the high model-predicted leakage rate is the fact that the planner optimally chooses to put 47.4% of each household’s resources into the completely illiquid account, 36.4% of household resources into 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%. This also explains why the presence of the partially illiquid account has such a small effect on total social welfare.

In the United States, the actual allocation to completely illiquid accounts is 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 completely liquid account plays a far more important role in practice than it does in our optimized model reported in Section 4.4. 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 now report an illustrative calibration of the model where we exogenously fix the account balance allocations (rather than endogenously optimizing them) to reflect the operation of the status quo system in the United States.

We base our calibrated model on empirical savings flows for a “typical” household and then translate them into the theoretical objects in our stylized two-period model. For households with median income, average defined contribution (gross) savings flows are 5.0% of income. For households with median income (who are therefore below the Social Security earnings cap), Social Security contributions are 12.4% of income. For households age 60-69 who are within five percentile points of median wealth in that age group, average home equity at retirement is approximately $100,000, which represents an average annual forced savings rate of \[
\frac{100,000}{50 \text{ years of working life}} \approx \frac{100,000}{51,940 \times (1+0.062+0.33 \times 0.05)} = 3.6\%.
\]

We now translate these savings flows from empirical data into stylized savings flows in our two-period model. Assuming that the real rate of (risk free) return is approximately zero, we therefore multiply savings flows by the ratio of the duration of working life to the duration of

\[41\] The National Compensation Survey - Employee Benefits (Table 2, March 2021) reports that 43% (53%) of civilian workers participate in their workplace defined contribution plan in occupations with average wages in the second (third) wage quartile. Averaging these two numbers yields a “middle income” participation rate of 48%. Brady (2017) (Figure A3) uses a representative sample of tax returns and finds that the average contribution rate to DC plans was 6.4% (7.4%) among contributing workers with gross-income between $40,000 and $50,000 ($50,000 and $100,000) in 2013. Averaging these numbers leads to a “middle income” contribution rate of 6.9%. Because this contribution data omits employer contributions, we scale this by a factor of 1.5, producing a total contribution rate of 10.4%. Our participation and contribution rates jointly imply an average contribution rate of 5.0% for workers with median income.

\[42\] This combines the 6.2% employer contribution and 6.2% employee contribution.

\[43\] Authors’ calculation using 2013 SCF for home equity and U.S. Census Bureau Data for median household income. This calculation assumes a risk-free return of zero percent. The denominator is scaled up to include employer Social Security contributions and matches as part of income.
(To complete this argument, we are also using the homotheticity of the CRRA utility class.) We find that illiquid savings translated into the model is $2 \times (12.4\% + 3.6\%) = 32.0\%$. Likewise, we find that total defined contribution savings translated into the model is $2 \times (5.0\%) = 10.0\%$. Finally, to capture the fact that some defined contribution savings can be withdrawn without paying a 10% penalty (e.g., for educational expenses or large medical costs), we assign 25% of the defined contribution savings to a 1% penalty account (reflecting transactions costs) and 75% to the standard 10% penalty account.

With this calibration, we obtain an aggregate leakage rate (total leakage divided by total balances in the two partially illiquid retirement accounts) of 40.7%, which is within the range of historical leakage rates in the United States (see Argento, Bryant, and Sabelhaus 2015). We find that the optimized funding for the completely illiquid asset (47.4%) is greater than the calibrated funding for completely illiquid assets in the U.S economy (32.0%). In our illustrative calibrated model, the (certainty equivalent) welfare gain of moving from the status quo U.S. system to the optimized system is 2.7% of lifetime income.

In summary, our analysis of the existing U.S. retirement savings system has generated two key results. First, the existing types of accounts in the U.S. system are approximately optimal. Second, the allocations to those accounts are not. Specifically, our model implies that a materially higher allocation should be made to the completely illiquid account (47.4% in our second-best optimum vs. 32.0% under existing U.S. policy).

---

44 For the purposes of our model, what matters is the ratio of years of employment (pre-retirement) to years of retirement. For example, consider a worker who starts working life at age 21, is out of work 20% of the time during “working life” (due to unemployment and withdrawal from the labor force), retires at age 65, and lives for 18 years in retirement (which matches life expectancy conditional on living to age 65). For this example, the ratio of years of employment to years of retirement is approximately two.
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 $\beta$ 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 0.082% of income. With respect to the second finding, the optimized penalty on the partially illiquid account ranges from 9% to 16% 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 80% to 99%.45

The detailed results are reported in Tables 3, 4, and 5, 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 C. For case (ii), in addition to the welfare gain, we also report the penalty $\pi_2$ and the leakage rate.46

45We also verify that our baseline results do not change when the support of the truncated $\theta$ distribution, $[1 - \chi, 1 + \chi]$, is expanded by increasing $\chi$ from $\chi = \frac{2}{3}$ to $\chi = \frac{9}{16}$.
46Note that the upper bound on the welfare gain—provided by the Relaxed Problem—is economically close to the N-account Systems that we study.
Table 3 varies the value of the coefficient of relative risk aversion \((\gamma)\) between \(\gamma = \frac{1}{2}\) and \(\gamma = 4\), with \(\gamma = 1\) being our benchmark (for comparison).

Table 4 varies the shape of the density of the taste shock \(\theta\). Table 4a changes the variance of the (untruncated) normal distribution between \(\sigma_\theta = 0.40\) and \(\sigma_\theta = 0.15\), with \(\sigma_\theta = 0.25\) being our benchmark. Table 4b varies the mean \(\mu_\theta\) of the untruncated \(\theta\) distribution while holding the truncation fixed. It varies \(\mu_\theta\) between \(\mu_\theta = 0.8\) (a more right-skewed distribution after truncation) and \(\mu_\theta = 1.2\) (a more left-skewed distribution after truncation), whereas our baseline model with \(\mu_\theta = 1\) has a symmetric truncated distribution. Table 4c repeats the same exercise but moves the truncated support in lockstep with changes in the untruncated mean, which varies the mean of the truncated distribution while maintaining symmetry.

In our benchmark calibration, we studied the case of a uniform distribution of \(\beta\) between 0.2 and 1.0. In Tables 5, we study a truncated normal distribution on \(\beta\) with support between 0.2 and 1.0. Our original benchmark is equivalent to the (truncated) normal case with \(\sigma_\beta = \infty\) and \(\mu_\beta = 0.6\). In Table 5a, we reduce \(\sigma_\beta\) to 1, 1/2 and 0 (holding the truncation points and \(\mu_\beta\) 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 3, results do not generalize to the degenerate case of homogeneous \(\beta\) (the last column of Table 5a), reinforcing that (at least partially unobservable) heterogeneity in \(\beta\) is necessary for a completely illiquid account to be optimal. In Table 5b, we vary \(\mu_\beta\), while holding the support fixed, between 0.45 (a more right-skewed distribution after truncation) and 0.75 (a more left-skewed distribution after truncation).

### 5.1 Observable types

Until now, we have assumed that heterogeneity in \(\beta\) is not directly observable by the planner (and can only be inferred from consumption decisions). However, there is evidence (see Lockwood (2020) for a review) that variation in \(\beta\) is positively correlated with household income. Summarizing the available evidence, households with high income are estimated to
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 System 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 System 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. This table varies the value of the coefficient of relative risk aversion $\gamma$.

<table>
<thead>
<tr>
<th>Value of $\gamma$</th>
<th>0.5</th>
<th>1.0</th>
<th>2.0</th>
<th>4.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$-account System: $N = 2$, $\pi_1 = 0$, $\pi_2 = 1$</td>
<td>8.853</td>
<td>6.106</td>
<td>3.261</td>
<td>1.623</td>
</tr>
<tr>
<td>$N$-account System: $N = 3$, $\pi_1 = 0$, $\pi_3 = 1$</td>
<td>8.920</td>
<td>6.137</td>
<td>3.274</td>
<td>1.633</td>
</tr>
<tr>
<td>Penalty $\pi_2^*$</td>
<td>0.13</td>
<td>0.13</td>
<td>0.11</td>
<td>0.13</td>
</tr>
<tr>
<td>Leakage Rate</td>
<td>0.89</td>
<td>0.90</td>
<td>0.99</td>
<td>0.88</td>
</tr>
<tr>
<td>Variation of the coefficient of relative risk aversion $\gamma$</td>
<td>8.935</td>
<td>6.145</td>
<td>3.279</td>
<td>1.634</td>
</tr>
</tbody>
</table>

The U.S. Social Security system has properties that reflect these qualitative predictions. Workers with very low earnings (i.e., $16,537 in 2023 dollars) make payments and receive benefits that yield an average Social Security retirement income replacement rate of 77.8%.

47Because the income level $Y$ rescales total utility under CRRA, we can leave the exact income levels unspecified.
### Table 4: Robustness checks for welfare gains, optimal penalties and leakage rates.

In each sub-table: row 1 contains welfare gains from an N-account System 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 System 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.

This table varies parameters of the distribution of \( \theta \). Table 4a varies the parameter \( \sigma_\theta \) of the truncated-normal distribution of \( \theta \). Table 4b varies the parameter \( \mu_\theta \) of the truncated-normal distribution of \( \theta \), without changing the truncated support. Table 4c varies the parameter \( \mu_\theta \) of the truncated-normal distribution of \( \theta \), while moving the truncated support to maintain a symmetric truncated distribution.

**Table 4a: Variation of the standard deviation \( \sigma_\theta \) of the taste shock**

<table>
<thead>
<tr>
<th>Value of ( \sigma_\theta )</th>
<th>0.40</th>
<th>0.35</th>
<th>0.30</th>
<th>0.25</th>
<th>0.20</th>
<th>0.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )-account System: ( N = 2, \pi_1 = 0, \pi_2 = 1 )</td>
<td>5.661</td>
<td>5.770</td>
<td>5.919</td>
<td>6.106</td>
<td>6.323</td>
<td>6.531</td>
</tr>
<tr>
<td>( N )-account System: ( N = 3, \pi_1 = 0, \pi_3 = 1 )</td>
<td>5.715</td>
<td>5.819</td>
<td>5.959</td>
<td>6.137</td>
<td>6.344</td>
<td>6.541</td>
</tr>
<tr>
<td>Penalty ( \pi_2^* )</td>
<td>0.16</td>
<td>0.15</td>
<td>0.14</td>
<td>0.13</td>
<td>0.12</td>
<td>0.09</td>
</tr>
<tr>
<td>Leakage Rate</td>
<td>0.80</td>
<td>0.84</td>
<td>0.84</td>
<td>0.90</td>
<td>0.89</td>
<td>0.99</td>
</tr>
<tr>
<td>Relaxed Problem</td>
<td>5.725</td>
<td>5.829</td>
<td>5.968</td>
<td>6.145</td>
<td>6.349</td>
<td>6.544</td>
</tr>
</tbody>
</table>

**Table 4b: Variation of the mean \( \mu_\theta \) of the taste shock holding fixed the support.**

<table>
<thead>
<tr>
<th>Value of ( \mu_\theta ) (Fixed Support)</th>
<th>0.8</th>
<th>1.0</th>
<th>1.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )-account System: ( N = 2, \pi_1 = 0, \pi_2 = 1 )</td>
<td>6.048</td>
<td>6.106</td>
<td>6.103</td>
</tr>
<tr>
<td>( N )-account System: ( N = 3, \pi_1 = 0, \pi_3 = 1 )</td>
<td>6.095</td>
<td>6.137</td>
<td>6.125</td>
</tr>
<tr>
<td>Penalty ( \pi_2^* )</td>
<td>0.15</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td>Leakage Rate</td>
<td>0.83</td>
<td>0.90</td>
<td>0.88</td>
</tr>
<tr>
<td>Relaxed Problem</td>
<td>6.095</td>
<td>6.145</td>
<td>6.136</td>
</tr>
</tbody>
</table>

**Table 4c: Variation of the mean \( \mu_\theta \) of the taste shock, with moving support.**

<table>
<thead>
<tr>
<th>Value of ( \mu_\theta ) (Moving Support)</th>
<th>0.8</th>
<th>1.0</th>
<th>1.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )-account System: ( N = 2, \pi_1 = 0, \pi_2 = 1 )</td>
<td>5.970</td>
<td>6.106</td>
<td>6.052</td>
</tr>
<tr>
<td>( N )-account System: ( N = 3, \pi_1 = 0, \pi_3 = 1 )</td>
<td>6.016</td>
<td>6.137</td>
<td>6.072</td>
</tr>
<tr>
<td>Penalty ( \pi_2^* )</td>
<td>0.15</td>
<td>0.13</td>
<td>0.11</td>
</tr>
<tr>
<td>Leakage Rate</td>
<td>0.82</td>
<td>0.90</td>
<td>0.95</td>
</tr>
<tr>
<td>Relaxed Problem</td>
<td>6.027</td>
<td>6.145</td>
<td>6.078</td>
</tr>
</tbody>
</table>
Table 5: Robustness checks for welfare gains, optimal penalties and leakage rates. In each sub-table: row 1 contains welfare gains from an $N$-account System 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 System 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. This table varies parameters of the distribution of $\beta$ for a truncated normal distribution of $\beta$ on support $[0, 1]$. Table 5a varies the parameter $\sigma_\beta$ of the truncated-normal distribution of $\beta$. Table 5b varies the parameter $\mu_\beta$ of the truncated-normal distribution of $\beta$. 

\begin{tabular}{|c|c|c|c|c|}
\hline
 & Value of $\sigma_\beta$ & & & \\
\hline & $+\infty$ & 1.0 & 0.5 & 0.0 \\
\hline N-account System: $N = 2$, $\pi_1 = 0$, $\pi_2 = 1$ & 6.106 & 6.020 & 5.773 & 2.542 \\
\hline N-account System: $N = 3$, $\pi_1 = 0$, $\pi_3 = 1$ & 6.137 & 6.053 & 5.811 & 2.841 \\
\hline --- --- --- --- & Penalty $\pi_2^*$ & 0.13 & 0.13 & 0.14 & 0.36 \\
\hline --- --- --- --- & Leakage Rate & 0.90 & 0.90 & 0.90 & 0.73 \\
\hline Relaxed Problem & 6.145 & 6.061 & 5.820 & 2.881 \\
\hline \end{tabular}

(a) Variation of the standard deviation $\sigma_\beta$ of the present bias distribution

\begin{tabular}{|c|c|c|c|c|}
\hline
 & Value of $\mu_\beta$ & & & \\
\hline 0.45 & 0.50 & 0.55 & 0.65 & 0.70 & 0.75 \\
\hline N-account System: $N = 2$, $\pi_1 = 0$, $\pi_2 = 1$ & 6.690 & 6.377 & 6.072 & 5.481 & 5.197 & 4.921 \\
\hline N-account System: $N = 3$, $\pi_1 = 0$, $\pi_3 = 1$ & 6.730 & 6.417 & 6.111 & 5.519 & 5.234 & 4.958 \\
\hline --- --- --- --- & Penalty $\pi_2^*$ & 0.15 & 0.15 & 0.15 & 0.14 & 0.14 \\
\hline --- --- --- --- & Leakage Rate & 0.91 & 0.90 & 0.89 & 0.88 & 0.87 & 0.85 \\
\hline Relaxed Problem & 6.741 & 6.427 & 6.120 & 5.527 & 5.242 & 4.966 \\
\hline \end{tabular}

(b) Variation of the mean $\mu_\beta$ of the present bias distribution
for the 1955 birth cohort retiring at normal retirement age. By contrast, workers in the same birth cohort and with same retirement age with earnings at the Social Security earnings cap (e.g., $160,200 in 2023 dollars) have an average income replacement rate of 27.6%).

5.2 Extensions

In the appendix, we consider three extensions of the model, and shed light on how they affect the key insights by studying the optimal mechanism in the quasilinear case with heterogeneous $\beta$ (i.e., the counterpart of Proposition 2).

First, we allow for the social planner to assign welfare weights $\omega(\theta, \beta)$ to household types, rather than assuming a utilitarian objective, in Appendix D.4. We show that the optimal penalties include not only a welfare-weight-adjusted conditional Pigouvian correction, similar to Proposition 2, but also an additive term reflecting the redistributive motive. When welfare weights are on average higher at higher values of $\phi$, the redistribution motives pushes for lower penalties at low $\phi$ and higher penalties at high $\phi$ (and conversely the opposite when welfare weights are lower at higher $\phi$).

Second, we allow for correlation between $\theta$ and $\beta$ in Appendix D.4. The optimal penalty formula is the same as Proposition 2, but the conditional expectation is affected by the correlation structure. Positive correlations between $\theta$ and $\beta$ tend to push for flatter penalties, while negative correlations tend to push for extreme penalties.

Third, in Appendix F we study an extension in which households have unobservable heterogeneity in earnings ability and choose their labor supply, receiving income allocated in the savings system. They then realize their consumption type and make withdrawal choices as in the baseline model. We show that the planner still sets penalties according to a conditional average Pigouvian correction, but upweights the magnitude of penalties for low ability (low income) households and downweights the magnitude of penalties for high ability (high income) households. In other words, the planner allows more liquidity for high income

48See Table D in Burkhalter and Chaplain (2023).
households and requires more illiquidity for low income households. Intuitively, allowing high ability households to consume closer to their preferred (no-correction) allocation encourages labor supply and allows the planner to achieve more income redistribution.\textsuperscript{49}

6 Conclusions and Directions for Future Work

We focus our summary on our benchmark 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 $\beta$ and \textit{no} inter-household transfers) turns out to be approximately optimal in our new setting (with heterogeneous $\beta$ 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 80\% to 99\%. By contrast, when we calibrate the model to match actual \textit{empirical} allocations to the completely illiquid account, the implied equilibrium leakage rate from the partially illiquid account drops to 40\%.

\textsuperscript{49}See Moser and Olea de Souza e Silva (2019) for a related insight on providing incentives via different savings rates by income level, in a model with heterogeneous earnings ability and present bias.
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 similarities, there are important differences between our stylized model and the rich institutional environment in which households make retirement savings decisions. Our theoretical model includes key simplifications. First, we assume a particular conceptual formulation of self-defeating behavior (present bias). Second, we assume only two periods (e.g., working life and retirement). 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. Fourth, we assume that households are naive with respect to their present bias parameter, β. Fifth, our quantitative results study a limited set of distributions of θ and β (see Appendix D.4 for analysis of correlated shocks). Sixth, our quantitative evaluations have assumed a fixed endowment, Y, which is not endogenously responsive to the tax and redistribution system (however, see Appendix F for a variation of the model that includes endogenous income). Seventh, although our quantitative evaluations omit observable type heterogeneity (e.g., high-income households tend to have higher values of β), we show how to conceptually incorporate such observable heterogeneity in subsection 5.1, which turns out to reinforce our findings.

50 In addition, the U.S. system contains some scope for tax arbitrage, which is not present in our model.
52 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).
53 We assume θu(c), but we could have instead assumed u(c − θ).
54 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).
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, raising the normative possibility that mandatory savings might be underutilized in the U.S. However, there is an active debate about both the existence and normative interpretation of the observed distribution of consumption changes for households transitioning from work life to retirement.

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.

References


---

56 In our model, mandatory savings are achieved through a funded system. Our model takes no position on the distinction between funded (e.g., the superannuation scheme in Australia) and unfunded (e.g., U.S. Social Security) mandatory savings systems.
57 See Beshears et al. (2018) for a recent review of the literature on consumption dynamics at and through retirement.
Ambrus, Attila and Georgy Egorov. 2013. “Comment on “Commitment vs. flexibility”.”


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 $\beta$ – 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:

$$c_1 + \frac{c_2}{R} \leq Y \text{ for each household.} \tag{A.1}$$

To simplify notation, we henceforth we set $\delta = 1$, $R = 1$ and $Y = 1$.\footnote{This involves no loss of generality because the utility function can be rescaled.} 

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

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

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

Next, we define the “pooling type” $\theta_1$ to be the minimum $\theta \in (0, \bar{\theta})$ such that

$$\frac{1}{\theta-t} \int_t^\bar{\theta} \Gamma(s) ds \geq 1$$

for all $t \in [\theta, \bar{\theta})$. Notice that $\theta_1$ is well-defined. Indeed, $\theta_1 > 0$ and – if we denote by $\Theta_1$ the set of all $\theta \in (0, \bar{\theta})$ such that $\frac{1}{\theta-t} \int_t^\bar{\theta} \Gamma(s) ds \geq 1$ for all $t \in [\theta, \bar{\theta})$ – then $\Theta_1$ is the non-empty half-open interval $[\theta_1, \bar{\theta})$. However, it is entirely possible that $\theta_1$ is a “hypothetical” type, in the sense that $\theta_1 < \bar{\theta}$.

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 $\theta_1$ would choose freely from the unconstrained budget line. Notice that $\Theta_S$ may be empty, but that $\Theta_P$ 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 $\theta_1 > \bar{\theta}$, then this will be the case if and only if: (i) the most patient of the relevant types, namely $\bar{\theta}$, would choose $c_1 > 0$ from the unconstrained budget line; and (ii) the least patient of the relevant types, namely $\theta_1$, would choose $c_2 > 0$ from the unconstrained budget line. If $\theta_1 \leq \bar{\theta}$, then the only relevant type is the pooling type $\theta_1$, and we need only require that this type chooses both $c_1 > 0$ and $c_2 > 0$ from the unconstrained budget line.

Finally, we need to ensure that the Lagrange multiplier used in the sufficiency argument is non-negative. To that end, we assume that $\Gamma$ is non-decreasing on the separating interval $\Theta_S = [\theta_1, \bar{\theta})$. Notice that, if $\theta_1 \leq \bar{\theta}$, then $\Theta_S$ is empty; so in that case this assumption

\[\theta_p = \max \{ \theta_1, \bar{\theta} \}.\]

It is helpful to compare our Assumption A4 with AWA’s (2006) Assumption A. AWA assume that $\Gamma$ is

\[u'(0+)^- = u_2'(0+) = +\infty.\]
places no restriction on \( \Gamma \).

We now enumerate all of our assumptions.

**A1** \( u_1, u_2 \) are twice continuously differentiable, with \( u'_1, u'_2 > 0 \) and \( u''_1, u''_2 < 0 \).

**A2** \( u'_1(0+) = u'_2(0+) = \infty \).

**A3** \( F' \) is a function of bounded variation.\(^5\)

**A4** \( \Gamma \) 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 \( \beta \) 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 \( \Theta_S \) – 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 \( \Theta_P \) – 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).\(^6\)

In other words, in the case of homogeneous \( \beta \), no inter-household transfers and a weak restriction on the distribution function of the taste shifter \( \theta \), the socially optimal allocation is achieved with an \( N \)-account System 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 \( \theta_1 \) is a hypothetical type that is not a member of the population \( \Theta \). Either way, all types non-decreasing on the interval \([\underline{\theta}, \theta_p]\).

\(^5\)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 A.1.3 and A.1.4 for a detailed discussion of functions of bounded variation.

\(^6\)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.
\( \theta \in \Theta \) with \( \theta \geq \theta_1 \) pool on the choice that type \( \theta_1 \) 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 \( \theta_1 > \theta \), whereas our analysis holds for all values of \( \theta_1 \).

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 Systems (equations 2-6). But the proposition relies on two strong assumptions – homogeneous \( \beta \) and no inter-household transfers. We next analyze the model in the case in which the latter assumption does not hold.

### A.1 Proof of Proposition 4

#### A.1.1 Formulation of the Proposition

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

The first step in formulating Proposition 4 is then to define \( \theta_1 \) in this more general setting. Recalling that the function \( \Gamma \) is given by the formula

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

we define \( \theta_1 \) to be the minimum \( \theta \in (0, \bar{\theta}) \) such that

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

for all \( t \in [\theta, \bar{\theta}) \). Assumptions A1-A5 are then assumed to hold exactly as stated in the main text.

Our proposition generalizes AWA’s analysis in two respects. First, AWA’s analysis covers the case \( \theta_1 > \bar{\theta} \), whereas our analysis holds for all values of \( \theta_1 \). 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. This is potentially important because many incentive-compatible consumption allocations are in fact discontinuous. For example, suppose that there is a type \( \theta_2 \in (\bar{\theta}, \theta) \) and two consumption bundles \( c \) and \( \tilde{c} \) such that all types in \( [\bar{\theta}, \theta_2) \) choose \( c \) and all types in \( (\theta_2, \theta] \) choose \( \tilde{c} \). Then there is a jump in the allocation at \( \theta_2 \).

---

7There 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 allocations, 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 \( \theta_2 \in (\bar{\theta}, \theta) \) and two consumption bundles \( c \) and \( \tilde{c} \) such that all types in \( [\bar{\theta}, \theta_2) \) choose \( c \) and all types in \( (\theta_2, \theta] \) choose \( \tilde{c} \). Then there is a jump in the allocation at \( \theta_2 \).
mal 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.

A.1.2 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 \( \theta \) in the separating interval \( \Theta_S = \{ \theta \mid \theta \in \Theta, \theta < \theta_1 \} = [\theta, \theta_1) \) 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 \( \theta \) in the pooling interval \( \Theta_P = \{ \theta \mid \theta \in \Theta, \theta \geq \theta_1 \} = [\max\{\theta, \theta_1\}, \theta] \) receive the allocation that the (possibly hypothetical) type \( \theta_1 \) would choose freely from the unconstrained budget line.\(^8\)

We transform the candidate consumption allocation \((c_1, c_2) : \Theta \rightarrow (0, \infty)\) into a candidate utility allocation \((r_1, r_2) : \Theta \rightarrow \mathbb{R}\) 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 \( \theta \) with a kink at \( \theta_1 \)), the alternative utility allocations \((v_1, v_2)\) may not even be continuous. We will get over this hurdle by using the one regularity property that incentive-compatible utility allocations do possess: they are monotonic. Hence they are functions of bounded variation.

A.1.3 Functions of Bounded Variation on \( \Theta \)

There are a number of competing definitions of a function of bounded variation. According to one elementary definition, a function \( f : \Theta \rightarrow \mathbb{R} \) is of bounded variation iff it is the difference of two bounded and non-decreasing functions \( f_+, f_- : \Theta \rightarrow \mathbb{R} \). The most serious drawback with this definition for our purposes is that the functions defined in this way do

\[^8\]The consumption allocation will be interior if and only if

\[ \frac{u_2'(Y)}{u_1'(0+)} < \frac{\min\{\beta_1, \beta_2\}}{\beta} \leq \frac{\theta_1}{\beta} < \frac{u_2'(0+)}{u_1'(Y)}. \]

Assumption A2 obviously implies this condition.
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 $\Theta$. 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 $\Theta$.

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

$$f_L(\theta) = f_L(\theta) + f'(\theta, \theta)$$

for all $\theta \in \Theta$ (including $\theta_0$) and the right-hand limits $f_R$ of $f$ by

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

for all $\theta \in \Theta$ (including $\theta_0$). And we endow the set of functions obtained in this way with the norm

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

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

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 $\theta_0$ and $\theta_0$.) A second subtlety is the fact that a function of bounded variation has limits from both the left and right at all points of $\Theta$, including a limit from the left at $\theta_0$ and a limit from the right at $\theta_0$. (This makes perfect sense if one views functions of bounded variation on $\Theta$ as restrictions to $\Theta$ 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.

A.1.4 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 \(\Theta\). It follows, first, that \(F'_L(\theta) = F'_R(\theta) = 0\). It follows, second, that the support of \(F''\) (as a measure) is contained in \(\Theta\). In other words, \(|F''|((0, \theta]) = |F''|((\theta, \infty)) = 0\).

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

\[
  \begin{align*}
  F'_L(\theta) &= F''((0, \theta]), \\
  F'_R(\theta) &= F''((0, \theta]]
  \end{align*}
\]

for all \(\theta > 0\). It then follows that

\[
  0 = F'_R(\bar{\theta}) = F''((0, \bar{\theta}]) = F''((0, \theta]) + F''([\theta, \bar{\theta}])
  = F'_L(\theta) + F''([\theta, \bar{\theta}]) = F''([\theta, \bar{\theta}])
\]

In other words, \(F''\) assigns total mass 0 to \(\Theta\).

### A.1.5 The Lagrangian

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

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

the subset of \(BV(\Theta, \mathbb{R})\) 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 \(\Omega\) the set of utility allocations

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

such that

\[
  \theta v'_1 + \beta v'_2 = 0 \quad (ICL)
\]

and

\[
  v'_1 \geq 0. \quad (ICM)
\]

(The idea here is to split the incentive-compatibility condition into the linear part ICL and the monotonic part ICM.) In other words, let \(\Omega\) be the set of incentive-compatible utility
allocations. Define the objective function

\[ M : \mathcal{BV}(\Theta, \mathbb{R})^2 \rightarrow \mathbb{R} \]

by the formula

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

where \( \ell \) is Lebesgue measure, and define the budget operator

\[ N : O_1 \times O_2 \rightarrow \mathcal{BV}(\Theta, \mathbb{R}) \]

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

\[ L : \Omega \times \mathcal{BV}(\Theta, \mathbb{R})^* \rightarrow \mathbb{R} \]

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 \( \Omega \), and not just on \( \Omega \) 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 \( v_t(0^+) < \min v_t \leq \max v_t < u_t(\infty-) \) and \( C_t \) is continuously differentiable on \( (u_t(0^+), u_t(\infty-)) \). Hence

\[ \| (C_t \circ v_t)' \|_{TV} \leq K \| v_t' \|_{TV}, \]

A.8
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 \( C(\Theta, \mathbb{R})^* \) of the space \( C(\Theta, \mathbb{R}) \) of continuous functions on \( \Theta \) can be represented by the space \( M(\Theta, \mathbb{R}) \) of bounded Borel measures on \( \Theta \). Unfortunately, there does not seem to be a correspondingly tractable representation for the dual \( BV(\Theta, \mathbb{R})^* \) of the space \( BV(\Theta, \mathbb{R}) \) of functions of bounded variation on \( \Theta \). 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.

### A.1.6 A Space of Lagrange Multipliers

One can associate continuous linear functionals in \( BV(\Theta, \mathbb{R})^* \) with bounded Borel measures in \( M(\Theta, \mathbb{R}) \) as follows. Suppose that we are given \( \Lambda \in M(\Theta, \mathbb{R}) \). Then we can construct \( \lambda_R \in BV(\Theta, \mathbb{R})^* \) 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 \( BV(\Theta, \mathbb{R})^* \). It turns out that this subspace is big enough for our purposes.

**Remark 10** By the same token, we can construct \( \lambda_L \in BV(\Theta, \mathbb{R})^* \) 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 \( \lambda_L \) and \( \lambda_R \) 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 \( \Lambda \). (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 \( \Lambda \).)
A.1.7 The Directional Derivative of the Lagrangian

Let us fix $\Lambda \in M(\Theta, \mathbb{R})$ 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 $\Lambda$.

In view of our choice of $\lambda_R$, we have

$$L(v; \lambda_R) = \int (v_1 R + v_2 R) F' \ell(d\theta) + \int (Y - C_1(v_1 R) - C_2(v_2 R)) \Lambda(d\theta).$$

Hence

$$\nabla_s L(r; \lambda_R) = \int (s_1 R + s_2 R) F' \ell(d\theta) - \int (C'_1(r_1 R) s_1 R + C'_2(r_2 R) s_2 R) \Lambda(d\theta).$$

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

$$\int s_2 R F' \ell(d\theta) = \left[s_2 F\right]_0^\theta - \int s_2 F' (d\theta),$$

where:

- $[s_2 F]_0^\theta$ denotes the difference between the right-hand limit of $s_2 F$ at $\theta$ and the left-hand limit of $s_2 F$ at $\theta$;
- $\int s_2 F' (d\theta)$ 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

$$\int F s'_2 (d\theta) = \int F' \frac{\theta}{\beta} s'_1 (d\theta) = -\frac{1}{\beta} \left[s_1 (\theta F)\right]_0^\theta + \frac{1}{\beta} \int s_1 R (\theta F)' \ell(d\theta),$$

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

$$\int (s_1 R + s_2 R) F' \ell(d\theta) = \int \theta s_1 R F' \ell(d\theta) + \int s_2 R F' \ell(d\theta)$$

$$= \int \theta s_1 R F' \ell(d\theta) + \left[s_2 F\right]_0^\theta + \frac{1}{\beta} \left[s_1 (\theta F)\right]_0^\theta - \frac{1}{\beta} \int s_1 R (\theta F)' \ell(d\theta)$$

$$= \left(\frac{\theta}{\beta} s_1 R (\theta) + s_2 R (\theta)\right) F(\theta) - \frac{1}{\beta} \int s_1 R ((1 - \beta) \theta F' + F) \ell(d\theta)$$

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

$$\int C_2'(r_{2R}) s_{2R} \Lambda(d\theta) = \int s_{2R} G'(d\theta)$$
$$= \left[ s_2 G' \right]_{\theta^{-}}^\theta \left[ G' \right]_{\theta^{-}}^\theta + \int G s'_2(d\theta) + \sum_{\theta \in [\theta, \theta]} \Delta s_2 \Delta G,$$

where $\Delta s_2$ and $\Delta G$ denote the jumps in $s_2$ and $G$ at $\theta$ (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

$$\int G s'_2(d\theta) = - \int G \frac{\theta}{\beta} s_1'(d\theta)$$
$$= -\frac{1}{\beta} \left[ s_1(\theta G) \right]_{\theta^{-}}^\theta + \frac{1}{\beta} \int s_{1R}(\theta G)'(d\theta) - \frac{1}{\beta} \sum_{\theta \in [\theta, \theta]} \Delta s_1 \Delta(\theta G)$$
$$= -\frac{1}{\beta} \left[ s_1(\theta G) \right]_{\theta^{-}}^\theta + \frac{1}{\beta} \int s_{1R}(\theta G)'(d\theta) - \frac{1}{\beta} \sum_{\theta \in [\theta, \theta]} \Delta s_1 \theta \Delta G$$

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

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

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

$$\int C'_2(r_{2R}) s_{2R} \Lambda(d\theta) = \left[ s_2 G' \right]_{\theta^{-}}^\theta + \frac{1}{\beta} \left[ s_1(\theta G) \right]_{\theta^{-}}^\theta - \frac{1}{\beta} \int s_{1R}(\theta G)'(d\theta)$$
$$= s_{2R}(\theta G)_R(\theta) + \frac{1}{\beta} s_{1R}(\theta G)_{\theta} G_R(\theta) - \frac{1}{\beta} \int s_{1R}(\theta G'(d\theta) + G \ell(d\theta))$$
$$= \left( \frac{\theta}{\beta} s_{1R}(\theta G') + s_{2R}(\theta G) \right) G_R(\theta) - \frac{1}{\beta} \int s_{1R}(\theta G'(d\theta) + G \ell(d\theta))$$

(where we have used the facts that $G_L(\theta) = 0$ and $(\theta G)'(d\theta) = \theta G'(d\theta) + G \ell(d\theta)$).
Finally, putting all of this information together, we have

\[ \nabla_s L(r; \lambda_R) = \left( \frac{\alpha}{\beta} s_{1R}(\theta) + s_{2R}(\theta) \right) F(\theta) - \frac{1}{\beta} \int s_{1R} \left( (1 - \beta \theta F' + F) \ell(d\theta) \right) \]

\[ - \int \frac{C'_1(r_{1R})}{C'_2(r_{2R})} s_{1R} G'(d\theta) \]

\[ - \left( \frac{\alpha}{\beta} s_{1R}(\theta) + s_{2R}(\theta) \right) G_R(\theta) + \frac{1}{\beta} \int s_{1R} (\theta G'(d\theta) + G \ell(d\theta)) \]

\[ = \left( \frac{\alpha}{\beta} s_{1R}(\theta) + s_{2R}(\theta) \right) (F(\theta) - G_R(\theta)) \]

\[ + \frac{1}{\beta} \int s_{1R} (G - (1 - \beta \theta F' - F) \ell(d\theta) \]

\[ + \frac{1}{\beta} \int s_{1R} \left( \theta - \beta \frac{C'_1(r_{1R})}{C'_2(r_{2R})} \right) G'(d\theta). \]

### A.1.8 A Candidate Lagrange Multiplier

We are now in a position to motivate our choice of Lagrange multiplier \( \Lambda \). 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 \( \Lambda \).

In choosing \( G \), the broad aim is to ensure that \( \nabla_s L(r; \lambda_R) \leq 0 \). 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

\[ \nabla_s L(r; \lambda_R) = \left( \frac{\alpha}{\beta} s_{1R}(\theta) + s_{2R}(\theta) \right) (F(\theta) - G_R(\theta)) \]

\[ + \frac{1}{\beta} \int s_{1R} (G - \Gamma) \ell(d\theta) \]

\[ + \frac{1}{\beta} \int s_{1R} \left( \theta - \beta \frac{C'_1(r_{1R})}{C'_2(r_{2R})} \right) G'(d\theta), \]

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

\[ G_R(\theta) = F(\theta). \]

This will ensure that the first term vanishes.

**Remark 13** At this point we have specified both \( G_L(\theta) \) and \( G_R(\theta) \). It remains to specify \( G \) in the interior of \( \Theta \).
Next, suppose that $\theta_1 > \bar{\theta}$. Then
\[
\frac{C_1'(r_1R)}{C_2'(r_2R)} = \left\{ \begin{array}{ll} \frac{\theta}{\bar{\theta}} & \text{for } \theta \in \Theta_S = [\bar{\theta}, \theta_1) \\
\frac{\theta_1}{\bar{\theta}} & \text{for } \theta \in \Theta_P = [\theta_1, \bar{\theta}] \end{array} \right\}.
\]
Hence the expression for $\nabla_s L(r; \lambda_R)$ simplifies to
\[
\frac{1}{\beta} \int_{\Theta_S \cup \Theta_P} s_1R (G - \Gamma) \ell(d\theta) + \frac{1}{\beta} \int_{\Theta_P} s_1R (\theta - \theta_1) G'(d\theta).
\]
Suppose further that we follow the suggestion of AWA (2006), and put $G = \Gamma$ on $\Theta_S$, where $\Gamma$ is the function defined in Section A.1.1 above. Then the contribution to $\nabla_s L(r; \lambda_R)$ from the separating interval $\Theta_S$ vanishes altogether, and all that is left is the contribution
\[
\frac{1}{\beta} \int_{\Theta_P} s_1R (G - \Gamma) \ell(d\theta) + \frac{1}{\beta} \int_{\Theta_P} s_1R (\theta - \theta_1) G'(d\theta)
\]
to $\nabla_s L(r; \lambda_R)$ from the pooling interval $\Theta_P$. Suppose finally that we follow the suggestion of AWA (2006), and put $G = F(\bar{\theta})$ on $\Theta_S$. Then the measure $G'$ will have an atom of size $F(\bar{\theta}) - \Gamma_L(\theta_1)$ at $\theta_1$, and it will vanish on $(\theta_1, \bar{\theta}]$. Since the term $\theta - \theta_1$ multiplying $G'(d\theta)$ vanishes at $\theta_1$, the second integral itself vanishes, and the first integral reduces to
\[
\frac{1}{\beta} \int_{\Theta_P} s_1R (F(\bar{\theta}) - \Gamma) \ell(d\theta).
\]
Next, suppose that $\theta_1 \leq \bar{\theta}$. In this case, the expression for $\nabla_s L(r; \lambda_R)$ simplifies to
\[
\frac{1}{\beta} \int s_1R (G - \Gamma) \ell(d\theta) + \frac{1}{\beta} \int s_1R (\theta - \theta_1) G'(d\theta).
\]
Suppose further that we follow the suggestion of AWA (2006), and put $G = F(\bar{\theta})$ on the whole of $(\bar{\theta}, \bar{\theta})$. Then the measure $G'$ will have an atom of size $F(\bar{\theta}) - \Gamma_L(\theta_1)$ at $\theta_1$, and it will vanish on $(\bar{\theta}, \bar{\theta}]$. Hence the expression for $\nabla_s L(r; \lambda_R)$ becomes
\[
\frac{1}{\beta} \int s_1R (F(\bar{\theta}) - \Gamma) \ell(d\theta) + \frac{1}{\beta} s_1R(\theta)(\theta - \theta_1) F(\bar{\theta}).
\]
In other words, compared with the case $\theta_1 > \bar{\theta}$, there is an extra term arising from the atom of $G'$ at $\bar{\theta}$.

A.13
Finally, we obtain the Lagrange multiplier $\Lambda$ itself from the formula

$$\Lambda = \frac{1}{C'_2(r_2R)} G'.$$

### A.1.9 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

$$G_L(\underline{\theta}) = 0$$

$$G' = \Gamma \text{ on } (\underline{\theta}, \theta_1)$$

$$G' = F(\overline{\theta}) \text{ on } (\theta_1, \overline{\theta})$$

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

Now, it follows from the formula for $\Gamma$ that

$$G_R(\underline{\theta}) = \Gamma_R(\underline{\theta}) = (1 - \beta) \theta F'_R(\underline{\theta}) + F(\underline{\theta}) = (1 - \beta) \theta F'_R(\underline{\theta}) \geq 0.$$  
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 $\Gamma$ 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 $\theta_1$ as the minimum $\theta \in (0, \overline{\theta})$ such that $\frac{1}{\theta - t} \int_t^\theta \Gamma(s)ds \geq F(\overline{\theta})$ 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

$$G_L(\underline{\theta}) = 0$$

$$G' = F(\overline{\theta}) \text{ on } (\underline{\theta}, \overline{\theta})$$

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

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

A.14
Suppose that $\theta_1 > \bar{\theta}$. Then, in the light of the discussion in Section A.1.8, we have

$$\nabla_s L(r; \lambda_R) = \frac{1}{2} \int_{\Theta_R} s_{1R} \left( F(\bar{\theta}) - \Gamma \right) \ell(d\theta).$$

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

$$H(\theta) = \int_{\theta}^{\bar{\theta}} (\Gamma - F(\bar{\theta})) \ell(d\theta).$$

Then

$$\int_{\Theta_R} s_{1R} \left( F(\bar{\theta}) - \Gamma \right) \ell(d\theta) = \int_{[\theta_1, \bar{\theta}]} s_{1R} \left( F(\bar{\theta}) - \Gamma \right) \ell(d\theta)$$

$$= \int_{[\theta_1, \bar{\theta}]} s_{1R} H' \ell(d\theta)$$

$$= \left[ s_{1} H \right]_{\theta_1}^{\bar{\theta}} - \int_{[\theta_1, \bar{\theta}]} H s_{1}'(d\theta)$$

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

$$\left[ s_{1} H \right]_{\theta_1}^{\bar{\theta}} = s_{1R}(\bar{\theta}) H(\bar{\theta}) - s_{1L}(\theta_1) H(\theta_1)$$

and

$$\int_{[\theta_1, \bar{\theta}]} H s_{1}'(d\theta) = H(\theta_1) \Delta s_{1}(\theta_1) + \int_{(\theta_1, \bar{\theta})} H s_{1}'(d\theta) + H(\bar{\theta}) \Delta s_{1}(\bar{\theta}).$$

Hence, overall,

$$\nabla_s L(r; \lambda_R) = -H(\theta_1) s_{1R}(\theta_1) - \int_{(\theta_1, \bar{\theta})} H s_{1}'(d\theta) + H(\bar{\theta}) s_{1L}(\bar{\theta})$$

$$= -\int_{(\theta_1, \bar{\theta})} H s_{1}'(d\theta)$$

(since $H(\bar{\theta}) = 0$ by construction and $H(\theta_1) = 0$ by definition of $\theta_1$). Now $v_1' \geq 0$ on the whole of $\Theta$, since $v_1$ is non-decreasing, and $r_1' = 0$ on $(\theta_1, \bar{\theta})$, since $r_1$ is constant there. Hence $s_{1}' \geq 0$ on $(\theta_1, \bar{\theta})$. On the other hand, for all $\theta \in [\theta_1, \bar{\theta}]$, we have

$$H(\theta) = \int_{\theta}^{\bar{\theta}} (\Gamma - F(\bar{\theta})) \ell(d\theta) = (\bar{\theta} - \theta) \left( \frac{1}{\bar{\theta} - \theta} \int_{\theta}^{\bar{\theta}} \Gamma \ell(d\theta) - F(\bar{\theta}) \right) \geq 0,$$
by definition of $\theta_1$. 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 $(\theta_1, \bar{\theta})$.

Suppose now that $\theta_1 \leq \underline{\theta}$. Then, in the light of the discussion in Section A.1.8, we have

$$
\nabla_s L(r; \lambda_R) = \frac{1}{\beta} s_{1R}(\theta) (\theta - \theta_1) r(\bar{\theta}) + \frac{1}{\beta} \int_{\theta_1}^{\bar{\theta}} s_{1R}(r(\bar{\theta}) - \Gamma) \ell(d\theta).
$$

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

$$
\int_{\theta_1}^{\bar{\theta}} s_{1R}(r(\bar{\theta}) - \Gamma) \ell(d\theta) = \int_{\theta_1}^{\bar{\theta}} s_{1R}(r(\bar{\theta}) - \Gamma) \ell(d\theta)
$$

$$
= \int_{\theta_1}^{\bar{\theta}} s_{1R} H' \ell(d\theta)
$$

$$
= [s_1 H]^\bar{\theta}_{\underline{\theta}} + \int_{\theta_1}^{\bar{\theta}} H s_1'(d\theta)
$$

$$
= -H(\underline{\theta}) s_{1R}(\underline{\theta}) - \int_{\theta_1}^{\bar{\theta}} H s_1'(d\theta) + H(\bar{\theta}) s_{1L}(\bar{\theta})
$$

$$
= -H(\underline{\theta}) s_{1R}(\underline{\theta}) - \int_{\theta_1}^{\bar{\theta}} H s_1'(d\theta)
$$

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

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

But

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

(by definition of $H$)

$$
= -\int_{\theta_1}^{\bar{\theta}} (\Gamma - F(\bar{\theta})) \ell(d\theta) - \int_{\theta_1}^{\bar{\theta}} (\Gamma - F(\bar{\theta})) \ell(d\theta)
$$

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

$$
= -H(\theta_1)
$$

(by definition of $H$ again)

$$
= 0.
$$

A.16
(by definition of $\theta_1$). Hence
\[
\beta \nabla_s L(r; \lambda_R) = -\int_{(\theta_1, \theta)} H s_1'(d\theta).
\]

Hence, arguing as in the case $\theta_1 > \bar\theta$, $\nabla_s L(r; \lambda_R) \leq 0$.

\section{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 $(\underline{\theta}, \overline{\theta})$.

\textbf{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).

\subsection*{B.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 $\theta$. There will exist $\theta_2 \in (0, \infty)$ such that: if $\theta < \theta_2$, then she consumes less than the balance $x_{\text{liquid}}$ in her liquid account: and, if $\theta \geq \theta_2$, then she consumes the whole of $x_{\text{liquid}}$. The cutoff $\theta_2$ 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 $x_{\text{liquid}}$, 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 $x_{\text{liquid}}$, $c_1$ will be strictly increasing in $\theta$ and $c_2$ will be strictly decreasing in $\theta$.

\footnote{In particular, both the right-hand limit $F'_R(\theta)$ of $F'$ at $\theta$ and the left-hand limit $F'_L(\overline{\theta})$ of $F'$ at $\overline{\theta}$ are strictly positive.}
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 $\theta$ for $\theta < \theta_2$, have a kink at $\theta_2$, 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

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

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

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

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

$$
\theta v_1' + \beta v_2' = 0 \quad (ICL)
$$

and a monotonic part

$$
v_2' \leq 0. \quad (ICM)
$$

**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.

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

Consider first the case in which $x_{\text{liquid}}$ and $x_{\text{illiquid}}$ 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 $v_1'_{IL} : [\underline{\theta}, \theta_2] \to \mathbb{R}$, $v_1'_{IR} : [\theta_2, \overline{\theta}] \to \mathbb{R}$. More precisely, we can put:
1. \( v_1(\theta) = v_1(\bar{\theta}) - \int^\theta_\bar{\theta} v'_1R(t) \, dt \) for \( \theta \in [\theta_2, \bar{\theta}] \);

2. \( v_1(\theta) = v_1(\theta_2) - \int^{\theta_2}_\theta v'_1L(t) \, dt \) for \( \theta \in [\theta, \theta_2] \);

3. \( v'_2R(\theta) = -\frac{\theta}{\beta} v'_1R(\theta) \) for \( \theta \in [\theta_2, \bar{\theta}] \);

4. \( v'_2L(\theta) = -\frac{\theta}{\beta} v'_1L(\theta) \) for \( \theta \in [\bar{\theta}, \theta_2] \);

5. \( v_2(\theta) = v_2(\bar{\theta}) - \int^\theta_{\bar{\theta}} v'_2R(t) \, dt \) for \( \theta \in [\theta_2, \bar{\theta}] \);

6. \( v_2(\theta) = v_2(\theta_2) - \int^{\theta_2}_\theta v'_2L(t) \, dt \) for \( \theta \in [\theta, \theta_2] \).

In other words: \( v_1 \) is the continuous function with continuous derivative \( v'_1L \) on \( [\theta, \theta_2] \), continuous derivative \( v'_1R \) on \( (\theta_2, \bar{\theta}) \) and value \( v_1(\bar{\theta}) \) at \( \bar{\theta} \); and \( v_2 \) is the continuous function with continuous derivative \( v'_2L \) on \( [\bar{\theta}, \theta_2] \), continuous derivative \( v'_2R \) on \( (\theta_2, \bar{\theta}] \) and value \( v_2(\bar{\theta}) \) at \( \bar{\theta} \).

**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) \( \theta_2 \) changes or \( v = (v_1, v_2) \) can be discontinuous.

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

\[
L(v_1(\bar{\theta}), v_2(\bar{\theta}), v'_1L, v'_1R, \lambda, \zeta_L, \zeta_R) = \int (\theta v_1(\theta) + v_2(\theta)) \, dF(\theta) \\
+ \lambda \int (Y - C_1(v_1(\theta)) - C_2(v_2(\theta))) \, dF(\theta) \\
- \int_{[\theta, \theta_2]} v'_2L(\theta) \, d\zeta_L(\theta) \\
- \int_{[\theta_2, \bar{\theta}]} v'_2R(\theta) \, d\zeta_R(\theta), \tag{A.2}
\]

where:

1. the arguments of \( L \) are the parameters \( v_1(\bar{\theta}), v_2(\bar{\theta}), v'_1L \) and \( v'_1R \), and the multipliers \( \lambda, \zeta_L \) and \( \zeta_R \);

2. \( \lambda \) is a scalar (namely the multiplier on the aggregate budget constraint);

3. \( \zeta_L \) is a finite non-negative Borel measure on \([\theta, \theta_2]\) (namely the multiplier associated with the non-positivity constraint on \( v'_2L \));
4. \( \zeta_R \) is a finite non-negative Borel measure on \([\theta_2, \bar{\theta}]\) (namely the multiplier associated with the non-positivity constraint on \(v'_{2R}\));

5. the variables \(v_1, v_2, v'_{2L} \) and \(v'_{2R}\) on the right-hand side are determined by the parameters \(v_1(\bar{\theta}), v_2(\bar{\theta}), v'_{1L}, v'_{1R}\) 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 \((r_1(\bar{\theta}), r_2(\bar{\theta}), r'_{1R}, r'_{1L})\) with the reference utility allocation \((r_1, r_2)\) and parameters \((v_1(\bar{\theta}), v_2(\bar{\theta}), v'_{1L}, v'_{1R})\) with the alternative utility allocation \((v_1, v_2)\) in the obvious way, and take the derivative of the Langrangian at the parameter values \((r_1(\bar{\theta}), r_2(\bar{\theta}), r'_{1L}, r'_{1R})\) in the direction \((s_1(\bar{\theta}), s_2(\bar{\theta}), s'_{1L}, s'_{1R})\), 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 \((v_1(\bar{\theta}), v_2(\bar{\theta}), v'_{1L}, v'_{1R})\). Hence we can simply take the derivative of the RHS at the point \((r_1, r_2, r'_{2L}, r'_{2R})\) in the direction \((s_1, s_2, s'_{2L}, s'_{2R})\) and only then substitute for \((s_1, s_2, s'_{2L}, s'_{2R})\) in terms of \((s_1(\bar{\theta}), s_2(\bar{\theta}), s'_{1L}, s'_{1R})\).

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

\[
0 = \int (\theta s_1 + s_2) dF - \lambda \int (C_1'(r_1) s_1 + C_2'(r_2) s_2) dF - \int_{[\theta_2, \bar{\theta}]} s'_{2L}(\theta) d\zeta_L(\theta) - \int_{[\theta_2, \bar{\theta}]} s'_{2R}(\theta) d\zeta_R(\theta) \tag{A.3}
\]

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

\[
0 = \int (Y - C_1(r_1(\theta)) - C_2(r_2(\theta))) dF(\theta),
\]

\[
0 \geq r'_{2L},
\]

\[
0 \geq r'_{2R}.
\]

Finally, constraint qualification must hold. That is,

\[
0 = \int_{[\theta_2, \bar{\theta}]} r'_{2L}(\theta) d\zeta_L(\theta), \tag{A.4}
\]

\[
0 = \int_{[\theta_2, \bar{\theta}]} r'_{2R}(\theta) d\zeta_R(\theta). \tag{A.5}
\]
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 \(\overline{F}(\theta) = \int_{[\theta]} F(t) \, dt\), and noting that \(\theta F - \overline{F}\) and \(s_1\) are both continuous, we can integrate this contribution by parts to obtain

\[
\int \theta s_1 dF(\theta) = \left[ (\theta F - \overline{F}) s_1 \right]_{\theta_1}^{\theta_2} - \int \left( \theta F - \overline{F} \right) s'_1 d\theta
\]

\[
= (\overline{F}(\overline{\theta}) - \overline{F}(\overline{\theta})) s_1(\overline{\theta}) - \int (\theta F - \overline{F}) s'_1 d\theta
\]

\[
= (\overline{F}(\overline{\theta}) - \overline{F}(\overline{\theta})) s_1(\overline{\theta})
- \int_{[\theta_1, \theta_2]} (\theta F - \overline{F}) s'_{1L} d\theta - \int_{[\theta_2, \theta_1]} (\theta F - \overline{F}) s'_{1R} d\theta.
\]

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

\[
\int s_2 dF(\theta) = [F s_2]_{\theta_2}^{\overline{\theta}} - \int F s'_2 d\theta
\]

\[
= F(\overline{\theta}) s_2(\overline{\theta}) - \int F s'_2 d\theta
\]

\[
= F(\overline{\theta}) s_2(\overline{\theta}) - \int_{[\overline{\theta}, \theta_2]} F s'_{2L} d\theta - \int_{[\theta_2, \overline{\theta}]} F s'_{2R} d\theta
\]

\[
= F(\overline{\theta}) s_2(\overline{\theta}) + \int_{[\overline{\theta}, \theta_2]} F \frac{\partial}{\partial \theta} s'_{1L} d\theta + \int_{[\theta_2, \overline{\theta}]} F \frac{\partial}{\partial \theta} s'_{1R} d\theta.
\]

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

\[
-\lambda \int C'_1(r_1) s_1 dF = -\int s_1 \lambda \Lambda'_1 d\theta
\]

\[
= -[s_1 \lambda \Lambda_1]_{\theta_1}^{\overline{\theta}} + \int \lambda \Lambda_1 s'_1 d\theta
\]

\[
= -s_1(\overline{\theta}) \lambda \Lambda_1(\overline{\theta}) + \int \lambda \Lambda_1 s'_1 d\theta
\]

\[
= -s_1(\overline{\theta}) \lambda \Lambda_1(\overline{\theta})
+ \int_{[\theta_1, \theta_2]} \lambda \Lambda_1 s'_{1L} d\theta + \int_{[\theta_2, \overline{\theta}]} \lambda \Lambda_1 s'_{1R} d\theta.
\]
Similarly, putting \( \Lambda_2(\theta) = \int_{[\theta, \theta]} C'_2(r_2(t)) \, dF(t) \),

\[
-\lambda \int C'_2(r_2) \, s_2 \, dF = -\int s_2 \, \lambda \, \Lambda'_2 \, d\theta \\
= -[s_2 \, \lambda \, \Lambda_2]_{\theta_-}^\theta + \int \lambda \, \Lambda_2 \, s'_2 \, d\theta \\
= -s_2(\overline{\theta}) \, \lambda \, \Lambda_2(\overline{\theta}) + \int \lambda \, \Lambda_2 \, s'_2 \, d\theta \\
= -s_2(\overline{\theta}) \, \lambda \, \Lambda_2(\overline{\theta}) \\
+ \int_{[\theta, \theta_2]} \lambda \, \Lambda_2 \, s'_{2L} \, d\theta + \int_{[\theta_2, \overline{\theta}]} \lambda \, \Lambda_2 \, s'_{2R} \, d\theta \\
= -s_2(\overline{\theta}) \, \lambda \, \Lambda_2(\overline{\theta}) \\
- \int_{[\theta, \theta_2]} \lambda \, \Lambda_2 \, \frac{\theta}{3} \, s'_{1L} \, d\theta - \int_{[\theta_2, \overline{\theta}]} \lambda \, \Lambda_2 \, \frac{\theta}{3} \, s'_{1R} \, d\theta.
\]

Finally, we have

\[
-\int_{[\theta, \theta_2]} s'_{2L}(\theta) \, d\zeta_L(\theta) = \int_{[\theta, \theta_2]} \frac{\theta}{3} \, s'_{1L}(\theta) \, d\zeta_L(\theta)
\]

and

\[
-\int_{[\theta_2, \overline{\theta}]} s'_{2R}(\theta) \, d\zeta_R(\theta) = \int_{[\theta_2, \overline{\theta}]} \frac{\theta}{3} \, s'_{1R}(\theta) \, d\zeta_R(\theta).
\]

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:

\[
0 = \overline{\theta} \, F(\overline{\theta}) - \overline{\theta} \, F(\overline{\theta}) - \lambda \, \Lambda_1(\overline{\theta}), \quad (A.6) \\
0 = F(\overline{\theta}) - \lambda \, \Lambda_2(\overline{\theta}), \quad (A.7) \\
0 = - (\theta \, F - \overline{F}) \, d\theta + \frac{\theta}{3} \, F \, d\theta + \lambda \, \Lambda_1 \, d\theta - \frac{\theta}{3} \, \lambda \, \Lambda_2 \, d\theta + \frac{\theta}{3} \, d\zeta_L, \quad (A.8) \\
0 = - (\theta \, F - \overline{F}) \, d\theta + \frac{\theta}{3} \, F \, d\theta + \lambda \, \Lambda_1 \, d\theta - \frac{\theta}{3} \, \lambda \, \Lambda_2 \, d\theta + \frac{\theta}{3} \, d\zeta_R. \quad (A.9)
\]

Now, we certainly have \( r'_{2L} < 0 \) on \([\theta, \theta_2] \). (This is because, if \( \theta < \theta_2 \), then self 1 consumes less than \( x_{\text{liquid}} \). Hence \( r'_{1L} > 0 \) and \( r'_{2L} < 0 \).) It therefore follows from constraint qualification (namely (A.4)) that \( \zeta_L = 0 \). Equation (A.8) therefore implies that

\[
\lambda \, (\theta \, \Lambda_2 - \beta \, \Lambda_1) = \theta \, F - \beta \, (\theta \, F - \overline{F}) = (1 - \beta) \, \theta \, F + \beta \, \overline{F} = \Gamma \quad (A.10)
\]

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

$$\frac{\theta \Lambda_2(\theta)}{\theta - \theta} \rightarrow \frac{\theta C_2'(r_2(\theta))}{\beta} F'(\theta^+),$$

$$\frac{\beta \Lambda_1(\theta)}{\theta - \theta} \rightarrow \frac{\beta C_1'(r_1(\theta))}{\beta} F'(\theta^+),$$

$$\frac{\Gamma}{\theta - \theta} \rightarrow \Gamma(\theta^+) = (1 - \beta) \frac{\theta F'(\theta^+)}{\beta}$$

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

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

We therefore have

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

as $\theta \downarrow \theta$. On the other hand,

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

as $\theta \downarrow \theta$. Passing to the limit in equation (A.10), we therefore obtain

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

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

**B.3 The Case $\theta_2 \in (\bar{\theta}, \infty)$**

Consider now the case in which $x_{\text{liquid}}$ and $x_{\text{illiquid}}$ are such that $\theta_2 \in [\bar{\theta}, \infty)$. In this case, we can derive equations (A.6, A.7 and A.8) exactly as in Section B.2 above. In particular, we can still derive equation (A.8). We can therefore derive a contradiction by essentially the same argument.

**B.4 The Case $\theta_2 \in (0, \theta]$**

Consider now the case in which $x_{\text{liquid}}$ and $x_{\text{illiquid}}$ are such that $\theta_2 \in (0, \theta]$. In this case, we can still derive equations (A.6, A.7 and A.9). However, we can no longer derive equation (A.8). We therefore need new arguments. The first point to note is that, since $\theta_2 \leq \theta$, all
types $\theta \in [\underline{\theta}, \bar{\theta}]$ choose the point that a hypothetical $\theta_2$ type would choose from the ambient budget set. We therefore have

$$
\Lambda_1(\bar{\theta}) = \int_{[\underline{\theta}, \bar{\theta}]} C_1'(r_1(t)) \, dF(t) = F(\bar{\theta}) C_1'(r_1(\theta_2)), \quad (A.11)
$$

$$
\Lambda_2(\bar{\theta}) = \int_{[\underline{\theta}, \bar{\theta}]} C_2'(r_2(t)) \, dF(t) = F(\bar{\theta}) C_2'(r_2(\theta_2)), \quad (A.12)
$$

Furthermore, since the $\theta_2$ 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 (A.6) and (A.7), we therefore obtain

$$
\frac{\bar{\theta} F(\bar{\theta}) - \bar{\theta} F(\bar{\theta})}{F(\bar{\theta})} = \frac{\Lambda_1(\bar{\theta})}{\Lambda_2(\bar{\theta})} = \frac{C_1'(r_1(\theta_2))}{C_2'(r_2(\theta_2))} = \frac{\theta_2}{\beta}. \quad (A.13)
$$

Hence

$$
(\bar{\theta} - \theta_2) F(\bar{\theta}) = \bar{\theta} F(\bar{\theta}) - \beta \left( \bar{\theta} F(\bar{\theta}) - \bar{\theta} F(\bar{\theta}) \right)
$$

$$
= (1 - \beta) \bar{\theta} F(\bar{\theta}) + \beta \bar{\theta} F(\bar{\theta})
$$

$$
= \bar{\Gamma}(\bar{\theta}), \quad (A.14)
$$

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

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

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

or

$$
\frac{1}{\theta - \theta_2} \int_{[\theta_2, \bar{\theta}]} \Gamma(t) \, dt = F(\bar{\theta}). \quad (A.15)
$$

The significance of this observation is that $\theta_1$ satisfies equation (A.15) 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 (A.9). It follows from this equation that

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

In other words, \( \zeta_R \) is absolutely continuous w.r.t. Lebesgue measure, with density

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

Furthermore:

\[
\Lambda_1(\theta) = \int_{[\theta, \overline{\theta}]} C_1'(r_1(t)) dF(t) = F(\theta) C_1'(r_1(\theta_2)) = \frac{F(\theta)}{F(\overline{\theta})} \Lambda_1(\overline{\theta})
\]

\[
= \frac{F(\theta)}{F(\overline{\theta})} \frac{\theta_2}{\beta} \Lambda_2(\overline{\theta}) = \frac{F(\theta)}{F(\overline{\theta})} \frac{\theta_2}{\beta} \frac{F(\overline{\theta})}{\lambda} = \frac{\theta_2}{\lambda} \frac{F(\theta)}{\lambda}
\]

(where the last line follows from (A.13) and (A.7)); and

\[
\Lambda_2(\theta) = \int_{[\theta, \overline{\theta}]} C_2'(r_2(t)) dF(t) = F(\theta) C_2'(r_2(\theta_2)) = \frac{F(\theta)}{F(\overline{\theta})} \Lambda_2(\overline{\theta})
\]

\[
= \frac{F(\theta)}{F(\overline{\theta})} \Lambda_2(\overline{\theta}) = \frac{F(\theta)}{F(\overline{\theta})} \Lambda_2(\overline{\theta})
\]

(where the last line follows from (A.7)). Hence

\[ \lambda (\theta \Lambda_2 - \beta \Lambda_1) = (\theta - \theta_2) F(\theta) \]

and

\[ \theta \zeta'_R = \beta (\theta F - F) - \theta F + (\theta - \theta_2) F = (\theta - \theta_2) F(\theta) - \Gamma. \]

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

\[ \frac{\theta \zeta'_R(\theta) - \theta \zeta'_R(\overline{\theta})}{\theta - \overline{\theta}} \geq 0. \]

Letting \( \theta \to \overline{\theta}^+ \), we therefore obtain

\[ (\theta \zeta'_R)'(\overline{\theta}^+) = (\beta \overline{\theta} - \theta_2) F'(\overline{\theta}^+) \geq 0. \]
Since \( F'(\theta+) > 0 \), it follows that
\[
\theta_2 \leq \beta \bar{\theta}.
\]
(A.16)

Similarly, (A.14) implies that \((\bar{\theta} - \theta_2) F(\bar{\theta}) - \Gamma(\bar{\theta}) = 0\). Hence \(\bar{\theta} \zeta'_R(\bar{\theta}) = 0\). Hence

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

Letting \(\theta \to \bar{\theta}^-\), we therefore obtain

\[
(\theta \zeta'_R(\bar{\theta}^-) = (\beta \bar{\theta} - \theta_2) F'(\bar{\theta}^-) \leq 0.
\]

Since \(F'(\bar{\theta}^-) > 0\), it follows that
\[
\theta_2 \geq \beta \bar{\theta}.
\]
(A.17)

But inequalities (A.16) and (A.17) 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 \(\zeta_R\) in such a way that this constraint is respected. Then, proceeding almost exactly as above, we will obtain

\[
(\theta \zeta'_R)^'(\bar{\theta}^-) = (\beta \bar{\theta} - \theta_2) F'(\bar{\theta}^-) \leq 0.
\]

Moreover we will have the boundary conditions \(\theta \zeta'_R(\theta) = 0\) and \(\bar{\theta} \zeta'_R(\bar{\theta}) = 0\). It follows that \(\theta_2 \in (\beta \bar{\theta}, \beta \bar{\theta})\) and \(\theta \zeta'_R < 0\) on \((\theta, \bar{\theta})\). Hence a small change in the direction of any incentive-compatible and fully separating mechanism is desirable. (This would have the effect of reducing \(r'_2\) from 0 – and increasing \(r'_1\) from 0 – at all points in the range \((\theta, \bar{\theta})\).) In other words, it is always desirable to allow some flexibility to the decision maker to respond to the information contained in \(\theta\).
C 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 $\beta$. The case of homogeneous $\beta$ is a simpler variant of the case studied in this section.

C.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

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

subject to the resource constraint

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

and the incentive-compatibility constraint

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

Here, $f$ is the density of $\theta$ (associated with distribution function $F$ in the main text); $g$ is the density of $\beta$ (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 $\Theta$; $g$ 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 $\mu$ and variance $\sigma^2$ truncated to the interval $[\underline{\theta}, \overline{\theta}]$; and $g$ might take the form
\[
g(\beta) = \frac{1}{\beta - \beta} \quad \text{for } \beta \in [\underline{\beta}, \overline{\beta}]
\]
and $g(\beta) = 0$ otherwise, i.e., $g$ might be the density of the uniform distribution on the interval $[\underline{\beta}, \overline{\beta}]$.

C.2 Transforming the Problem

The first step in solving this problem is to note that
\[
(c_1, c_2) \in \arg\max_{(c_1, c_2) \in C} \{\theta u_1(\tilde{c}_1) + \beta u_2(\tilde{c}_2)\}
\]
iff
\[
(c_1, c_2) \in \arg\max_{(\tilde{c}_1, \tilde{c}_2) \in C} \left\{\frac{\theta}{\beta} u_1(\tilde{c}_1) + u_2(\tilde{c}_2)\right\}.
\]
The set of optimal choices of the individual therefore depends only on $\phi = \theta / \beta$. Combining this fact with the assumed continuity of the distribution functions $F$ and $G$ of $\theta$ and $\beta$ implies that, if we put $\Phi = [\underline{\phi}, \overline{\phi}]$ where $\underline{\phi} = \underline{\theta} / \overline{\beta}$ and $\overline{\phi} = \overline{\theta} / \overline{\beta}$, 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 $(\theta, \beta)$ with $(\phi, \beta)$.

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,

\[ 0 = \phi v'_1(\phi) + v'_2(\phi) \]  \hspace{1cm} \text{(ICL)}

and a monotonic part,

\[ 0 \leq -v'_2(\phi). \]  \hspace{1cm} \text{(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 \( \phi \) and the conditional density \( j \) of \( \beta \) given \( \phi \), namely

\[ h(\phi) = \int \beta f(\beta \phi) g(\beta) \, d\beta \]  \hspace{1cm} \text{(A.18)}

and

\[ j(\beta \mid \phi) = \frac{\beta f(\beta \phi) g(\beta)}{h(\phi)}. \]  \hspace{1cm} \text{(A.19)}

We can also introduce the conditional expectation of \( \beta \), namely

\[ b(\phi) = \int \beta j(\beta \mid \phi) \, d\beta. \]  \hspace{1cm} \text{(A.20)}

**Remark 23** The limits of integration in the definition of \( h \) (namely (A.18)) 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 \( \max\{\beta, \theta / \phi\} \) and \( \min\{\beta, \theta / \phi\} \}. \) In particular, the support of the conditional distribution of \( \beta \) varies with \( \phi \):

1. For \( \phi \in [\phi, \min\{\theta / \beta, \theta / \beta\}] \}, \) the support of \( \beta \) is \( [\theta / \phi, \beta] \}. \) In other words: the range of \( \beta \) types that is consistent with \( \phi \) is increasing in \( \phi \), and this range always includes \( \beta \}. \) By the same token, the range of \( \theta \) types that is consistent with \( \phi \) is increasing in \( \phi \), and this range always includes \( \theta \}.

2. For \( \phi \in [\max\{\theta / \beta, \theta / \beta\}, \phi] \}, \) the support of \( \beta \) is \( [\beta, \theta / \phi] \}. \) In other words: the range of \( \beta \) types that is consistent with \( \phi \) is decreasing in \( \phi \), and this range always includes \( \beta \).
3. If $\frac{\theta}{\beta} < \frac{\bar{\theta}}{\bar{\beta}}$ then, for $\phi \in [\min\{\frac{\theta}{\beta}, \frac{\bar{\theta}}{\bar{\beta}}\}, \max\{\frac{\theta}{\beta}, \frac{\bar{\theta}}{\bar{\beta}}\}]$, the support of $\beta$ is $[\beta, \bar{\beta}]$. In other words, if the range of $\theta$ types is large relative to the range of $\beta$ types, then all $\beta$ types are consistent with intermediate values of $\phi$.

4. If $\frac{\theta}{\beta} > \frac{\bar{\theta}}{\bar{\beta}}$ then, for $\phi \in [\min\{\frac{\theta}{\beta}, \frac{\bar{\theta}}{\bar{\beta}}\}, \max\{\frac{\theta}{\beta}, \frac{\bar{\theta}}{\bar{\beta}}\}]$, the support of $\beta$ is $[\frac{\theta}{\phi}, \frac{\bar{\theta}}{\phi}]$. In other words, if the range of $\theta$ types is small relative to the range of $\beta$ types, then there is no value of $\phi$ for which all $\beta$ types are consistent with that value.

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

$$\int \left( b(\phi) \phi v_1(\phi) + v_2(\phi) \right) h(\phi) d\phi$$ \hspace{1cm} (W)

and

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

$$\int \left( b(\phi) \phi v_1(\phi) + v_2(\phi) \right) h(\phi) d\phi$$

$$+ \lambda \int \left( Y - C_1(v_1(\phi)) - \frac{1}{R} C_2(v_2(\phi)) \right) h(\phi) d\phi$$

$$- \int (\phi v'_1(\phi) + v'_2(\phi)) \mu(\phi) h(\phi) d\phi$$

$$- \int v'_2(\phi) \nu(\phi) h(\phi) d\phi,$$

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 $\nu : \Phi \to \mathbb{R}$.

C.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

\[- \int (\phi v_1' + v_2') \mu h d\phi = - \int ((\phi v_1)' - v_1 + v_2') \mu h d\phi = \int v_1 \mu h d\phi - \int ((\phi v_1)' + v_2') \mu h d\phi,\]

where we have dropped the dependence of \(v_1, v_2, \mu\) and \(h\) on \(\phi\). Moreover

\[- \int ((\phi v_1)' + v_2') \mu h d\phi = - \left[ \frac{\phi v_1 + v_2}{\phi} \right] + \int ((\phi v_1)' + v_2') (\mu h)' d\phi = \int ((\phi v_1) + v_2') (\mu h)' d\phi,\]

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

\[- \int v_2' \nu h d\phi = - \left[ \frac{\nu h}{\phi} \right] + \int v_2 (\nu h)' d\phi = \int v_2 (\nu h)' d\phi.\]

The Langrangian can therefore be written

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

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

\[0 = (b \phi + \mu - \lambda C_1'(v_1)) h + \phi (\mu h)'\]

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

\[0 \leq \int \left( Y - C_1(v_1) - \frac{1}{R} C_2(v_2) \right) h d\phi \}

\[0 \leq \lambda\]
and the complementary slackness condition associated with (IC2), namely

\[
\begin{align*}
0 & \leq -v'_2 \\
0 & \leq \nu
\end{align*}
\]

C.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 \(\nu\) from the equations and tackle the three differential equations

\[
\begin{align*}
0 &= \left(b \phi + \mu - \lambda C'_1(v_1)\right) h + \phi (\mu h)', \quad (A.21) \\
0 &= \left(1 - \lambda \frac{1}{R} C'_2(v_2)\right) h + (\mu h)', \quad (A.22) \\
0 &= \phi v'_1 + v'_2 \quad (A.23)
\end{align*}
\]

and the integral equation

\[
0 = \int \left(Y - C_1(v_1) - \frac{1}{R} C_2(v_2)\right) h d\phi. \quad (A.24)
\]

The first step is to make \(v_1\) and \(v_2\) the subjects of equations (A.21) and (A.22). Putting \(U_1 = (C'_1)^{-1}\) and \(U_2 = (C'_2)^{-1}\), we obtain

\[
\begin{align*}
v_1 &= U_1\left(\frac{a_1}{\lambda}\right), \quad (A.25) \\
v_2 &= U_2\left(\frac{a_2}{\lambda}\right), \quad (A.26)
\end{align*}
\]

where

\[
\begin{align*}
a_1 &= b \phi + \mu + \phi (\mu h)', \quad (A.27) \\
a_2 &= R \left(1 + \frac{(\mu h)'}{h}\right). \quad (A.28)
\end{align*}
\]
C.5 Solving (A.21-A.23) where \( b \) and \( h \) are Smooth

Consider the equations (A.21-A.23) in the open region \( \Phi = \Phi \setminus \{ \phi, \theta / \beta, \bar{\theta} / \bar{\beta}, \bar{\phi} \} \). In this region, both \( b \) and \( h \) are smooth. Hence we may differentiate (A.25,A.26) to obtain

\[
\begin{align*}
v_1' &= U_1' \left( \frac{a_1}{\lambda} \right) \frac{a_1'}{\lambda}, \\
v_2' &= U_2' \left( \frac{a_2}{\lambda} \right) \frac{a_2'}{\lambda}
\end{align*}
\]

and, substituting (A.29,A.30) in (A.23),

\[0 = \phi U_1' \left( \frac{a_1}{\lambda} \right) \frac{a_1'}{\lambda} + U_2' \left( \frac{a_2}{\lambda} \right) \frac{a_2'}{\lambda}.
\]

Next, provided that \( u_1 \) and \( u_2 \) have the same coefficient of relative risk aversion \( \gamma \), the latter equation is homogeneous in \( \lambda \). 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 \( \gamma \), 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 \( \mu'' \), \( \mu' \) and \( \mu \), we obtain

\[
0 = (\phi^2 U_1'(a_1) + RU_2'(a_2)) h^2 \mu'' + (\phi (\phi h' + 2 h) U_1'(a_1) + h' RU_2'(a_2)) h \mu' + (\phi (h (h'' + h') - \phi h'^2) U_1'(a_1) + (h h'' - h'^2) RU_2'(a_2)) \mu + \phi (\phi b' + b) U_1'(a_1) h^2. \tag{A.31}
\]

In other words, in the region \( \Phi \), equations (A.21-A.23) reduce to a second-order ordinary differential equation for \( \mu \).

C.6 Solving (A.21-A.23) where \( b \) and \( h \) have Kinks

Now consider the equations (A.21-A.23) at the points \( \phi_1 = \frac{\theta}{\beta} \) and \( \phi_2 = \frac{\bar{\theta}}{\bar{\beta}} \), where both \( b \) and \( h \) have kinks. We cannot differentiate (A.25,A.26) at these points. However, we do have

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

A.33
where

\[ a_1(\phi_i^+) = b \phi_i + \mu(\phi_i^+) + \frac{\phi (\mu h)'(\phi_i^+)}{h(\phi_i)}, \]

\[ a_1(\phi_i^-) = b \phi_i + \mu(\phi_i^-) + \frac{\phi (\mu h)'(\phi_i^-)}{h(\phi_i)}, \]

\[ a_2(\phi_i^+) = R 1 + \frac{(\mu h)'(\phi_i^+)}{h(\phi_i)}, \]

\[ a_2(\phi_i^-) = R 1 + \frac{(\mu h)'(\phi_i^-)}{h(\phi_i)}. \]

Hence, at \( \phi_i \), we can impose the value-matching condition

\[ 0 = \Delta \mu(\phi_i) = \mu(\phi_i^+) - \mu(\phi_i^-) \quad (A.32) \]

and the incentive condition

\[ 0 = \phi_i (U_1(a_1(\phi_i^+)) - U_1(a_1(\phi_i^-))) + (U_2(a_2(\phi_i^+)) - U_2(a_2(\phi_i^-))). \quad (A.33) \]

### C.7 Solving (A.21-A.23) at the Endpoints

Assuming for concreteness that \( \phi_1 < \phi_2 \), we now have the second-order ordinary differential equation (A.31) in the three open intervals \((\phi, \phi_1), (\phi_1, \phi_2)\) and \((\phi_2, \phi)\). Moreover, we have two boundary conditions at each of \( \phi_1 \) and \( \phi_2 \). (Cf. (A.32) and (A.33).) The obvious way of completing the equation would therefore be to require that \( \mu \) take on appropriate values at the boundaries \( \phi \) and \( \phi \). However, \( h \) decays linearly to 0 at both \( \phi \) and \( \phi \). Moreover, inspection of (A.31) shows that:

1. the coefficient of \( \mu'' \) is positive and of order \( h^2 \) near \( \phi \) and \( \phi \);
2. the coefficient of \( \mu' \) is positive and of order \( h \) near \( \phi \), and negative and of order \( h \) near \( \phi \);
3. the coefficient of \( \mu \) is negative and of order 1 near \( \phi \) and \( \phi \).

Hence \( \mu \) will not take on boundary values at \( \phi \) and \( \phi \) in the usual way.\(^{10}\) On the other hand, the inhomogeneous term, namely

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

\(^{10}\)Intuitively speaking, the dynamics of \( \phi \) move away from the endpoints \( \phi \) and \( \phi \).
is of order $h^2$ near $\phi$ and $\bar{\phi}$. In particular, it is bounded. Hence the relevant solution of the equation is the one that is bounded near $\phi$ and $\bar{\phi}$.

C.8 Solving for $\lambda$

As we have seen, we can find $\mu$ by solving the second-order o.d.e. (A.31) with the required boundary conditions at the internal boundaries $\phi_1$ and $\phi_2$ and the required boundedness properties at the endpoints $\phi$ and $\bar{\phi}$. Like $b$ and $h$, $\mu$ can be expected to have kinks at $\phi_1$ and $\phi_2$. The next step is to solve for $\lambda$. This can be done using the resource equation (A.24).

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

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

Hence, substituting in (A.24),

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

or

$$\lambda^{\frac{1}{\gamma}} = \frac{\int \left(a_1^{\frac{1}{\gamma}} + \frac{1}{R} a_2^{\frac{1}{\gamma}}\right) h \, d\phi}{\int Y h \, d\phi}. \quad (A.34)$$

Bearing in mind that $a_1$ and $a_2$ are given in terms of $\mu$ by equations (A.27) and (A.28), this gives us a formula for $\lambda$ in terms of $\mu$.

C.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 $\lambda$ and $\mu$ by (A.25) and (A.26); and $c_1$ and $c_2$ are given in terms of $v_1$ and $v_2$ by the formulae $c_1 = C_1(v_1)$ and $c_2 = C_2(v_2)$.

C.10 Numerical implementation

We generate a numerical solution (using Matlab’s bvp4c function\footnote{See https://www.mathworks.com/help/matlab/ref/bvp4c.html}) for the second-order differential equation for $\mu$ (equation A.31) with the boundary conditions described in section C.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 (A.34), is a definite integral. Its value can be accurately obtained by adding an appropriate expression to the system of differential equations. Let:

$$
Num\lambda(\phi) = \int_{\phi}^{\phi} \left( a_1(x)^{\frac{1}{\gamma}} + \frac{1}{R} a_2(x)^{\frac{1}{\gamma}} \right) h(x) dx
$$

$$
\frac{\partial Num\lambda(\phi)}{\partial \phi} = \left( a_1(\phi)^{\frac{1}{\gamma}} + \frac{1}{R} a_2(\phi)^{\frac{1}{\gamma}} \right) h(\phi)
$$

(A.35)

Going by these definitions, we are interested in calculating $Num\lambda(\phi)$, which is exactly the terminal condition that one obtains when solving the o.d.e. given by (A.35). The boundary conditions for $Num\lambda(\phi)$ are straightforward to obtain and are given by:

$$
Num\lambda(\phi) = 0
$$

$$
Num\lambda(\phi_i+) - Num\lambda(\phi_i-) = 0
$$

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 $\lambda$ and $\lambda$ 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 (A.25), (A.26), $U_1$, and $U_2$ into the definition of $W^{opt}$ and $\lambda$ has been factorized out of the RHS:

$$
W^{opt} \cdot \lambda^{\frac{1}{\gamma}} + \frac{\lambda^{\frac{1}{\gamma}}}{1 - \gamma} \left( \int \phi b(\phi) h(\phi) d\phi + 1 \right) = \\
\int \left( b(\phi)\phi \left( a_1(\phi) \right)^{\frac{1}{\gamma}} + \left( a_2(\phi) \right)^{\frac{1}{\gamma}} \right) h(\phi) d\phi_0
$$

(A.36)

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 (A.36) to recover $W^{opt}$. 

A.36
C.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 C.3, where one cannot use the result that $h(\bar{\phi}) = h(\phi) = 0$. Solving the Relaxed Problem without using that result leads to the exact same second-order differential equation for $\mu$ (A.31). This occurs because the new first-order conditions of the problem directly imply that $\mu(\phi) = \mu(\bar{\phi}) = 0$. 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 C.6 does not apply to the case with homogeneous present bias. This happens because each of the open intervals $(\phi, \phi_1)$ 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 (A.31) can be solved in a single interval in $\Phi$. 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.

D Analysis of the Quasi-Linear Limit Case

D.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

$$\theta u_1'(c_1) = u_2'(c_2),$$
which is the planner’s Euler Equation (if the planner observed $\theta$).

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.

D.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

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

subject to

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

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

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

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

We now turn to proving Proposition 2.

D.2.1 Implementability

Given the representation of the problem in the space of $\phi$, 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 $\frac{d}{d\phi} (\phi v_1(\phi) + v_2(\phi)) = v_1(\phi)$, and we
obtain integral incentive compatibility by integrating:

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

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 $\phi$ space:

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

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

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

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

D.2.2 Completing the Model

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

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

Substituting in the characterization of $v_2$ above, we get:

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

That is, the planner chooses a non-decreasing function $v_1$, with the implementability conditions above defining the function $v_2$ that implements this outcome.
From here, we solve the relaxed problem, not subject to monotonicity. The relaxed problem is simply given by

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

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

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

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

**D.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:

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

**D.2.4 Homogeneous $\beta$**

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

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

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

**D.2.5 Heterogeneous $\beta$**

If $\beta$ is heterogeneous and the regularity condition of Proposition 2 is satisfied, then as mentioned before we have:

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

That is, we have an “average Pigouvian tax”: the optimal tax rate on the margin for a type-$\phi$ agent is the average tax rate in that population.
We know that $\pi(\phi)$ must be close to $1 - \beta$ near $\phi$, where the highest $\beta$ types are the only ones with that $\phi$ type. Similarly, we know that $\pi(\phi) \simeq 1 - \beta$ near $\phi$. This suggests a large degree of flexibility over initial withdrawals, and much tighter restrictions on flexibility for households withdrawing a lot.

D.2.6 Corollary 3

The joint density of $(\phi, \theta)$ 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

$$E[\theta|\phi] = \begin{cases} 
\frac{\int_0^{\beta} \theta^2 \phi^{-2} f(\theta) g(\theta\phi^{-1}) d\theta}{\int_0^{\beta} \theta^{-2} f(\theta) g(\theta\phi^{-1}) d\theta}, & \phi < \frac{\theta}{\beta} \\
\frac{\int_\beta^{\phi} \theta^2 \phi^{-2} f(\theta) g(\theta\phi^{-1}) d\theta}{\int_\beta^{\phi} \theta^{-2} f(\theta) g(\theta\phi^{-1}) d\theta}, & \frac{\theta}{\beta} \leq \phi \leq \frac{\theta}{\beta} \\
\frac{\int_\phi^{\infty} \theta^2 \phi^{-2} f(\theta) g(\theta\phi^{-1}) d\theta}{\int_\phi^{\infty} \theta^{-2} f(\theta) g(\theta\phi^{-1}) d\theta}, & \frac{\theta}{\beta} < \phi
\end{cases}$$

If $\beta$ is uniformly distributed, this simplifies to

$$E[\theta|\phi] = \begin{cases} 
\frac{\int_0^{\beta} \theta^2 f(\theta) d\theta}{\int_0^{\beta} \theta^{-2} f(\theta) d\theta}, & \phi < \frac{\theta}{\beta} \\
\frac{\int_\beta^{\phi} \theta^2 f(\theta) d\theta}{\int_\beta^{\phi} \theta^{-2} f(\theta) d\theta}, & \frac{\theta}{\beta} \leq \phi \leq \frac{\theta}{\beta} \\
\frac{\int_\phi^{\infty} \theta^2 f(\theta) d\theta}{\int_\phi^{\infty} \theta^{-2} f(\theta) d\theta}, & \frac{\theta}{\beta} < \phi
\end{cases}$$

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.

A.41
D.3 Derivation of Equation 10

Since from the resource constraint we have \( R \left[ c_1(\phi) + c_2(\phi) \right] dH = Y \), then substituting into the objective function social welfare is

\[
\int \left[ \mathbb{E}[\theta|\phi]u_1(c_1(\phi)) - c_1(\phi) \right] dH(\phi) + Y
\]

Now, consider the consumption policy of a type-\( \phi \) agent. Given quasilinear preferences, absence of a corner solution implies that consumption at date 1 is between 0 and \( Y \) under Autarky, and so we can without loss of generality restrict the balance on the first account to be \( x_1 \leq Y \). The two-account system is thus defined as a pair \((\pi_2, x_1)\) of the penalty on the second account and the contribution of the first account. We can define welfare

\[
W^*(\pi_2) = \max_{x_1 \leq Y} \int \left[ \mathbb{E}[\theta|\phi]u_1(c_1(\phi)) - c_1(\phi) \right] dH(\phi) + Y.
\]

Consider the consumption policy of households faced with a \((x_1, \pi_2)\) system. For values \( \phi \leq \phi_1 \), where \( \phi_1u'_1(x_1) = 1 \), these households withdraw only from the first account and have consumption that solves \( \phi u'_1(c_1(\phi)) = 1 \). For values \( \phi \geq \phi_2 \), where \( \phi_2(1 - \pi_2)u'_1(x_1) = 1 \), these households withdraw from the second account and have consumption that solves \( \phi_2(1 - \pi_2)u'_1(c_1(\phi)) = 1 \). For values \( \phi \in [\phi_1, \phi_2] \), these households are at the corner solution \( c_1(\phi) = x_1 \). Note that consumption policy is continuous and that, holding fixed \( x_1 \), we have \( \frac{\partial c_1(\phi)}{\partial \pi_2} = 0 \) for \( \phi < \phi_2 \). Therefore by Envelope Theorem,

\[
\frac{\partial W^*(\pi_2)}{\partial \pi_2} = \int_{\phi_2}^{\phi} \left[ \mathbb{E}[\theta|\phi]u'_1(c_1(\phi)) - 1 \right] \frac{\partial c_1(\phi)}{\partial \pi_2} dH(\phi).
\]

From the consumption FOC, \( u'_1(c_1(\phi)) = \frac{1}{\phi(1-\pi_2)} \) and so

\[
\frac{\partial W^*(\pi_2)}{\partial \pi_2} = \frac{1}{1 - \pi_2} \int_{\phi_2}^{\phi} \left[ \mathbb{E}[\beta|\phi] - (1 - \pi_2) \right] \frac{\partial c_1(\phi)}{\partial \pi_2} dH(\phi),
\]

where we used \( \theta/\phi = \beta \). The optimal penalty sets \( \frac{\partial W^*(\pi_2)}{\partial \pi_2} = 0 \), which simplifying gives

\[
\int_{\phi_2}^{\phi} \left[ \mathbb{E}[1 - \beta|\phi] - \pi_2 \right] \cdot \frac{\partial c_1(\phi)}{\partial \pi_2} dH(\phi) = 0.
\]
Thus re-expressing as a conditional expectation, we have

\[ E \left[ \mathbb{E}[1 - \beta|\phi] - \pi_2 \right] \cdot \frac{\partial c_1(\phi)}{\partial \pi_2} \bigg| \phi \geq \phi_2 = 0 \]

which is equation (10).

**D.4 Extension: Pareto Weights, Correlated Shocks**

We extend our analysis of the quasilinear model to allow for Pareto (welfare) weights and correlation between shocks, and discuss their implication for the structure of optimal penalties \( \pi(\phi) \).

We introduce shock correlation by allowing for a joint distribution \( F(\theta, \beta) \) for \((\theta, \beta) \in [\theta, \overline{\theta}] \times [\beta, \overline{\beta}]\). We assume that the joint distribution is continuous and has a differentiable density \( f \). As in the baseline model, define \( \phi = \theta/\beta \). We let \( H \) be the marginal distribution of \( \phi \) and \( h \) its density on \( \Phi = [\phi, \overline{\phi}] \).

We introduce welfare weights by assuming that the social planner places a welfare weight \( \omega(\theta, \beta) > 0 \) on household type \((\theta, \beta)\). The social planner’s objective is therefore

\[
\int \int \omega(\theta, \beta) \left[ \theta u_1(c_1) + u_2(c_2) \right] f(\theta, \beta) d\theta d\beta.
\]

Without loss of generality, we normalize \( E[\omega(\theta, \beta)] = 1 \). We can apply the usual transformation of variables to represent social welfare in terms of \( \phi \). Defining \( v_1 : \Phi \to \mathbb{R} \) by \( v_1(\phi) = u_1(c_1(\phi)) \), then social welfare is

\[
\int \left[ E[\omega(\theta, \beta)\theta|\phi]v_1(\phi) + E[\omega(\theta, \beta)|\phi]v_2(\phi) \right] h(\phi) d\phi.
\]

We obtain the following characterization of the optimal penalty in the quasilinear case, that is for \( u_2(c_2) = c_2 \).

**Proposition 24** In the quasilinear case, the optimal penalty in the relaxed problem is

\[
\pi(\phi) = E[\omega(\theta, \beta)(1 - \beta)|\phi] + \int_\phi^\theta (w(x) - 1) dH(x) \frac{\phi h(\phi)}{\phi h(\phi)}
\]

where \( w(\phi) = E[\omega(\theta, \beta)|\phi] \).

The first term in Proposition 24 is an average Pigouvian correction, weighted by the welfare weight assigned to each household. The second term captures a redistributive motive.
Intuitively if $\omega$ is low for $x < \phi$, types below $\phi$ are less valued than types above $\phi$. Thus the planner wants to increase information rents arising to households with higher types. To do so, the planner increases promised utility $v_1(\phi)$ to this lower type, in other words allowing such agents to consume more (i.e., as-if this agent had a higher type). Allowing more overconsumption is achieved by lowering penalties.

**Implications of Welfare Weights.** Consider a planner whose welfare weight varied with $\theta$. Suppose first that the welfare weight was increasing in $\theta$, so that the planner places higher weight on higher $\theta$ agents. Since $\phi$ increases in $\theta$, this also implies a tendency to value higher $\phi$ types more. This creates a force for a smaller corrective tax on low $\phi$ agents (high $\beta$) for two reasons. First, the lower welfare weight at low $\phi$ directly downweights Pigouvian correction in the first term. Second, the increasing welfare weights in $\phi$ prompt a redistributive motive to allow low-$\phi$ agents to overconsume. Conversely, both effects go in the opposite direction at high $\theta$ (high $\phi$), pushing up corrective taxes at high $\theta$ (high $\phi$). This pushes for a system closer to a liquid account and illiquid account, with low penalties at the bottom and high penalties at the top. Conversely, higher welfare weights at low $\theta$ push for greater correction at low $\theta$ (low $\phi$) both for the Pigouguian motive and for the redistributive motive, and conversely for less correction at high $\theta$ (high $\phi$). This pushes for a more Pigouguian-like system.

**Implications of Shock Correlation.** Consider a utilitarian planner, $\omega(\theta, \beta) = 1$, so that the penalty is $\pi(\phi) = \mathbb{E}[1 - \beta|\phi]$. Recall that the penalty tends to be increasing in $\phi$ absent shock correlation, because higher $\phi$ types are achieved by either higher $\theta$ or lower $\beta$. Suppose first that $\beta$ is positively correlated with $\theta$, that is higher $\theta$ types tend to have higher $\beta$. This tends to exacerbate the inference problem of distinguishing high $\theta$ from low $\beta$, pushing towards a more Pigouguian system (and enhancing the motivation for pooling identified in Corollary 3). First, suppose that $\theta$ and $\beta$ go to a limiting case of positive correlation in which $\beta$ is independent of $\phi$. Then, the planner would not be able to learn anything about $\beta$ from eliciting $\phi$. As a result, the optimal penalty would go to the a constant unconditional Pigouguian correction, $\pi(\phi) = \mathbb{E}[1 - \beta]$. On the other hand, if $\beta$ is negatively correlated with $\theta$, then the planner becomes even more certain that low $\phi$ types have high $\beta$ and high $\phi$ types have low $\beta$, leading to very low penalties at low $\phi$ and very high penalties at high $\phi$. This reinforces the motivation for separation with extreme penalties, where low $\phi$ types face small corrective taxes and high $\phi$ types face large corrective taxes.
D.4.1 Proof of Proposition 24

The social planner’s problem is to maximize social welfare subject to the standard incentive and resource constraints. Since the integral incentive and resource constraints are identical to the baseline model, so that as in the proof of Proposition 2 we have

$$\int_{\phi}^{\overline{\phi}} \left( u_1^{-1}(v_1(\phi)) + \int_{\phi}^{\overline{\phi}} v_1(x)dx - \phi v_1(\phi) \right) dH(\phi) + \overline{\phi} v_1(\overline{\phi}) + v_2(\overline{\phi}) = Y$$

Integrating by parts with $$u = \int_{\phi}^{\overline{\phi}} v_1(x)dx$$ and $$dv = dH$$, we have

$$\int_{\phi}^{\overline{\phi}} \int_{\phi}^{\overline{\phi}} v_1(x)dx dH(\phi) = \int_{\phi}^{\overline{\phi}} v_1(x)dx H(\phi) \bigg|_{\phi}^{\overline{\phi}} - \int_{\phi}^{\overline{\phi}} v_1(\phi)H(\phi)d\phi$$

Since $$H(\overline{\phi}) = 0$$, then we have

$$\int_{\phi}^{\overline{\phi}} \int_{\phi}^{\overline{\phi}} v_1(x)dx dH(\phi) = -\int_{\phi}^{\overline{\phi}} v_1(\phi)H(\phi)d\phi,$$

which therefore yields

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

Finally, substituting into the IC constraint yields a definition of the function $$v_2$$ in terms of the function $$v_1$$,

$$v_2(\phi) = -\phi v_1(\phi) + Y + \int_{\phi}^{\overline{\phi}} v_1(\phi)H(\phi)d\phi - \int_{\phi}^{\overline{\phi}} \left( u_1^{-1}(v_1(\phi)) - \phi v_1(\phi) \right) dH(\phi) + \int_{\phi}^{\overline{\phi}} v_1(x)dx.$$

We substitute this into the social welfare function to obtain an optimization problem in terms of $$v_1$$. Defining $$t^\omega(\phi) \equiv E[\omega(\theta, \beta)\theta|\phi]$$ and $$w(\phi) \equiv E[\omega(\theta, \beta)|\phi]$$, we obtain the objective

$$\int t^\omega(\phi)v_1(\phi)dH$$

$$+ \int w(\phi) \left[ -\phi v_1(\phi) + Y + \int_{\phi}^{\overline{\phi}} v_1(\phi)H(\phi)d\phi - \int_{\phi}^{\overline{\phi}} \left( u_1^{-1}(v_1(\phi)) - \phi v_1(\phi) \right) dH(\phi) + \int_{\phi}^{\overline{\phi}} v_1(x)dx \right] dH$$

Integrating by parts with $$u = \int_{\phi}^{\overline{\phi}} v_1(x)dx$$ and $$dv = \omega(\phi)dH(\phi)$$, we have

$$\int w(\phi) \int_{\phi}^{\overline{\phi}} v_1(x)dx dH = -\int v_1(\phi)H^\omega(\phi)d\phi$$

where $$H^\omega(\phi) = \int_{\phi}^{\overline{\phi}} \omega(\phi)dH(\phi)$$. Substituting back in yields the optimization-
Thus we obtain the FOC for optimality

\[
\left( t^w(\phi) - \phi w(\phi) \right) u_1'(c_1(\phi)) h(\phi) + \left[ \mathbb{E}[w] H(\phi) - H^w(\phi) \right] u_1'(c_1(\phi)) - \mathbb{E}[w] \left( 1 - \phi u_1'(c_1(\phi)) \right) h(\phi) = 0
\]

which rearranges to the solution

\[
\left[ \frac{t^w(\phi)}{\mathbb{E}[w]} \right] h(\phi) + \left[ H(\phi) - \frac{H^w(\phi)}{\mathbb{E}[w]} \right] + \left( 1 - \frac{w(\phi)}{\mathbb{E}[w]} \right) \phi h(\phi) \right] u_1'(c_1(\phi)) = h(\phi).
\]

Since we have, without loss, normalized \( \mathbb{E}[w] = 1 \), then

\[
\left[ \mathbb{E}[\omega(\theta, \beta)|\phi] h(\phi) + \int_\phi^\phi (1 - \omega(\phi)) dH + \left( 1 - w(\phi) \right) \phi h(\phi) \right] u_1'(c_1(\phi)) = h(\phi)
\]

Thus defining the penalty by

\[
(1 - \pi(\phi)) \phi u_1'(c_1(\phi)) = 1
\]

we obtain

\[
\pi(\phi) = \mathbb{E}[w(\phi) - \omega(\theta, \beta)] + \frac{\int_\phi^\phi (w(x) - 1) dH(x)}{\phi h(\phi)}
\]

from which the result follows from the definition of \( w(\phi) \).

**E A Pareto Efficient Completely Illiquid Account**

In this Appendix, we shed additional theoretical light on the insights revealed in Section 4.2. Another property revealed in Figure 2 is that starting from autarky, low-\( \beta \) types benefit significantly from introducing a completely illiquid account – that is, from having a two-account system with \( \pi_1 = 0 \) and \( \pi_2 = 1 \) – while high-\( \beta \) types are only slightly hurt by it.

We show in this appendix that starting from autarky, there is always a Pareto improvement that can be achieved by introducing a completely illiquid account with a positive balance. Lower \( \beta \) types benefit more from the introduction of this account. Moreover, increasing the illiquid account balance above this Pareto efficient level delivers large welfare gains for low \( \beta \) types and only small losses for high \( \beta \) types.
Note that in this setup, Pareto efficiency is defined relative to the planner’s normative preferences for a household.

We study the general setup with concave utility. For simplicity, we let $\beta = 1$. As a preliminary, we define $c^A_1(\theta, \beta)$ to be the autarky consumption profile of a type $(\theta, \beta)$ agent. The following proposition is the formal counterpart of the statements above.

**Proposition 25** Starting from autarky (a completely liquid account only):

1. It is Pareto efficient to introduce a completely illiquid account with balance $x_2 = c^A_2(\theta, 1)$.
2. Average welfare gains for type-$\beta$ households, relative to autarky, increase in their time inconsistency $1 - \beta$.
3. A marginal increase in $x_2$ above $c^A_2(\theta, 1)$ generates a second order welfare loss for $\beta = 1$ households and a first order welfare gain for all $\beta < 1$ households.

The first part of Proposition 25 shows that it is Pareto efficient to introduce a completely illiquid account with a balance equal to the lowest date 2 consumption level chosen by any $\beta = 1$ agent (specifically, that chosen by the type $(\theta, \beta) = (\bar{\theta}, 1)$). Intuitively, any date 2 consumption level below this cannot be rationalized by any (positive density) realization of $\theta$ and so must be associated with overconsumption at date 1. Thus mandating a minimum date 2 consumption at this level generates strict welfare gains for households with $\beta < 1$. Since $\beta = 1$ households never desire to consume above this level anyway, the minimum savings level does not impose any welfare losses on them, hence guaranteeing (weak) Pareto efficiency.

The second part of Proposition 25 shows that welfare gains from moving to this system are larger for more time inconsistent households. For $\beta = 1$ households, there are no welfare gains, since they are consuming at the optimal level already. Welfare gains increase the more inconsistent the household is, reflecting that more low-$\beta$ households are more prone to overconsumption and so more bound by the minimum savings level relative to high-$\beta$ households. This matches the pattern of Figure 2, in which welfare gains from the completely illiquid account are largest for low-$\beta$ households.

Finally, the third part of Proposition 25 implies that from a utilitarian perspective, it is always desirable to increase minimum savings above the Pareto efficient level. Intuitively, the higher savings level reduces overconsumption by all $\beta < 1$ households, but also (unnecessarily) ties the hands of the highest-$\theta$ type $\beta = 1$ household. Since this household, type $(\bar{\theta}, 1)$, was already at the optimal consumption policy, by Envelope Theorem the welfare

A.47
loss from forcing it to scale back its consumption is second order. Thus as in Figure 2, the utilitarian planner increases the completely illiquid account balance, leading to small welfare losses for high-\(\beta\) agents but large welfare gains for low-\(\beta\) agents.

E.0.1 Proof of Proposition 25

First, we show that the system is Pareto efficient for all household types \((\theta, \beta)\). Pareto efficiency is trivial for types \((\theta, \beta)\) who chose \(c_1 \leq x_1\) under Autarky since their consumption policy does not change, so consider a type \((\theta, \beta)\) that chose \(c_1 > x_1\) under autarky. By construction, any such household must have \(\beta < 1\) since \(x_1 = c_1^A(\overline{\theta}, 1)\). Therefore, this household’s new consumption policy is \((c_1, c_2) = (x_1, x_2)\), with \(x_1 < c_1^A(\theta, \beta)\) and \(x_2 > c_2^A(\theta, \beta)\). For any \(c_1 \geq x_1 = c_1^A(\overline{\theta}, 1) \geq c_1^A(\theta, 1)\), by concavity we have

\[
\frac{\partial[\theta u_1(c_1) + u_2(1-c_1)]}{\partial c_1} < \overline{\theta} u_1'(c_1) - u_2'(c_2) \leq 0
\]

and therefore we have \(\theta u_1(x_1) + u_2(x_2) \geq \theta u_1(c_1^A(\theta, \beta)) + u_2(c_2^A(\theta, \beta))\). Hence, every household type \((\theta, \beta)\) is weakly better off, giving Pareto efficiency.

To prove the second part, the welfare gain for a type \((\theta, \beta)\) household with \(c_1^A(\theta, \beta) \geq x_1\) is

\[
\Delta(\theta, \beta) = \theta u_1(x_1) + u_2(x_2) - \left(\theta u_1(c_1^A(\theta, \beta)) + u_2(c_2^A(\theta, \beta))\right)
\]

observe that if a type \((\theta, \beta)\) household has \(c_1^A(\theta, \beta) \geq x_1\), then for \(\beta' < \beta\) a type \((\theta, \beta')\) household has \(c_1^A(\theta, \beta') > c_1^A(\theta, \beta) \geq x_1\). Therefore by concavity (using the first step), we have \(\Delta(\theta, \beta') > \Delta(\theta, \beta)\), and so low-\(\beta\) households benefit more than high-\(\beta\) households.

The third part is immediate from Envelope Theorem since \(x_1 = c_1^A(\overline{\theta}, 1)\) and \(x_2 = c_2^A(\overline{\theta}, 1)\) (with strict gains for \(\beta < 1\) immediately following from the prior two steps and from the second order loss for a \(\theta = \overline{\theta}\) household).

F Extension: Labor Supply and Income Heterogeneity

In this appendix, we extend our baseline model to allow for endogenous labor supply and unobservable heterogeneity in earnings ability, which leads to income heterogeneity. The social planner’s problem is to choose not only optimal consumption penalties ("retirement system") but also income redistribution ("tax-transfer system"), subject to incentive compatibility of labor supply choices. We analyze the quasilinear case with welfare weights on the different ability types, and show that the motive to redistribute income while maintaining labor supply incentives leads the planner to adopt a system with larger-than-Pigou
withdrawal penalties for low ability (low income) types and smaller-than-Pigou withdrawal penalties for high ability (high income) types. Intuitively, allowing high ability types to withdraw more flexibly moves them closer to their preferred allocation, increasing incentives for labor supply and facilitating income redistribution.

F.1 Introducing Earnings Ability and Labor Supply

There is, for simplicity, measure 2 of households. The household decision problem unfolds in three steps. At dates 1 and 2, households face the consumption choice problem of the baseline model. At date 0 (before learning their date 1 type), households are equally likely to have a high or low unobservable earnings ability type \( j \in \{H, L\} \) that determines their disutility \( \kappa^j \) of labor supply. A household of type \( j \) that chooses to work hours \( \ell^j \geq 0 \) generates date 1 income of \( Y(\ell^j) \) at a private utility cost of \(-\kappa^j \ell^j \). High earnings ability types have a lower cost of labor supply, that is \( \kappa^H < \kappa^L \). Household ability is unobservable to the planner and independent of the consumption type.\(^{12}\)

The expected utility of a type \( j \) household who will have a consumption schedule \((c^j_1, c^j_2)\), from the household’s perspective, is given by\(^{13}\)

\[
-\kappa^j \ell^j + \mathbb{E}[\phi u_1(c^j_1) + u_2(c^j_2)]
\]

The resource constraint of consumption for a type \( j \) household is given by

\[
\mathbb{E}[c^j_1 + c^j_2] = Y^j,
\]

where \( Y^j \) is the after-tax income of type \( j \) households. Finally, after-tax income to each

---

\(^{12}\)We interpret the timing of the labor supply choice before the consumption type is revealed as corresponding to the household placing income in a savings system (or the government doing so on their behalf), with the possibility of making pre-retirement withdrawals. Withdrawals and their corresponding penalties are reflected in the consumption profile \((c^j_1, c^j_2)\) and implied penalties we characterize. Assuming that \( \kappa \) independent of \((\theta, \beta)\) avoids technical complexity associated with persistent private information.

\(^{13}\)In specifying these preferences, we have assumed that a naive household perceives the distribution of its type to be that of \( \phi \). This interpretation of naivete is consistent with the idea that the household has type \( \phi \) and, being naive, treats variation in \( \phi \) as though it were variation in \( \theta \). This could also be seen more broadly as the household having a type \( \phi \) but the planner perceiving a bias \( \beta = \theta/\phi \) in the household’s type. Another interpretation of naivete would be that households believe the distribution of its type to be that of \( \theta \). Under this interpretation, the planner would be motivated to use the consumption profile to incentivize labor supply for \( \phi < \bar{\theta} \), but would not have the motive for \( \phi > \bar{\theta} \) since households believe these states are zero probability. Implementing this interpretation is technically complicated because households’ belief that states \( \phi > \bar{\theta} \) occur with zero probability would lead the planner to pool different ability types at the same allocation when \( \phi > \bar{\theta} \), allowing “free” redistribution and early withdrawal penalties above this point except as bound by consumption incentive compatibility and monotonicity.
household type have to add to aggregate output, that is

\[ Y^H + Y^L = Y(\ell^H) + Y(\ell^L). \]

If the planner engineers a system with \( Y^H < Y(\ell^H) \), we interpret the planner as instituting a tax-transfer system that taxes high-income households and subsidizes low-income households.

**Incentive Compatibility of Labor Supply.** Consider a social planner that designs a system \( \{\ell^j, c_1^j, c_2^j\}_{j \in \{H, L\}} \), with after-tax income \( Y^j \) defined implicitly from consumption profiles. Because the earnings ability type \( j \) is not observable, labor supply choices have to be incentive compatible. A high ability household \( j = H \) prefers their allocated profile \( (\ell^H, c_1^H, c_2^H) \) to the alternative \( (\ell^L, c_1^L, c_2^L) \) if

\[
\kappa^H (\ell^H - \ell^L) \leq \mathbb{E} \left[ \phi u_1(c_1^H) + u_2(c_2^H) \right] - \mathbb{E} \left[ \phi u_1(c_1^L) + u_2(c_2^L) \right],
\]

that is if the disutility cost of working more is compensated for by a better consumption profile. This limits the scope for income redistribution to the extent that the planner wishes to maintain higher labor supply levels at higher-ability workers. As long as the high ability type’s incentive constraint binds and \( \ell^H \geq \ell^L \), then since \( \kappa^H < \kappa^L \) the low ability household’s incentive constraint also binds.

**F.2 Optimal Penalties in the Quasilinear Case with Welfare Weights**

To illustrate how endogenous labor supply and unobservable earnings ability heterogeneity affects the design of the optimal withdrawal penalty system (and its interaction with income redistribution), we study the optimal system in the quasilinear case, \( u_2(c_2^j) = c_2^j \), when there is a redistributive motive due to a higher social welfare weight on low-ability households. Formally, the social planner chooses a system \( \{\ell^j, c_1^j, c_2^j\}_{j \in \{H, L\}} \) to maximize social welfare,

\[
\sum_{j \in \{H, L\}} \omega^j \left( -\kappa^j \ell^j + \mathbb{E} \left[ \mathbb{E} \left[ \theta | \phi \right] u_1(c_1^j) + u_2(c_2^j) \right] \right)
\]

subject to incentive compatibility of labor supply, incentive compatibility of the consumption choices, and the resource constraints. \( \omega^j \) is the welfare weight placed on a type-\( j \) household, and we assume that \( \omega^L \geq \omega^H \) so that there is a motive to redistribute from high-ability (high income) to low-ability (low income) households.

We obtain the following characterization of the socially optimal consumption penalty system (“retirement system”). In its proof, we provide details on optimal labor supply and
Proposition 26 The optimal consumption penalties for type $j$ households are

$$\pi^j(\phi) = \frac{\omega^j}{\mathbb{E}[\omega]} \mathbb{E}[1 - \beta|\phi]$$

To understand Proposition 26, consider a planner that started from Pigouvian correction, $\pi^j(\phi) = \mathbb{E}[1 - \beta|\phi]$. In this case the labor supply choice boils down to a comparison of the income level against the cost of labor. However, at this consumption profile the social planner is indifferent on the margin to any high-earnings-ability household consuming slightly more at any $\phi$, while high-ability households perceives a first order cost to unnecessary constraints on their consumption choices. Thus the planner is willing to allow high-ability households to consume slightly more for any $\phi$, which generates a first order benefit in promoting labor supply and income redistribution at only a second order cost (from the planner’s perspective) in distorting the household’s intertemporal consumption allocation. Thus, the planner adopts a penalty system for high income households with lower penalties than the simple conditional Pigouvian correction. Conversely, higher consumption penalties for low-ability households also encourages labor supply by making the consumption profile of low-ability households less attractive to high-ability households. This once again allows for a first-order gain via redistribution at a second-order cost via intertemporal consumption patterns.

F.3 Proof of Proposition 26

The social planner’s problem is to choose a system $\{c^j_1, c^j_2, \ell^j, Y^j\}$ to maximize social welfare,

$$\sum_{j \in H,L} \omega^j \int \left[ \mathbb{E}[\theta|\phi]u_1(c^j_1) + u_2(c^j_2) \right] dH$$

subject to labor supply incentive compatibility,

$$r^H(\ell^H - \ell^L) \leq \mathbb{E} \left[ \phi u_1(c^H_1) + u_2(c^H_2) \right] - \mathbb{E} \left[ \phi u_1(c^L_1) + u_2(c^L_2) \right]$$

to the resource constraints,

$$\sum_{j \in \{H,L\}} Y^j \leq \sum_{j \in \{H,L\}} Y(\ell^j)$$

$$\mathbb{E}[c^j_1 + c^j_2] \leq Y^j$$

A.51
and to incentive compatibility of the consumption choices at date 1.

Following the proof of Proposition 2, for a monotone consumption profile \( c^j_1 \) we can internalize incentive compatibility to write

\[
E \left[ E[\theta|\phi]u_1(c^j_1) + c^j_2 \right] = E \left[ E[\theta|\phi]u_1(c^j_1) - c^j_1 \right] + Y^j
\]

\[
E \left[ \phi u_1(c^j_1) + c^j_2 \right] = E \left[ \phi u_1(c^j_1) - c^j_1 \right] + Y^j
\]

Therefore, the maximization problem of the social planner is equivalently written as

\[
\max_{\{c^j_1, \ell, Y\}} \sum_{j \in \{H,L\}} \omega^j \left[ -\kappa^j \ell^j + E \left[ E[\theta|\phi]u_1(c^j_1) - c^j_1 \right] + Y^j \right]
\]

subject to

\[
\kappa^H (\ell^H - \ell^L) \leq E \left[ \phi u_1(c^H_1) - c^H_1 \right] + Y^H - E \left[ \phi u_1(c^L_1) - c^L_1 \right] - Y^L
\]

\[
\sum_{j \in \{H,L\}} Y^j \leq \sum_{j \in \{H,L\}} Y(\ell^j)
\]

Letting \( \mu \) be the Lagrange multiplier on labor incentive compatibility and \( \lambda \) the Lagrange multiplier on the resource constraint, the Lagrangian is

\[
\mathcal{L} = \sum_{j \in \{H,L\}} \omega^j \left[ -\kappa^j \ell^j + E \left[ E[\theta|\phi]u_1(c^j_1) - c^j_1 \right] + Y^j \right]
+ \mu \left( E \left[ \phi u_1(c^H_1) - c^H_1 \right] + Y^H - E \left[ \phi u_1(c^L_1) - c^L_1 \right] - Y^L - \kappa^H (\ell^H - \ell^L) \right)
+ \lambda \left( \sum_{j \in \{H,L\}} Y(\ell^j) - \sum_{j \in \{H,L\}} Y^j \right)
\]

The FOCs for \( Y^H \) and \( Y^L \) are, respectively,

\[
0 = \omega^H + \mu - \lambda
\]

\[
0 = \omega^L - \mu - \lambda
\]

and so differencing the two equations, we obtain

\[
\mu = \frac{\omega^L - \omega^H}{2}
\]
\[ \lambda = \frac{\omega_L + \omega_H}{2} = \mathbb{E}[\omega] \]

Next, we can take the FOC in \( c^H_1(\phi) \) to obtain

\[ 0 = \omega_H \left( \mathbb{E}[\theta|\phi] u'_1(c^H_1) - 1 \right) + \mu \left( \phi u'_1(c^H_1) - 1 \right) \]

Rearranging, we have

\[ \left( \frac{\omega_H}{\omega_H + \mu} \mathbb{E}[\beta|\phi] + \frac{\mu}{\omega_H + \mu} \right) \phi u'_1(c^H_1) = 1 \]

Thus we have the optimal marginal penalty

\[ \pi^H(\phi) = 1 - \left( \frac{\omega_H}{\omega_H + \mu} \mathbb{E}[\beta|\phi] + \frac{\mu}{\omega_H + \mu} \right) = \frac{\omega_H}{\omega_H + \mu} \mathbb{E}[1 - \beta|\phi] = \pi(\phi) = \frac{\omega_H}{\mathbb{E}[\omega]} \mathbb{E}[1 - \beta|\phi] \]

which gives the result for \( j = H \).

Next taking the FOC for \( c^L_1(\phi) \), we have

\[ 0 = \omega_L \left( \mathbb{E}[\theta|\phi] u'_1(c^L_1) - 1 \right) - \mu \left( \phi u'_1(c^L_1) - 1 \right) \]

\[ \left( \frac{\omega_L}{\omega_L - \mu} \mathbb{E}[\beta|\phi] - \frac{\mu}{\omega_L - \mu} \right) \phi u'_1(c^L_1) = 1 \]

and therefore the penalty is

\[ \pi^L(\phi) = 1 - \left( \frac{\omega_L}{\omega_L - \mu} \mathbb{E}[\beta|\phi] - \frac{\mu}{\omega_L - \mu} \right) = \frac{\omega_L}{\omega_L - \mu} \mathbb{E}[1 - \beta|\phi] = \omega_L \mathbb{E}[\omega] \mathbb{E}[1 - \beta|\phi] \]

which gives the result for \( j = L \).

Finally, we can complete the model through the optimal labor supply FOC, where we have for \( j = H \)

\[ 0 = -\omega_H \kappa^H - \mu \kappa^H + \lambda Y'(\ell^H) \]

\[ Y'(\ell^H) = \kappa^H \]

Analogously for low ability types, we have

\[ 0 = -\omega_L \kappa^L + \mu \kappa^H + \lambda Y'(\ell^L) \]

\[ Y'(\ell^L) = \kappa^L + \frac{\mu}{\mathbb{E}[\omega]} (\kappa^L - \kappa^H) \]

A.53
Lastly, income redistribution is then defined by the binding labor supply incentive compatibility constraint.