Optimal Illiquidity*

John Beshears, James Choi, Christopher Clayton
Christopher Harris, David Laibson, Brigitte Madrian

September 12, 2017

Abstract

We calculate the socially optimal level of illiquidity in a stylized retirement savings system. We solve the planner's problem in an economy in which time-inconsistent households face a tradeoff between commitment and flexibility (Amador, Werning and Angeletos, 2006). We assume that the planner can set up multiple accounts for households: a perfectly liquid account and/or partially illiquid retirement savings accounts with early withdrawal penalties. Revenue from penalties is collected by the government and redistributed through the tax system. We solve for the socially optimal values of these penalties, and the socially optimal allocations to these accounts. When agents have heterogeneous present-biased preferences, the socially optimal system has three accounts: (i) a liquid account, (ii) an account with an early withdrawal penalty of ≈ 100%, and (iii) an account with an early withdrawal penalty of ≈ 10%. With heterogeneous preferences, the socially optimal retirement savings system in our stylized model looks surprisingly like the existing U.S. system: (i) a liquid account, (ii) an illiquid Social Security account (and defined benefit pensions), and (iii) a 401(k)/IRA account with a 10% penalty. The socially optimal allocations to these accounts and the predicted equilibrium flows of early withdrawals – "leakage" – also match the U.S. system.

*This research was supported by the U.S. Social Security Administration through grant #RRC08098400-09 to the National Bureau of Economic Research as part of the SSA Retirement Research Consortium. This project also received funding from the TIAA Institute and Wharton School's Pension Research Council/Boettner Center. The findings and conclusions expressed are solely those of the authors are do not represent the views of SSA, any agency of the Federal Government, the TIAA Institute, Wharton School's Pension Research Council/Boettner Center, or the NBER. The authors are grateful for comments from David Richardson and seminar participants at Harvard University, the SSA RRC, and the NBER. Beshears: Harvard Business School and NBER. Choi: Yale School of Management and NBER. Clayton: Harvard University, Department of Economics. Harris: University of Cambridge. Laibson: Harvard University, Department of Economics and NBER. Madrian: Harvard Kennedy School and NBER.

1 Introduction

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

If illiquidity is optimal, how should it be implemented? Possible forms of illiquidity include a perfectly illiquid retirement claim (like a typical defined benefit pension or Social Security) or a partially illiquid account (like an IRA or 401(k) plan). In theory, an optimal system might combine different types of illiquid accounts.

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

But that's where agreement ends. For example, in most developed countries defined contribution (DC) savings accounts are usually completely illiquid before age 55 (Beshears et al 2015). By contrast, in the US, certain types of withdrawals from DC accounts are allowed without penalty, and, for IRAs, withdrawals may be made for any reason if a 10% penalty is paid. Liquidity allows significant pre-retirement "leakage": for every \$1 contributed to the accounts of US savers under age 55, \$0.40 simultaneously flows out of the 401(k)/IRA system, not counting loans (Argento, Bryant, and Sabelhaus 2014). Until now, no normative model has been used to determine whether such leakage is good or bad from the perspective of overall social welfare. Nevertheless, most media coverage bemoans leakage.

Our paper evaluates the optimality of an N-account retirement savings system with a combination of liquid, partially illiquid, and/or fully illiquid accounts. Within this framework, we focus on two special cases: systems with two accounts and systems with three accounts. In all of our analysis we will assume that the first account is fully liquid, so our two-account system has a fully liquid account and a partially (or fully) illiquid account. Likewise, our three-account system has a fully liquid account and two partially illiquid accounts (one of which might be fully illiquid). We show that the three-account system is a good

approximation (with respect to expected welfare) for a completely general mechanism design solution.

We study preferences that include both normatively legitimate taste shifters and normatively undesirable self-control problems. The self-control problems are modeled as the consequence of present bias (Phelps and Pollak 1968, Laibson 1997): i.e., a discount function with weights $\{1, \beta\delta^2, \ldots, \beta\delta^t\}$, where the degree of present bias is $1-\beta$. Our model is an aggregate version (with interpersonal transfers) of the flexibility/commitment framework of Angeletos, Werning, and Amador (2006; hereafter refered to as AWA). Our model is also closely related to the model of Moser and Olea de Souza e Silva (2017), who also generalize AWA. They study the case of unobservable earnings ability and unobservable β, whereas we study the case of unobservable taste shocks (with exogenous earnings) and unobservable β. Moser and Olea de Souza e Silva (2017) find that second-best optimal savings institutions have many of the properties of the U.S. retirement savings system, a theme that also emerges in our analysis.

We divide our analysis into the cases of homogeneous present bias and heterogeneous present bias. In the homogeneous case, we assume that all agents have the same degree of present bias – in other words, the same value of β. Under homogeneous β, our model implies that partially illiquid accounts with penalties $\pi \simeq 1-\beta$ play an economically significant role in improving social welfare.

We then relax the homogeneity assumption, and assume that agents have *heterogeneous* present-bias. In this heterogeneous-preference case, we find that fully illiquid accounts play an important role in improving welfare, whereas partially illiquid accounts matter relatively little. We show that the socially optimal degree of illiquidity mostly caters to the households with the lowest β values. Completely illiquid retirement savings generates large welfare gains for these low-β agents and these welfare gains swamp the welfare losses of the high-β agents (who are made slightly worse off by shifting some of their wealth from perfectly liquid accounts to perfectly illiquid accounts).

To the extent that there is a role for partially illiquid accounts in the heterogeneous-β

economy, we find that they should have low early withdrawal penalties – approximately 10%. This implies that the partially illiquid accounts look much like a typical 401(k) account. Moreover, these partially illiquid accounts display a high level of leakage in equilibrium. In other words, early withdrawals (i.e., pre-retirement withdrawals) are common from this partially illiquid account. This leakage is a two-edged sword: it results in part from legitimate taste shocks and in part from self-control problems (i.e., low β). The costs of the partially illiquid account to low β types (who end up paying most of the early withdrawal penalties) and benefits to high β types (who are net recipients of these penalties) are nearly off-setting, although the net effect for all of society is slightly positive.

Section 2 describes the planner's problem – i.e., account allocations and early withdrawal penalties that maximize social welfare subject to information asymmetries between the planner and the households. Section 3 analyzes the solution to the planner's problem in the case of autarky (inter-household transfers are not permitted) and present bias that is homogeneous. Section 4 analyzes the solution to the planner's problem when inter-household transfers are admitted and present bias is homogeneous. Section 5 analyzes the solution to the planner's problem in the case of inter-household transfers and heterogeneous present-bias. This is the most realistic benchmark that we study. We also show that our optimized retirement savings system is characterized by very high rates of leakage from the partially illiquid retirement savings account, suggesting that the US system, which exhibits high leakage in practice, is not necessarily suboptimal (though it is 'second-best' because of information assymetries). In section 6, we conclude and discuss the limitations of our existing analysis and our goals for future work.

2 Model

We study a two-period model of consumption for a continuum of households with unit mass. Households are indexed by taste shocks (θ) and present bias (β). In period one, a household consumes $c_1(\theta, \beta)$. In period two, a household consumes $c_2(\theta, \beta)$. For example, think of period 1 as working life and period 2 as retirement. When the values of θ and β are implicit,

we will simply refer to c_1 and c_2.

In this model, we give households access to N savings accounts with initial balances $\{x_n\}_{n=1}^N$, early withdrawal penalties $\{\pi_n\}_{n=1}^N$, and allow households to withdraw from these accounts in whatever order they choose – in equilibrium they will choose to withdraw from the low-penalty accounts first. This N-account model is a special case of the fully general mechanism design problem, which is discussed in Appendix A and quantitatively analyzed in subsection 4.1, where we show that the welfare that arises from the N-account framework (with $N = 3$) is close to the welfare for the completely general mechanism. We choose to focus on the N-account framework because of its similarity to the actual retirement savings systems that are currently in use around the world.

2.1 Preferences of the social planner

The *social planner* has the following preferences over consumption in periods 1 and 2:

$$\theta u(c) + \delta v(c)$$

where θ is a taste shifter[1], δ is a discount factor, and u and v are both strictly increasing and strictly concave functions. We assume that u' and v' converge to ∞ as their respective arguments fall to zero. Following AWA and Beshears et al (2015), we assume that θ has bounded support with closure $[\underline{\theta}, \overline{\theta}]$, where $0 < \underline{\theta}$.

2.2 Preferences of households

We now describe the preferences of the *households* in this economy. Self 1 has the same preferences as the social planner, but Self 1 also has present bias:

$$\theta u(c) + \beta\delta v(c)$$

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

where

$$0 < \beta \leq 1.$$

Bounding β at 1 is without loss of generality.

2.3 Information structure.

We assume that households are naive in the sense that they don't anticipate their own present bias and hence won't use commitment strategies.

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

2.4 Timing.

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

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

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

The planner must satisfy intertemporal, economy-wide budget balance (under the assumption that the gross interest rate is one). We state the budget constraint in two equivalent ways. First, the integral of equilibrium consumption over states, must equal the overall resources in the economy, which are normalized to one. Our framework assumes a continuum population of consumers (with measure one), so that integrating over taste-shock states is the same as integrating over consumers. Accordingly, the budget constraint can be written:

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

An equivalent way of describing budget balance is to relate allocations to resources. Allocations are the accounts given to each consumer. Resources are both the initial unit endowment and the revenue raised from penalties paid in equilibrium. Let $\omega_n(\theta,\beta)$ be equilibrium period-1 withdrawals from account n across the population of consumers. Then the budget constraint can be written:

$$\begin{aligned}
\int c_1(\theta,\beta)\, dF(\theta)\, dG(\beta) &= \sum_{n=1}^{N} \left[(1-\pi_n) \int \omega_n(\theta,\beta)\, dF(\theta)\, dG(\beta) \right]. \\
\int c_2(\theta,\beta)\, dF(\theta)\, dG(\beta) &= \sum_{n=1}^{N} \left[x_n - \int \omega_n(\theta,\beta)\, dF(\theta)\, dG(\beta) \right]. \\
\int \left(c_1(\theta,\beta) + \frac{c_2(\theta,\beta)}{R} \right) dF(\theta) dG(\beta) &= \sum_{n=1}^{N} x_n - \sum_{n=1}^{N} \left[\pi_n \int \omega_n(\theta,\beta)\, dF(\theta)\, dG(\beta) \right] = 1.
\end{aligned}$$

Time 1: Self 1 maximizes her perceived welfare from the perspective of time 1 (which includes present bias). This will generate withdrawals from the accounts established at date 0.

Time 2: Self 2 spends any remaining funds in the accounts.

2.5 Summary of the N-account mechanism design problem

We can now jointly express both the planner's problem and the consumer's problem. We begin with the consumer's problem, since consumer behavior is an input to the planner's

problem. In essence, the consumer has only one decision to make. In period 1, the consumer with parameters θ and β faces this problem:

$$\max_{\{\omega_n\}_{n=1}^N} \theta u(c_1) + \beta v(c_2) \tag{2}$$

subject to the constraints,

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

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

This problem generates equilibrium policy functions $c_1(\theta,\beta)$ and $c_2(\theta,\beta)$.

In period 0, the planner faces the following problem:

$$W = \max_{\{x_n\}_{n=1}^N,\, \{\pi_n\}_{n=1}^N} \int \left[\theta u(c_1(\theta,\beta)) + v(c_2(\theta,\beta))\right]\, dF(\theta)dG(\beta) \tag{5}$$

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

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

which is equivalent to equation 1.

The problem summarized in the last subsection is a restricted version of a completely general mechanism design problem. We compare our results to the solution of the general mechanism design problem in the Appendix.

3 Optimal Illiquidity with Homogeneous Present Bias and Autarky

We begin our analysis by reviewing a special case that requires two key assumptions. First, we assume that all agents share a common value of β – i.e., a common degree of present bias. Hence, the distribution function $G(\beta)$ is degenerate.

Second, we assume "autarky." By this we mean, no interpersonal transfers are possible, so each household can consume no more than their original endowment. Note that our autarky assumption allows interactions between households and the government, but those interactions can't engender transfers across households. In other words, the government needs to burn any resources that it collects from households (rather than transferring those resources to other households). We relax this extreme assumption starting in the next section and for the remainder of the paper.

With the assumption of autarky, our problem can be expressed using our standard notation with the aggregate budget constraint replaced by an autarkic budget constraint:

$$c_1(\theta, \beta) + c_2(\theta, \beta) \leq 1 \text{ for each household with parameters } \theta \text{ and } \beta.$$

Assuming a degenerate $G(\beta)$, autarky, and a weak restriction on $F(\theta)$, the socially optimal allocation is achieved with only two accounts: an account that is completely liquid and an account that is completely illiquid in period 1 and completely liquid in period 2. Any additional accounts (with intermediate levels of liquidity) do not have value to the planner.

Assume that F is differentiable and let $G(\theta) = (1-\beta)\theta F'(\theta) + F(\theta)$. Assume there exists $\theta_M \in [\underline{\theta}, \overline{\theta}]$ such that: (i) $G' \geq 0$ on $(0, \theta_M)$; and (ii) $G' \leq 0$ on (θ_M, ∞). Assume that F' is bounded away from zero on $[\underline{\theta}, \overline{\theta}]$. We refer to these as the assumptions on F.

Theorem 1 *(AWA) Assume there is a homogeneous population-wide value of β. Assume that households are in autarky. Welfare is maximized by giving self 1 two accounts: a completely liquid account and a completely illiquid account.*

This result is a corollary of a result in AWA.[3] The result depends critically on the autarky assumption, as we will show in the next section. With autarky (and the density assumption in Proposition 1), intermediate penalties (i.e., $0\% < \pi < 100\%$) are socially inefficient because they force resources to be destroyed – the revenue from penalties can't be transferred to other households and must be wasted.

This proposition implies that no welfare benefits are achieved by increasing the number of accounts beyond $N = 2$ in the N-Account Mechanism Design Problem (equations 2-6). But the proposition relies on an extremely strong (and unrealistic) assumption, notably autarky.

4 Optimal Liquidity with Homogeneous Present Bias and Transfers

The current section relaxes the autarky assumption. Specifically, we now revert to overall budget balance rather than consumer-by-consumer budget balance. With overall budget balance, a perfectly liquid and a perfectly illiquid account are not jointly sufficient to maximize social surplus. We continue to make the same assumptions on F that we adopted in the previous section.

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

In other words, when transfers are allowed, a completely liquid account and a completely illiquid account are not jointly sufficient to obtain the social optimum.

This result is proven in the appendix.

4.1 Optimal Policy with N Accounts.

If a perfectly liquid account and a perfectly illiquid account do *not* jointly obtain the social optimum, what happens under other account structures? In this subsection we answer this

[3] See Ambrus et al (2013) for related information about this argument. The Theorem is stated here with a slightly stronger condition on F than the condition used by AWA (see Beshears et al 2015).

question with simulation results. Each simulation has a different assumption on the number of accounts and the scope that the planner has to set withdrawal penalties on those accounts. In our benchmark simulations we make the following assumptions.

A1. The per period utility functions are $u(c) = v(c) = \ln(c)$;

A2. The density of the multiplicative taste shocks is truncated normal, with mean $\mu = 1$ and standard deviation $\sigma = 1/4$, and support $[1-\chi, 1+\chi]$, where $\chi = 2/3$. Note that $E[\theta] = 1$.

In the analysis reported in Figure 1, we plot total welfare (W) for the case of two accounts: a completely liquid account and a partially illiquid account with the penalty (π_2), which is represented by the horizontal axis. For each point on the figure, we optimize the allocation to the liquid account (x_1) and the allocation to the partially illiquid account (x_2), holding fixed the penalty on the horizontal axis. Figure 1 assumes that $\beta = 0.7$. The peak welfare is obtained at a value of $\pi_2 = 0.28$, implying that the optimal withdrawal penalty is 28% (in this two-account system).

Figure 2 repeats this exercise for ten cases: in each case β is fixed at a population-wide (homogeneous) value. For every case, we assume that all agents in the economy have the same value of β, and we plot total welfare for the case of one fully liquid account and one partially illiquid account (with the penalty, π_2, represented by the horizontal axis, and the account allocation optimized for this particular penalty and the assumed population-wide value of β). Figure 2 reveals that the optimal value of π_2 is approximately equal to $1-\beta$. This near match between π_2 and $1-\beta$ is easy to see on Figure 3, where we plot the optimal level of π_2 as a function of the population-wide value of β; now β is on the horizontal axis.

Table 1 reports the welfare consequences of other account structures, including the solution to the non-linear mechanism design problem. (Each column in Table 1 represents a homogeneous-β economy.)

Row 1 is the money metric welfare gain by moving from a system with one fully liquid account to two accounts (one purely liquid and one partially liquid). As expected, the welfare

gains are enormous for $\beta = 0.1$ (a welfare gain of 71.65% of wealth), modest for large β values (e.g., a welfare gain of 1.29% of wealth for the case of $\beta = 0.7$) and non-existent for $\beta = 1$.

Row 2 is the money metric welfare gain by moving from all-liquid to 3 accounts (one purely liquid and two partially liquid). For this case, the optimal penalties continue to track $1 - \beta$. In particular, it is approximately the case that $\frac{\partial \pi_n}{\partial \beta} = -1$ for all $n = \{2, 3\}$. The welfare gains in row 2 closely mimic the welfare gains in row 1.

Row 3 is the money metric welfare gain by moving from one fully liquid account to the (completely general) solution to the non-linear mechanism design problem. We discuss the formal set-up for this problem in Appendix A. These welfare gains also mimic the welfare gains in row 1, with one interesting conceptual exception (which is not quantitatively important). When $\beta = 1$ there is a welfare gain of 0.02% of wealth (2/100th's of 1%). This small welfare gain derives from redistribution. The mechanism is set-up to effectively transfer resources to high θ types, which requires a non-convex budget set. Such a budget set is not possible in the N-account system (because we constrain the penalties to be positive).

Row 4 is the money metric welfare gain by moving from one fully liquid account to two accounts: one completely liquid and one completely illiquid. These welfare gains are similar to rows 1-3 for low β values. However, for larger β values (i.e., $\beta \in \{0.5, 0.6, 0.7, 0.8\}$) the welfare gains are meaningfully smaller than the welfare gains in rows 1-3. For example, for $\beta = 0.7$, row 1 reports a welfare gain of 1.33% of wealth, while row 4 has a welfare gain of 1.02% of wealth. In other words, meaningful welfare gains can be obtained by using accounts with intermediate penalties (rows 1 and 2) or general mechanisms (row 3), instead of being constrained to use only accounts that are fully illiquid and fully illiquid (row 4).

Some readers may also be interested in the specific penalty values and account allocations for the cases studied in rows 1-4. These penalties and allocations are reported in appendix B.

5 Optimal Liquidity with Transfers and Heterogeneous Present Bias

In this section, we continue to assume overall budget balance (rather than consumer-by-consumer budget balance). In addition, we now *relax* the assumption that consumers have a homogeneous discount parameter, β. As in the previous section, we begin with a theoretical result and then provide quantitative simulations. Our theoretical result studies the case where the heterogeneous population of consumers is concentrated at two boundary points, $\beta = 0$ and $\beta = 1$. This 'limiting case' of heterogeneity turns out to be analytically tractable and sheds light on the general case of heterogeneity.

Theorem 3 *Let the mass of agents with $\beta = 1$ be μ and the mass of agents with $\beta = 0$ be $1 - \mu$. Assume that the planner is constrained by aggregate budget balance (rather than autarkic budget balance). Then the socially optimal N-account allocation is achieved by a two-account system with one perfectly liquid account and one perfectly illiquid account.*

The intuition for this result follows from two observations. First, adding partially illiquid accounts transfers resources from $\beta = 0$ households (who pay the penalties) to $\beta = 1$ households (who are net recipients of the penalties because of aggregate budget balance). Second, this additional source of inequality is welfare-reducing because utility is concave.

It turns out that this exact result is still approximately true with far less extreme distributions of β values.

5.1 Optimal policy with transfers and heterogeneous present bias: the general case.

The previous theorem reports the case in which heterogeneity is extreme: some housholds have $\beta = 0$ and others have $\beta = 1$. In this case, a system with a fully liquid account and completely illiquid account achieves the (constrained-efficient) social optimum. In the current and the following subsections, we explore the robustness of this result using numerical

simulations. Specifically, we study cases in which β has a distribution of values over the points $\{0.1, 0.2, 0.3, ..., 1\}$.

As before, each simulation has a different assumption on the (finite) number of accounts and the scope that the planner has to set withdrawal penalties on those accounts. In all of these cases we make assumptions A1 and A2 from the previous section (pinning down the utility functions, u and v, and the density on the taste shock, $f(\theta)$).

In the current section, we make an additional assumption on the distribution of β values.

A3. β takes only the discrete values in the set $B = \{0.1, 0.2, 0.3, ..., 1\}$ (with probabilities that sum to one).

Our benchmark distribution is plotted in Figure 4, which is generated from the continuous density $f(\beta) = \beta^{a-1}(1-\beta)^{b-1}$, with calibratin $a = 2.3$ and $b = 1$. This distribution/calibration was chosen with the following goals in mind. First, we wanted a density that was bounded between 0 and 1. Second, we wanted a distribution that had a mean value close to 0.7.[4] Third, we wanted a distribution that had declining mass as β falls from 1 to 0, implying that the modal agent in the economy has no self-control problem. This is the first calibration that we tried. (We report additional distributions – i.e., robustness checks – in the next subsection and find that the results change very little.)

Figure 5 reports the case of one fully liquid account and one partially illiquid account, where the penalty of the partially illiquid account is plotted on the horizontal axis. The resulting welfare for every type in the economy is plotted on the vertical axis. Figure 5 is analogous to Figure 2, except that in Figure 5 the account allocations $\{x_1, x_2\}$ are the same for every agent (because an agent's β-type is not known by the government). Figure 5 has two panels. Panel A uses a vertical axis scale that encompasses all of the data, including the large welfare improvements for low-β types. Panel B uses a truncated scale that improves

[4]The mean of the continuous density is exactly 0.7. But we discretize the density, so that the actual probabilities on B are

$$\frac{f(\beta)}{\sum_{\beta \in B} f(\beta)}.$$

This discretized density has mean 0.73.

the visibility of welfare fluctuations for the high-β types. Unlike the low-β types, who experience monotonically rising welfare as π_2 rises, the high-β types have welfare that rises with π_2 for low levels of π_2, but then peaks and starts falling with π_2. This single-peaked property is driven by penalty payments, which are disproportionately made by low-β types (see Figure 7 below). Low-β types make larger and larger penalty payments as π_2 rises from 0 to intermediate values. High-β types are the implicit beneficiaries of these penalty payments. Eventually, rising π_2 squelches the penalty payments of low-β types, suppressing the cross-subsidy to high-β types, and thereby causing the welfare of high-β types to start falling. By the time π_2 reaches 1, the cross-subsidy has been completely eliminated, and high-β types are slightly worse off than they were when $\pi_2 = 0$. On a money metric basis the $\beta = 1$ types have a welfare loss equivalent to -0.23% of their income (approximately 1/2 the vertical distance on Figure 5), as they moved from $\pi_2 = 0$ to $\pi_2 = 1$. However, on a money-metric basis, the $\beta = 0.1$ types have gained welfare equivalent to 71.35% of their income.

Figure 6 aggregates welfare for the 10 types of agents, $\beta \in B = \{0.1, 0.2, ..., 1.0\}$, using the population weights in Figure 4. Figure 6 reveals that the enormous welfare gains for low-β types swamp the modest welfare losses for high-β types, an example of assymetric paternalism (Camerer et al 2003). Accordingly, the optimal policy (in this two-account system) is to have the partially illiquid account be fully illiquid: $\pi_2 = 1$.

Figure 7 reports the gross penalties paid by all β-types. As anticipated above, the penalties are hump-shaped in π_2, and the penalties are overwhelmingly paid by the low-β households.

Figure 8 reports the accounts allocations chosen by the planner as a function of π_2. Note that the account allocations asymptote to a nearly 50/50 split, implying that the fully illiquid account is used aggressively to achieve population-wide consumption smoothing. By implication, the high-β types are being constrained for the good of the low-β types.

Table 2 shows that the welfare implications of Figures 5-8 are not specific to the two account system. The first row of Table 2 reports population weighted welfare for the case of

a fully liquid account and a fully illiquid account (with optimized account allocations). This is essentially the socially optimized welfare in Figures 5-8 (with $\pi_2 = 1$). Welfare in this row (and all of the rows to follow) is expressed in terms of the money metric gain relative to population-weighted welfare in the case with only one fully liquid account.

The second row of Table 2 reports population weighted welfare for the case of a fully liquid account, a fully illiquid account, *and* a third account that has a flexible intermediate penalty: $0 < \pi_3 < 1$. We find that the optimal value of the penalty on this third account is 9%, remarkably close to the penalties that we see in 401(k) and IRA accounts. Table 2 reveals that the addition of this third account does almost nothing for welfare, which increases by 0.018% of income (18/1000ths of 1% of income). Accordingly, a 401(k)/IRA account is socially optimal in our model (i.e., it raises welfare), but its welfare consequences are de minimis relative to a system with a fully liquid account and a fully illiquid account (row 1). The balances in the account with a 9% penalty represent 14% of partially and fully illiquid assets. This percentage compares favorably to the one in the US economy (treating Social Security claims and DB claims as perfectly illiquid retirement savings and 401(k)/IRA claims as partially illiquid retirement savings).

The third row of Table 2 reports population-weighted welfare in the case with two fully flexible accounts – i.e., no account has a penalty that is ex-ante pinned down. We find that the welfare gain (relative to the system with one fully liquid account and one fully illiquid account) is 0.015% of income (15/1000ths of 1% of income). Hence, the account system in row three generates essentially no welfare gain relative to the case of one fully liquid and one fully illiquid account.

The fourth row of Table 2 reports population-weighted welfare in the case with three fully flexible accounts – i.e., no account has a penalty that is ex-ante pinned down. We find that the welfare gain (relative to the system with one fully liquid account and one fully illiquid account) is 0.020% of income (20/1000ths of 1% of income). Hence, the account system in row four also generates essentially no welfare gain relative to the case of one fully liquid and one fully illiquid account.

We conclude that one fully liquid account and one fully liquid account generates essentially the same welfare gains as more systems with more flexibility and more accounts.

5.2 Leakage

We can also use this model to study leakge. The most empirically relevant case is the one in which we have three accounts: a fully liquid account ($\pi_1 = 0$), , a partially illiquid account with optimized penalty π_2, and a fully illiquid account ($\pi_3 = 1$). When we optimize this system, we obtain $\pi_2 = 0.09$. (This corresponds to row 2 in Table 2.) This case (which approximately achieves the social optimum for the family of N-account models), admits penalized leakage from the second account. For this case we find that 74% of the dollars allocated to the second account 'leak' from that account in period 1 (and incur a penalty of π_2).

This high leakage rate is even higher than the leakage rate reported in the U.S. system (where the early withdrawal penalty is 10%). One potential explanation for the difference is that dollars in the model are deposited by government fiat, whereas many of the dollars in the U.S. system are voluntarily deposited into the 401(k)/IRA system, implying that they are coming from households with higher β values in the first place.

Our robustness analysis reveals that high leakage is a robust feature of our model. None of our dozens of calibrations generate leakage rates below 40% (see Appendix C).

5.3 Optimal policy with transfers and heterogeneous present bias: robustness.

In the previous subsection, we described a benchmark calibration in the economy with inter-household transfers and heterogeneous present bias. Two 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 (and extreme) two-account system.

2. If a third account is added, its optimized early-withdrawal penalty is 9%.

3. The equilibrium leakage rate from the third account is 74%.

In the current subsection, we document the robustness of these three findings. With respect to the first finding, the largest welfare gain that we generate in our robustness checks (by extending the system of savings accounts beyond one perfectly liquid account and one perfectly illiquid account) is 0.06% of income (6/100ths of 1%).

With respect to the second finding, the optimized penalty on the partially illiquid account ranges from 6% to 13% across our calibrated economies.

With respect to the third finding, the equilibrium leakage rate ranges from 65% to 90%.

The actual results are reported in the three panels of Table 3, which reports the results for (i) the two account system $\pi_1 = 0$ and $\pi_2 = 1$ and (ii) the three-account system with $\pi_1 = 0$, $0 < \pi_2 < 1$, and $\pi_3 = 1$. In every table we report the vector of penalties, the leakage rate (where appropriate), and welfare using a money metric improvement relative to the case of only one (fully liquid) account.

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

Table 3b varies the shape of the density of θ. First we vary the standard deviation in the distribution of taste shifters (σ). In our benchmark calibration we set $\sigma = 1/5$. In Table 3b we study the case $\sigma = 1/4$ and the case $\sigma = 1/6$. Second, we vary the support of the distribution of taste shifters $[1 - \chi, 1 + \chi]$. In our benchmark calibration we set $\chi = 2/3$. In Table 3b we also study the case $\chi = 1/3$.

Table 3c varies the mean and standard deviation of the distribution of β values. In our benchmark calibration we set $E[\beta] = 0.73$ and $\sigma_\beta = 0.23$. In Table 3c we study the case $E[\beta] = 0.70$ and $\sigma_\beta = 0.25$ and the case $E[\beta] = 0.79$ and $\sigma_\beta = 0.20$.

6 Conclusions and Directions for Future Work

Three findings emerge from the analysis of our stylized two-period model for the case of heterogeneous present bias (which allows for mechanisms that admit interpersonal transfers):

1. The constrained-efficient social optimum is well-approximated by a two-account sys-

tem, 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. Stated differently, the two account system identified in Amador, Werning and Angeletos (2006) turns out to be approximately optimal.

2. If a third account is added, its optimized early-withdrawal penalty is between 6% and 13%.

3. In equilibrium, the leakage rate from this (partially illiquid) third account ranges from 65% to 90%.

These properties have analogs in the retirement savings systems in the US. The US has fully liquid accounts (i.e., a standard checking account), fully illiquid accounts (i.e., Social Security), and a partially illiquid defined contribution system with a 10% penalty for early withdrawals. This partially illiquid DC system has a leakage rate of 40%.

Despite these similarities, it is inappropriate to conclude that our findings demonstrate the social optimality of the US system. Our simulation framework is highly stylized. For example, we assume only two periods (e.g., working life and retirement). We assume a particular form of multiplicative taste shifter.[5] We assume that households are naive with respect to their present bias. We study a limited set of distributions of the present bias parameter, β.[6] We only study N-account retirement savings systems (instead of studying a fully general mechanism design framework).[7]

Much more robustness work is needed to interrogate the three findings that we have generated. It is not yet clear whether our results – which, to our surprise, seem to rationalize the fundamental institutional structure of the US retirement savings system – will continue to hold as we enrich and expand our analysis.

[5] We assume $\theta u(c)$ but we could have instead assumed $u(c-\theta)$. In onging work, we are studying this case.

[6] Little is known about the distribution of present bias.

[7] Future drafts of this paper will contain such analysis. We are able to show that the fully general mechanism design solution is well-approximated (in terms of welfare) by our two account system.

7 References

Amador, Manuel, Ivan Werning, and George-Marios Angeletos, 2006. "Commitment vs. Flexibility." *Econometrica* 74, pp. 365-396.

Ambrus, Attila, and Georgy Egorov. 2013. "Comment on "Commitment vs. flexibility" *Econometrica* 81.5: 2113-2124.

Angeletos, George-Marios, David Laibson, Andrea Repetto, Jeremy Tobacman, Stephen Weinberg, 2001. "The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation." Journal of Economic Perspectives 15(3), pp. 47-68.

Argento, Robert, Victoria L. Bryant and John Sabelhaus, forthcoming, "Early Withdrawals from Retirement Accounts During the Great Recession," *Contemporary Economic Policy.*

Ariely, Dan and Klaus Wertenbroch, 2002. "Procrastination, Deadlines, and Performance: Self-Control by Precommitment." Psychological Science 13, pp. 219-224.

Ashraf, Nava, Dean Karlan, and Wesley Yin, 2006. "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Phillipines." Quarterly Journal of Economics 121, pp. 635-672.

Atkeson, Andrew, and Robert E. Lucas. "On efficient distribution with private information." *The Review of Economic Studies* 59.3 (1992): 427-453.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger, 2012. "Working Over Time: Dynamic Inconsistency in Real Effort Tasks." Working paper.

Bernheim, B. Douglas and Antonio Rangel, 2004. "Addiction and Cue-Triggered Decision Processes." American Economic Review 94, pp. 1558-1590.

Beshears, John, James Choi, Christopher Harris, David Laibson, Brigitte Madrian, Jung Sakong. 2015. "Self Control and Commitment: Can Decreasing the Liquidity of a Savings Account Increase Deposits?" NBER Working Paper #21474

Colin Camerer, Samuel Issacharoff, George Loewenstein, Ted O'Donoghue & Matthew Rabin. 2003. "Regulation for Conservatives: Behavioral Economics and the Case for 'Asymmetric Paternalism'." 151 University of Pennsylvania Law Review 101: 1211–1254.

Chow, Vinci Y. C. "Demand for a Commitment Device in Online Gaming." Working paper.

DellaVigna, Stefano and Ulrike Malmendier, 2006. "Paying Not to Go to the Gym." American Economic Review 96, pp. 694-719.

DellaVigna, Stefano and Daniele Paserman, 2005. "Job Search and Impatience." Journal of Labor Economics 23, pp. 527-588.

Dupas, Pascaline and Jonathan Robinson, forthcoming. "Why Don't the Poor Save More? Evidence from Health Savings Experiments." American Economic Review.

Fudenberg, Drew and David K. Levine, 2006. "A Dual-Self Model of Impulse Control." American Economic Review 96, pp. 1449-1476.

Gine, Xavier, Dean Karlan, and Jonathan Zinman, 2010. "Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation." American Economic Journal: Applied Economics 2, pp. 213-235.

Gul, Faruk and Wolfgang Pesendorfer, 2001. "Temptation and Self-Control." Econometrica 69, pp. 1403-1435.

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

Houser, Daniel, Daniel Schunk, Joachim Winter, and Erte Xiao, 2010. "Temptation and Commitment in the Laboratory." Working paper.

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan, 2010. "Self-Control at Work: Evidence from a Field Experiment." Working paper.

Laibson, David, 1997. "Golden Eggs and Hyperbolic Discounting." Quarterly Journal of Economics 112, pp. 443-477.

Laibson, David, Andrea Repetto, and Jeremy Tobacman, 2007. "Estimating Discount Functions with Consumption Choices over the Lifecycle." Working paper.

Loewenstein, George, 1996. "Out of Control: Visceral Influences on Behavior." Organizational Behavior and Human Decision Processes 65, pp. 272-292.

Loewenstein, George and Drazen Prelec, 1992. "Anomalies in Intertemporal Choice: Evidence and an Interpretation." Quarterly Journal of Economics 107, pp. 573-597.

McClure, Samuel M., David Laibson, George Loewenstein, and Jonathan D. Cohen, 2004. "Separate Neural Systems Value Immediate and Delayed Monetary Rewards." Science 306, pp. 503-507.

McClure, Samuel M., Keith M. Ericson, David Laibson, George Loewenstein, and Jonathan D. Cohen, 2007. "Time Discounting for Primary Rewards." Journal of Neuroscience 27, pp. 5796-5804.

Meier, Stephan and Charles Sprenger, 2010. "Present-Biased Preferences and Credit Card Borrowing." American Economic Journal: Applied Economics 2, pp. 193-210.

Milkman, Katherine L., Julia A. Minson, and Kevin G. M. Volpp, 2012. "Holding the Hunger Games Hostage at the Gym: An Evaluation of Temptation Bundling." Working paper.

Milkman, Katherine L., Todd Rogers, and Max H. Bazerman, 2008. "Harnessing Our Inner Angels and Demons: What We Have Learned About Want/Should Conflicts and How That Knowledge Can Help Us Reduce Short-Sighted Decision Making." Perspectives on Psychological Science 3, pp. 324-338.

Milkman, Katherine L., Todd Rogers, and Max H. Bazerman, 2009. "Highbrow Films Gather Dust: Time-Inconsistent Preferences and Online DVD Rentals." Management Science 55, pp. 1047-1059.

Milkman, Katherine L., Todd Rogers, and Max H. Bazerman, 2010. "I'll Have the Ice Cream Soon and the Vegetables Later: A Study of Online Grocery Purchases and Order Lead Time." Marketing Letters 21, pp. 17-36.

Moser, Christian and Olea de Souza e Silva, Pedro,2017. "Optimal Paternalistic Savings Policies" Columbia Business School Research Paper No. 17-51. Available at SSRN: https://ssrn.com/abstract=2959844 or http://dx.doi.org/10.2139/ssrn.2959844

O'Connor, Kathleen M., Carsten K. W. De Dreu, Holly Schroth, Bruce Barry, Terri R. Lituchy, and Max H. Bazerman, 2002. "What We Want To Do Versus What We Think We Should Do: An Empirical Investigation of Intrapersonal Conflict." Journal of Behavioral Decision Making 15, pp. 403-418.

O'Donoghue, Ted and Matthew Rabin, 1999. "Doing It Now or Later." American Economic Review 89, pp. 103-124.

Oster, Sharon M. and Fiona M. Scott Morton, 2005. "Behavioral Biases Meet the Market: The Case of Magazine Subscription Prices." Berkeley Electronic Press: Advances in Economic Analysis & Policy 5(1).

Phelps, E. S. and R. A. Pollak, 1968. "On Second-Best National Saving and Game-Equilibrium Growth." Review of Economic Studies 35, pp. 185-199.

Read, Daniel and Barbara van Leeuwen, 1998. "Predicting Hunger: The Effects of Appetite and Delay on Choice." Organizational Behavior and Human Decision Processes 76, pp. 189-205.

Read, Daniel, George Loewenstein, and Shobana Kalyanaraman, 1999. "Mixing Virtue and Vice: Combining the Immediacy Effect and the Diversification Heuristic." Journal of Behavioral Decision Making 12, pp. 257-273.

Redelmeier, Donald A. and Eldar Shafir, 1995. "Medical Decision Making in Situations That Offer Multiple Alternatives." Journal of the American Medical Association 273, pp. 302-305.

Reuben, Ernesto G., Paola Sapienza, and Luigi Zingales, 2009. "Procrastination and Impatience." Working paper.

Royer, Heather, Mark Stehr, and Justin Sydnor, 2012. "Incentives, Commitments and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company." Working paper.

Shafir, Eldar, Itamar Simonson, and Amos Tversky. "Reason-based choice."Cognition 49.1 (1993): 11-36.

Shapiro, Jesse, 2005. "Is There a Daily Discount Rate? Evidence from the Food Stamp Nutrition Cycle." Journal of Public Economics 89, pp. 303-325.

Simonson, Itamar, 1989. "Choice Based on Reasons: The Case of Attraction and Compromise Effects." Journal of Consumer Research 16, pp. 158-174.

Steyer, Robert, 2011. "DC Plan Leakage Problem Alarming, Solutions Evasive." Pensions & Investments, April 4, 2011. Available at http://www.pionline.com/apps/pbcs.dll/article?AID=/20110404 accessed April 18, 2011.

Strotz, R. H., 1955. "Myopia and Inconsistency in Dynamic Utility Maximization." Review of Economic Studies 23, pp. 165-180.

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

Wertenbroch, Klaus, 1998. "Consumption Self-Control by Rationing Purchase Quantities of Virtue and Vice." Marketing Science 17, pp. 317-337.

Table 1: Homogeneous Welfare Gains Relative to Decentralized (% Wealth Equivalent)

	Value of β									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1 Liquid, 1 Flexible	71.65	32.67	17.54	9.80	5.40	2.81	1.29	0.45	0.07	0
1 Liquid, 2 Flexible	71.67	32.72	17.63	9.89	5.48	2.86	1.31	0.46	0.07	0
General Mechanism	71.68	32.75	17.66	9.93	5.51	2.88	1.33	0.46	0.07	0.02
1 Liquid, 1 Illiquid	71.65	32.66	17.49	9.68	5.2	2.54	1.02	0.26	0.01	0

Table 2: Heterogeneous Welfare Gains (% Wealth Equivalent)

	Relative to Decentralized	Relative to 1 Liquid, 1 Illiquid
1 Liquid, 1 Illiquid	3.397	0
1 Liquid, 1 Flexible, 1 Illiquid	3.415	0.018
2 Flexible	3.412	0.015
3 Flexible	3.417	0.020

Table 3: Heterogeneous Robustness

	Parameter Varied	1 Liquid, 1 Illiquid	1 Liquid, 1 Flexible, 1 Illiquid		
		Welfare Gain	Welfare Gain	π_2^*	Leakage Rate (%)
Baseline	—	3.397	3.415	0.09	74
Vary γ	$\gamma = \frac{1}{3}$	4.537	4.599	0.09	75
	$\gamma = \frac{1}{2}$	4.601	4.641	0.09	74
	$\gamma = 2$	1.899	1.907	0.13	65
	$\gamma = 3$	1.279	1.285	0.09	73
Vary $f(\theta)$	$\chi = \frac{1}{3}$	3.622	3.632	0.06	90
	$\sigma = \frac{1}{2}$	3.050	3.074	0.09	67
	$\chi = \frac{1}{3}$, $\sigma = \frac{1}{2}$	3.569	3.586	0.11	72
Vary $g(\theta)$	$E(\beta) = 0.66$, $sd(\beta) = 0.26$	5.854	5.871	0.1	79
	$E(\beta) = 0.79$, $sd(\beta) = 0.20$	1.862	1.879	0.07	70

The table benchmarks against the Baseline heterogeneous case, which has parameters $\gamma = 1$, $\chi = \frac{2}{3}$, $\sigma = \frac{1}{2}$, $E(\beta) = 0.73$, *and* $sd(\beta) = 0.23$. *Each variation considered varies only the parameters specified in the "Simulation Details" column.*

Figure 1: $\beta = 0.7$ **Homogeneous Utility Against** π_2

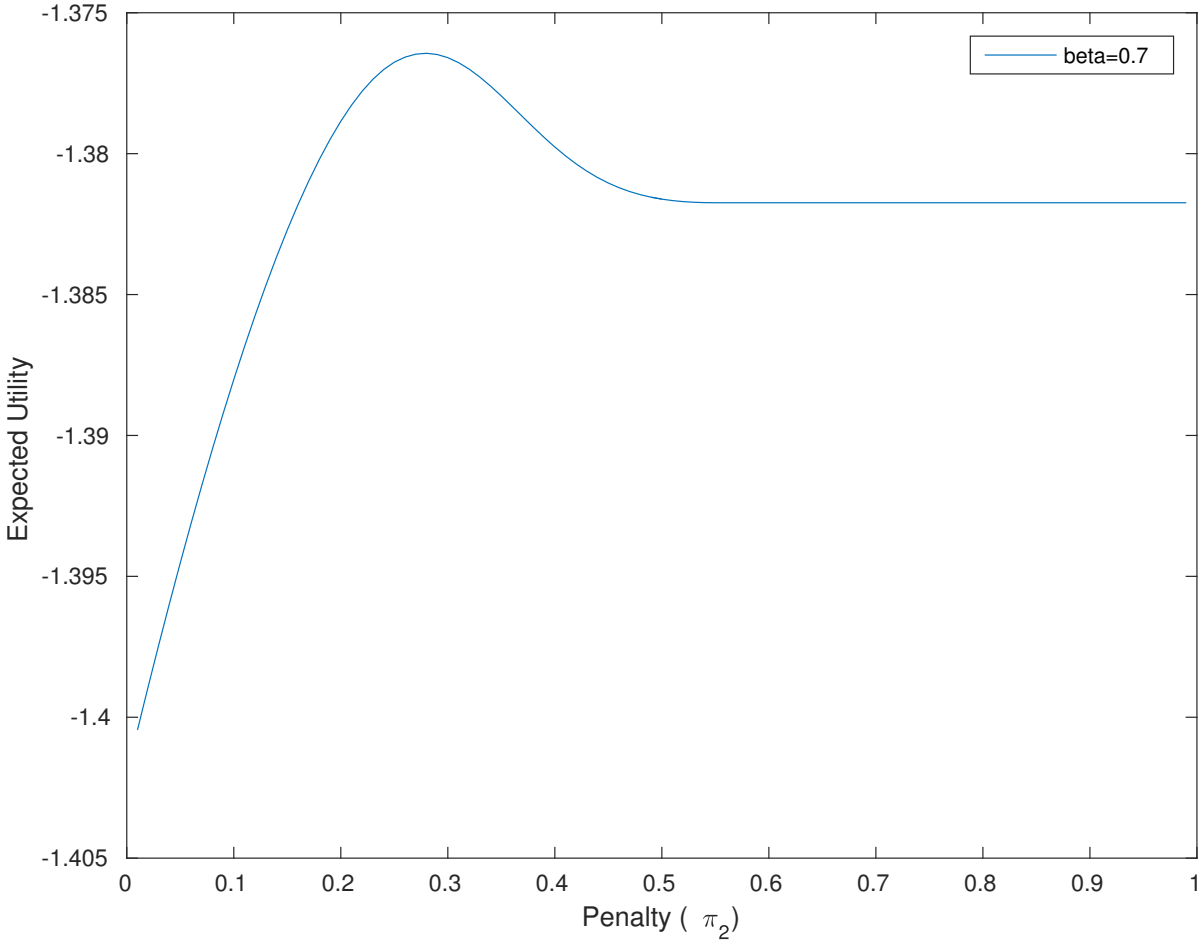


Figure 2: $\beta \in \{0.1, ..., 1.0\}$ **Homogeneous Utility Against** π_2

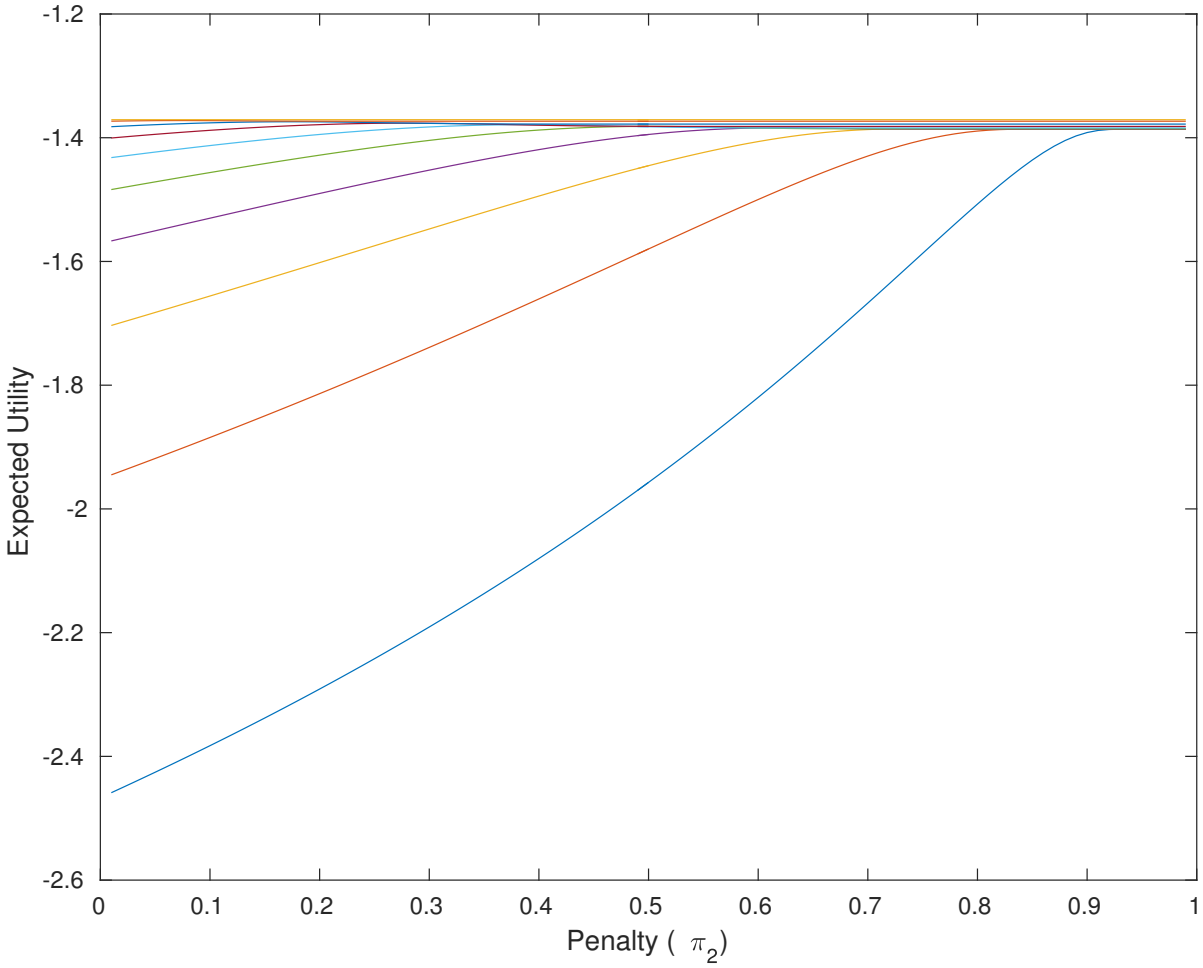


Figure 3: Optimal Homogeneous Penalty π_2^* against β

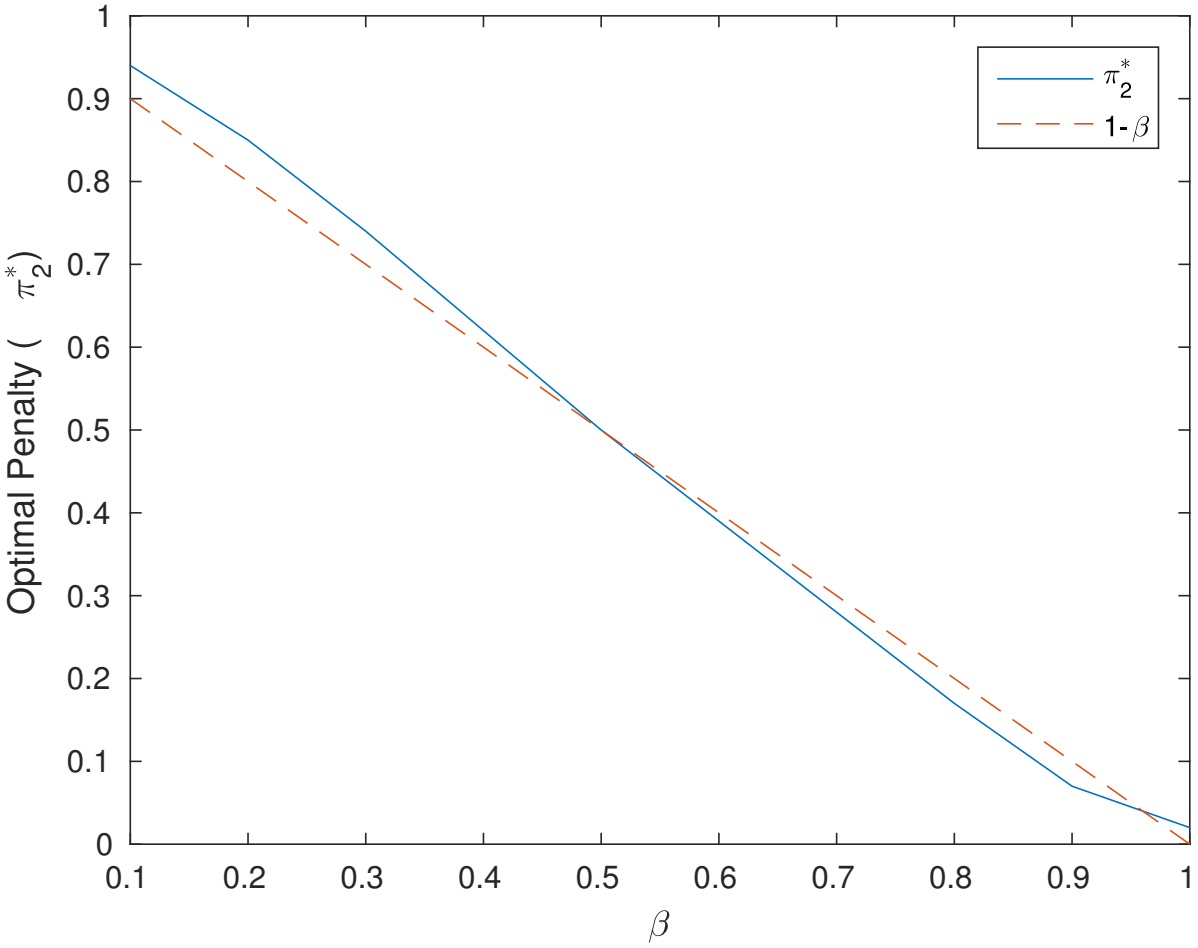


Figure 4: Heterogeneous Population Density $g(\beta)$

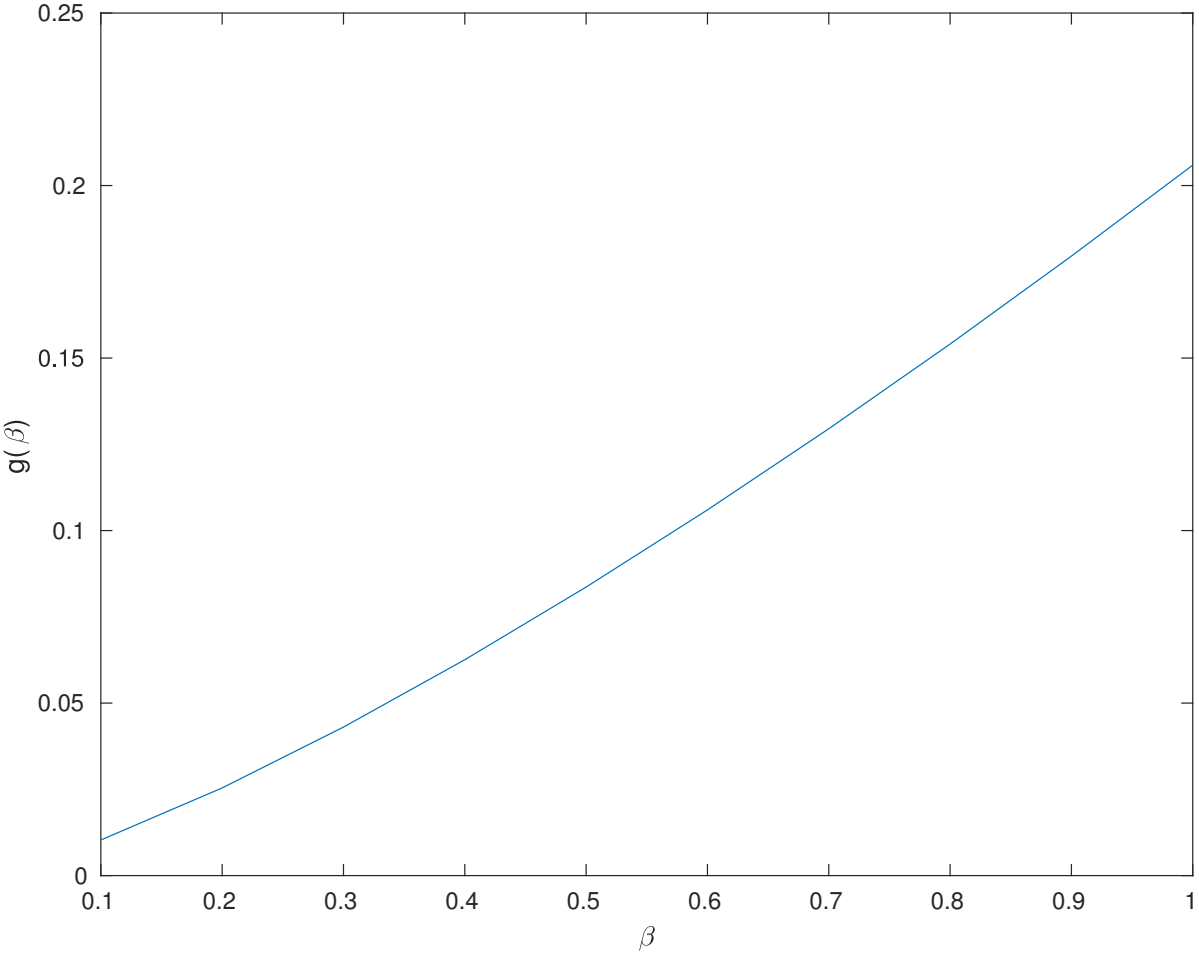


Figure 5: Heterogeneous Population: Subpopulation Utilities against π_2

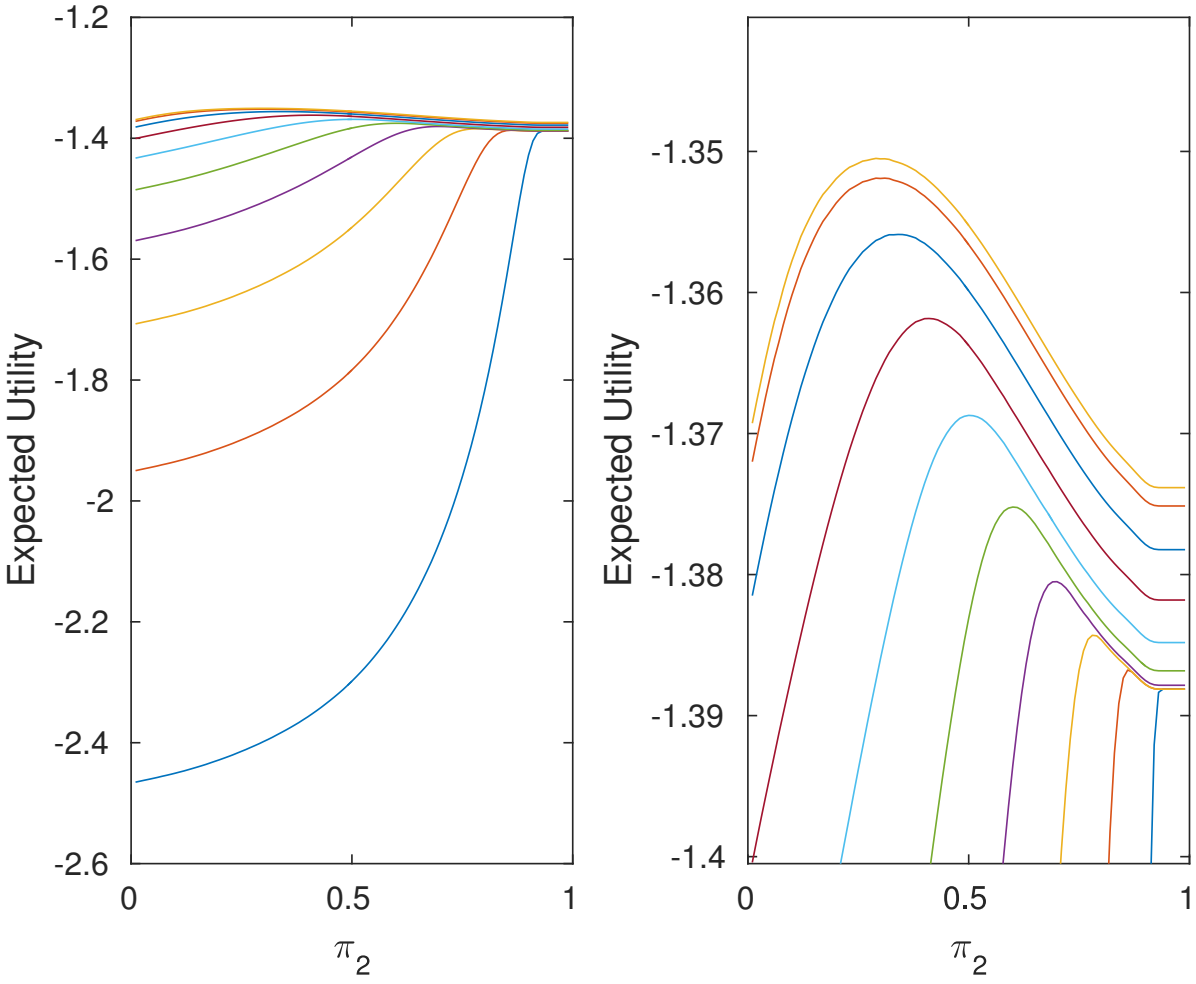


Figure 6: Heterogeneous Population: Population Utility against π_2

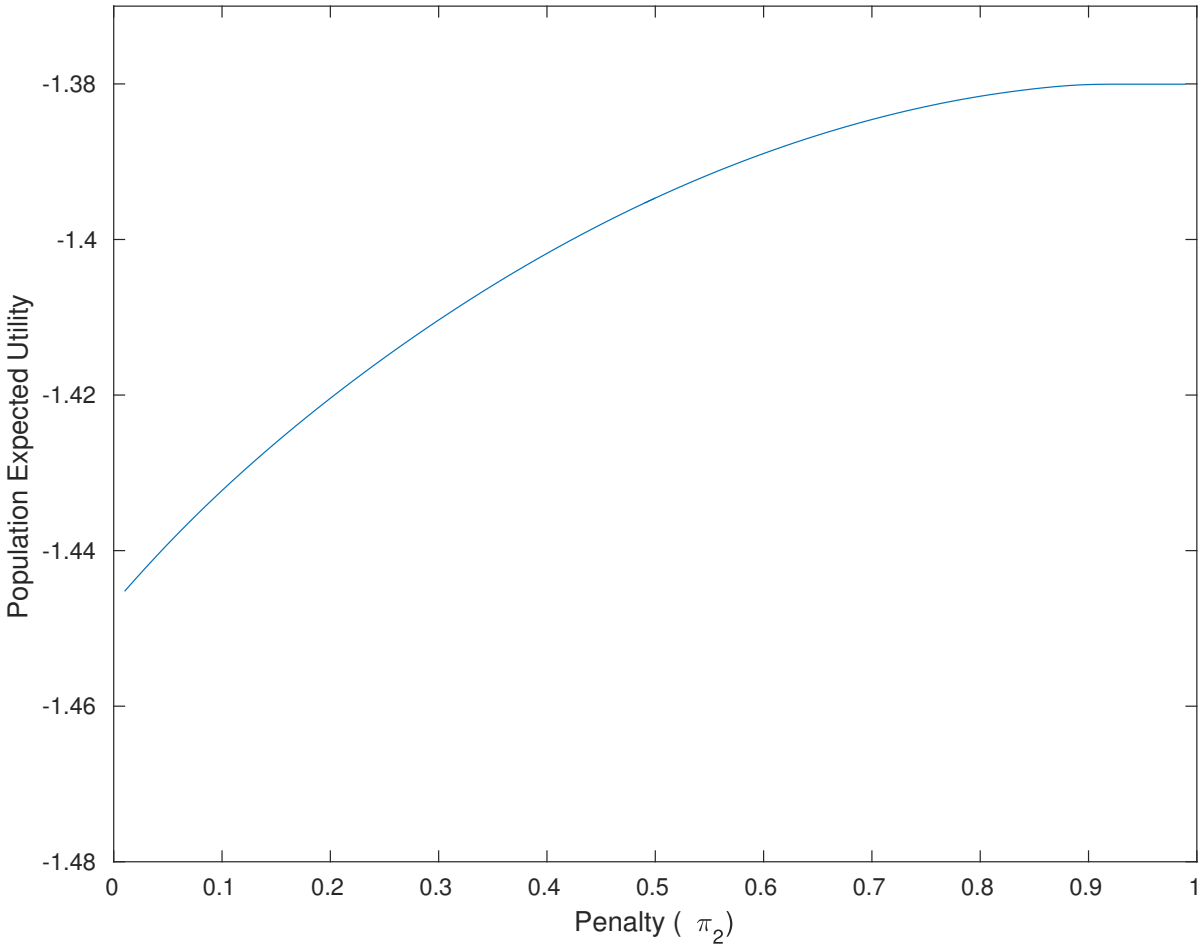


Figure 7: Heterogeneous Population: Subpopulation Penalties Paid

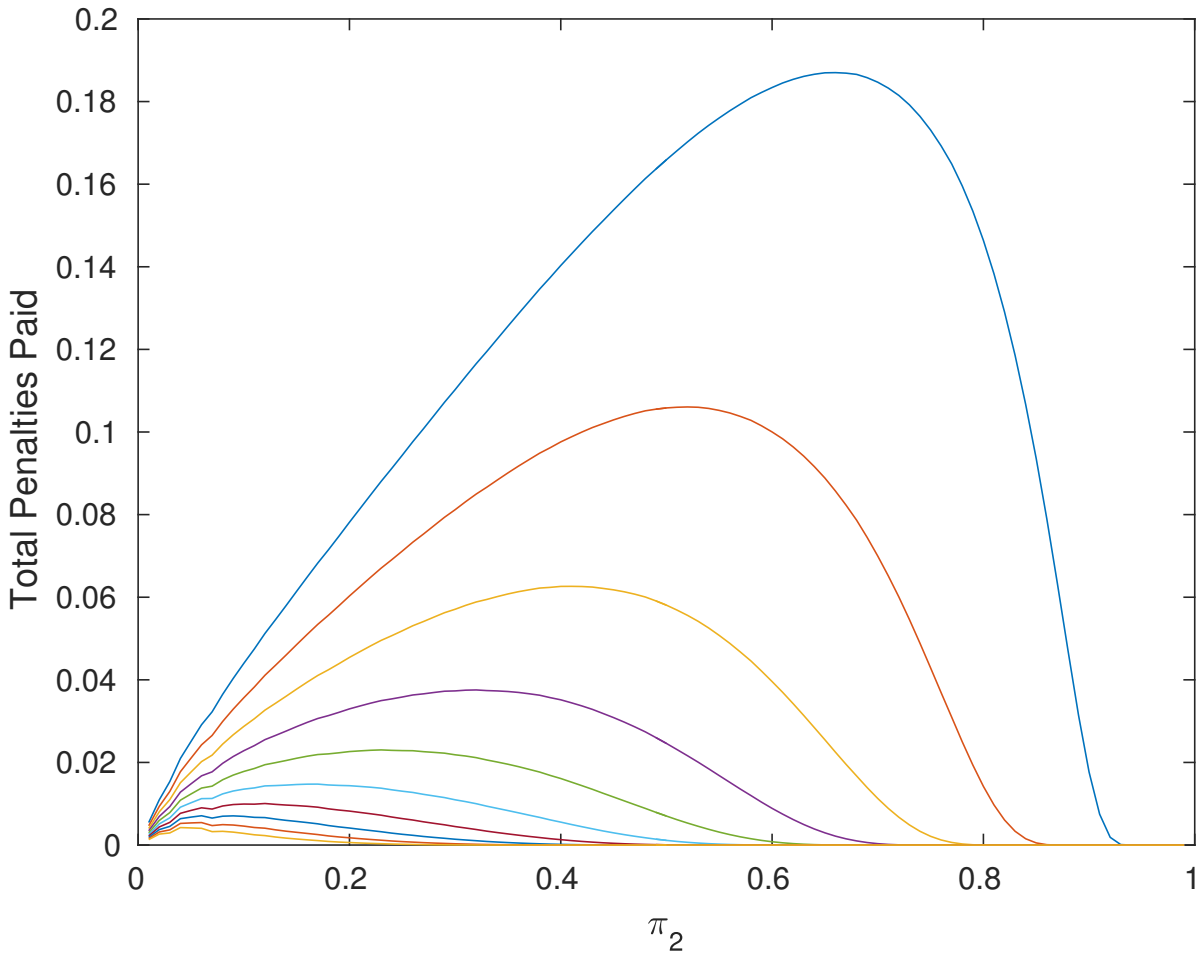
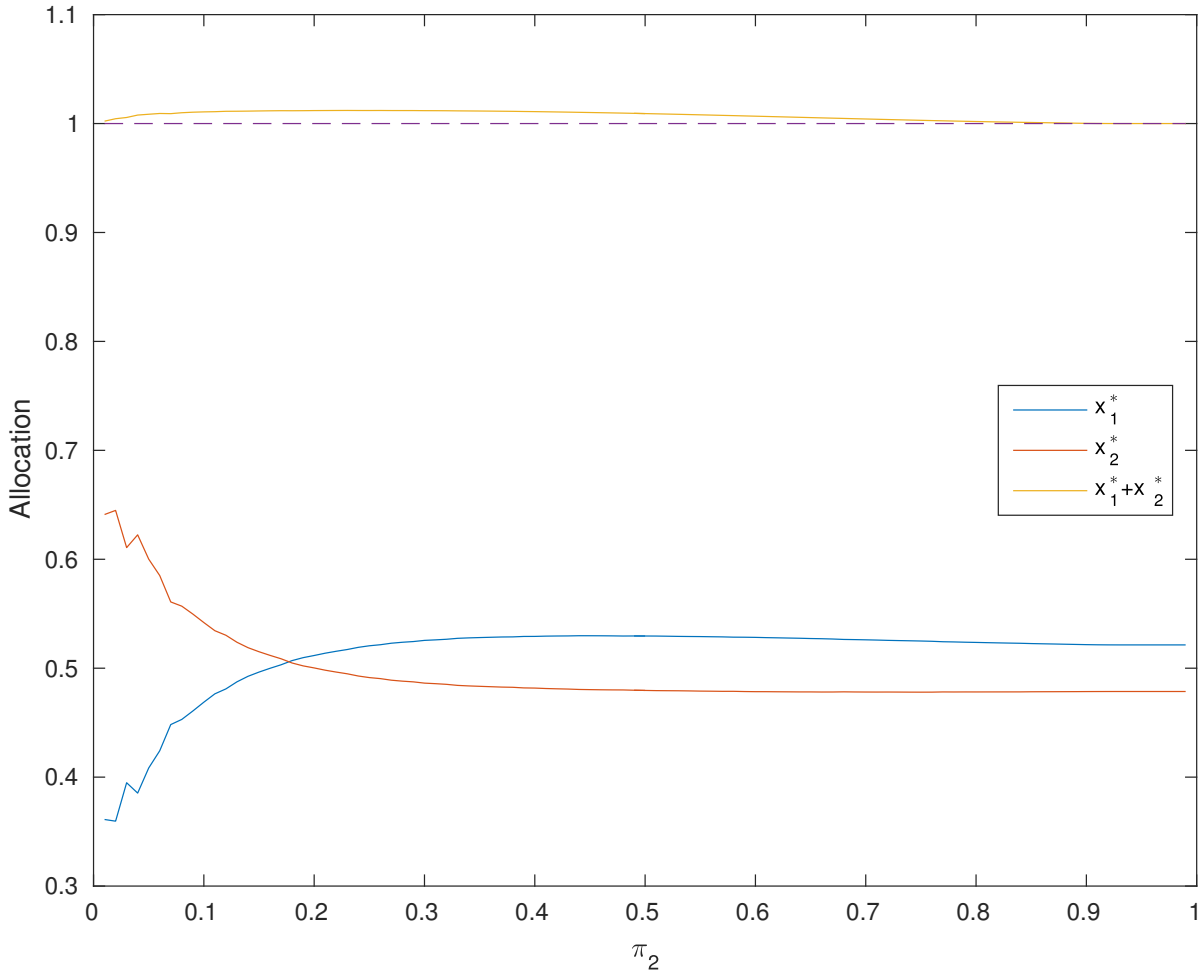


Figure 8: Heterogeneous Population: Account Allocations and Total Wealth



Is a Simple Two-Account Policy Optimal when Redistribution is Feasible?

Proof of Theorem 2

1. Introduction

We assume that: $U, W : [0, \infty) \to [-\infty, \infty)$; $U', W' > 0$ and $U'', W'' < 0$ on $(0, \infty)$; and $U'(0+) = W'(0+) = \infty$.

We make the following assumptions on the distribution function F of the taste shock θ:

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

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

A3′ F' is bounded away from 0 on $\left(\underline{\theta}, \overline{\theta}\right)$. In particular, $F'(\underline{\theta}+), F'(\overline{\theta}-) > 0$.

We assume that $0 < \beta < 1$.

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

Theorem 2. *Suppose that interpersonal transfers are possible. Then a two-account system with one completely liquid account and one completely illiquid account never maximizes welfare.*

2. Full Proof

If self 1 is presented with two accounts, a perfectly liquid account containing the amount $x_{\text{liquid}} > 0$ and a perfectly illiquid account (containing the amount $x_{\text{illiquid}} \geq 0$), then the outcome will depend on her type θ. There will exist $\theta_1 \in (0, \infty)$ such that: if $\theta < \theta_1$, then she consumes less than the balance x_{liquid} in her liquid account: and, if $\theta \geq \theta_1$, then she consumes the whole of x_{liquid}. This gives rise to a utility curve (u_0, w_0) given by the formulae $u_0(\theta) = U(c(\theta))$ and $w_0(\theta) = W(k(\theta))$. This utility curve is a smooth function of θ for $\theta < \theta_1$, has a kink at θ_1, and is constant for $\theta \geq \theta_1$. The idea behind the proof is to find necessary conditions for this curve to be optimal.

Remark 3. *We emphasize that there is no reason why θ_1 should lie in $\left[\underline{\theta}, \overline{\theta}\right]$.*

The first step is to formulate the optimization problem of the planner. The planner seeks to maximize social welfare

$$\int (\theta\, u(\theta) + w(\theta))\, dF(\theta)$$

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

$$\int (y - C(u(\theta)) - K(w(\theta)))\, dF(\theta) \geq 0. \tag{BC}$$

Incentive compatibility breaks down into two parts, a linear part

$$\theta\, u' + \beta\, w' = 0 \tag{ICL}$$

and a monotonic part

$$w' \leq 0. \tag{ICM}$$

We parameterize candidate solutions to this problem in terms of $u(\overline{\theta})$, $w(\overline{\theta})$ and a piecewise continuous function $u' : \Theta \to \mathbb{R}$. More precisely, if $\theta_1 \in \left(\underline{\theta}, \overline{\theta}\right)$ then we require that:

1. u' is continuous in the left-hand interval $[\underline{\theta}, \theta_1)$;
2. $u'(\theta_1-)$ (the limit from the left of u' at θ_1) exists;
3. u' is continuous in the right-hand interval $\left(\theta_1, \overline{\theta}\right]$;
4. $u'(\theta_1+)$ (the limit from the right of u' at θ_1) exists.

We then:

1. put $w' = -\frac{\theta}{\beta}\, u'$;
2. let $u : \Theta \to \mathbb{R}$ be the function with derivative u' that takes the value $u(\overline{\theta})$ at $\overline{\theta}$ (as the notation already suggests we should);
3. let $w : \Theta \to \mathbb{R}$ be the function with derivative w' that takes the value $w(\overline{\theta})$ at $\overline{\theta}$ (as the notation already suggests we should).

If $\theta_1 \in (0, \underline{\theta}]$ or $\theta_1 \in \left[\overline{\theta}, \infty\right)$, then we require that u' be continuous on the whole of $\left[\underline{\theta}, \overline{\theta}\right]$.

The third step is to formulate the Lagrangean. In the case $\theta_1 \in \left(\underline{\theta}, \overline{\theta}\right)$, this can be written

$$\begin{aligned} L(u, w, \lambda, \nu_L, \nu_R) &= \int (\theta\, u(\theta) + w(\theta))\, dF(\theta) \\ &\quad + \lambda \int (y - C(u(\theta)) - K(w(\theta)))\, dF(\theta) \\ &\quad - \int_{[\underline{\theta}, \theta_1]} w'_L(\theta)\, d\nu_L(\theta) - \int_{\left[\theta_1, \overline{\theta}\right]} w'_R(\theta)\, d\nu_R(\theta), \end{aligned}$$

where u and w are detemined by $u(\overline{\theta})$, $w(\overline{\theta})$ and u' as described above, λ is the scalar multiplier on the aggregate budget constraint, w'_L is the extension of $w'|_{[\underline{\theta}, \theta_1)}$ to $[\underline{\theta}, \theta_1]$, w'_R is the extension of $w'|_{\left(\theta_1, \overline{\theta}\right]}$ to $\left[\theta_1, \overline{\theta}\right]$, ν_L is a finite non-negative Borel measure on $[\underline{\theta}, \theta_1]$ and ν_R is a finite non-negative Borel measure on $\left[\theta_1, \overline{\theta}\right]$.

Remark 4. *Notice that the Lagrangean no longer includes terms corresponding to (ICL), since we are working with a reduced form.*

A necessary condition for (u_0, w_0) to be optimal is that there exist multipliers λ, ν_L and ν_R such that three sets of conditions hold. First, the derivative of $L(u, w, \lambda, \nu_L, \nu_R)$ at (u_0, w_0) in the direction (u, w) is 0 for all (u, w). That is,

$$\begin{aligned} 0 &= \int (\theta\, u + w)\, dF - \lambda \int (C'(u_0)\, u + K'(w_0)\, w)\, dF \\ &\quad - \int_{[\underline{\theta}, \theta_1]} w'_L(\theta)\, d\nu_L(\theta) - \int_{\left[\theta_1, \overline{\theta}\right]} w'_R(\theta)\, d\nu_R(\theta) \end{aligned} \tag{1}$$

for all feasible (u, w). Second, the constraints must all be satisfied. That is,

$$\begin{aligned} 0 &= \int (y - C(u_0(\theta)) - K(w_0(\theta)))\, dF(\theta), \\ 0 &\leq w'_{0L}, \\ 0 &\leq w'_{0R}. \end{aligned}$$

Third, constraint qualification must hold. That is,

$$0 = \int_{[\underline{\theta}, \theta_1]} w'_{0L}(\theta)\, d\nu_L(\theta), \tag{2}$$

$$0 = \int_{\left[\theta_1, \overline{\theta}\right]} w'_{0R}(\theta)\, d\nu_R(\theta). \tag{3}$$

Now, we can rearrange (1) in terms of the underlying parameters $u(\overline{\theta})$, $w(\overline{\theta})$ and u' as follows. Putting $\overline{F}(\theta) = \int_{[\underline{\theta}, \theta]} F(t)\, dt$, and noting that $\theta\, F - \overline{F}$ and u are both

continuous, we can integrate by parts to obtain

$$\begin{aligned}\int \theta\, u\, dF(\theta) &= \big[(\,\theta\, F - \overline{F}\,)\, u\big]_{\underline{\theta}-}^{\overline{\theta}} - \int (\,\theta\, F - \overline{F}\,)\, u'\, d\theta \\ &= \big(\overline{\theta}\, F(\,\overline{\theta}\,) - \overline{F}(\,\overline{\theta}\,)\big)\, u(\,\overline{\theta}\,) - \int (\,\theta\, F - \overline{F}\,)\, u'\, d\theta \\ &= \big(\overline{\theta}\, F(\,\overline{\theta}\,) - \overline{F}(\,\overline{\theta}\,)\big)\, u(\,\overline{\theta}\,) \\ &\quad - \int_{[\,\underline{\theta},\theta_1\,]} (\,\theta\, F - \overline{F}\,)\, u'_L\, d\theta - \int_{\left[\,\theta_1,\overline{\theta}\right]} (\,\theta\, F - \overline{F}\,)\, u'_R\, d\theta.\end{aligned}$$

Similarly,

$$\begin{aligned}\int w\, dF(\theta) &= [F\, w]_{\underline{\theta}-}^{\overline{\theta}} - \int F\, w'\, d\theta \\ &= F(\,\overline{\theta}\,)\, w(\,\overline{\theta}\,) - \int F\, w'\, d\theta \\ &= F(\,\overline{\theta}\,)\, w(\,\overline{\theta}\,) \\ &\quad - \int_{[\,\underline{\theta},\theta_1\,]} F\, w'_L\, d\theta - \int_{\left[\,\theta_1,\overline{\theta}\right]} F\, w'_R\, d\theta.\end{aligned}$$

Next, putting $\Lambda_u(\theta) = \int_{[\underline{\theta},\theta]} C'(u_0(t))\, dF(t)$, we have

$$\begin{aligned}-\lambda \int C'(u_0)\, u\, dF &= -\int u\, \lambda\, \Lambda'_u\, d\theta \\ &= -\,[u\, \lambda\, \Lambda_u]_{\underline{\theta}-}^{\overline{\theta}} + \int \lambda\, \Lambda_u\, u'\, d\theta \\ &= -u(\,\overline{\theta}\,)\, \lambda\, \Lambda_u(\,\overline{\theta}\,) + \int \lambda\, \Lambda_u\, u'\, d\theta \\ &= -u(\,\overline{\theta}\,)\, \lambda\, \Lambda_u(\,\overline{\theta}\,) \\ &\quad + \int_{[\,\underline{\theta},\theta_1\,]} \lambda\, \Lambda_u\, u'_L\, d\theta + \int_{\left[\,\theta_1,\overline{\theta}\right]} \lambda\, \Lambda_u\, u'_R\, d\theta.\end{aligned}$$

Similarly, putting $\Lambda_w(\theta) = \int_{[\underline{\theta},\theta]} K'(w_0(t))\, dF(t)$,

$$\begin{aligned}-\lambda \int K'(w_0)\, w\, dF &= -\int w\, \lambda\, \Lambda'_w\, d\theta \\ &= -\,[w\, \lambda\, \Lambda_w]_{\underline{\theta}-}^{\overline{\theta}} + \int \lambda\, \Lambda_w\, w'\, d\theta \\ &= -w(\,\overline{\theta}\,)\, \lambda\, \Lambda_w(\,\overline{\theta}\,) + \int \lambda\, \Lambda_w\, w'\, d\theta \\ &= -w(\,\overline{\theta}\,)\, \lambda\, \Lambda_w(\,\overline{\theta}\,) \\ &\quad + \int_{[\,\underline{\theta},\theta_1\,]} \lambda\, \Lambda_w\, w'_L\, d\theta + \int_{\left[\,\theta_1,\overline{\theta}\right]} \lambda\, \Lambda_w\, w'_R\, d\theta.\end{aligned}$$

Finally, the terms $-\int_{[\underline{\theta},\theta_1]} w'_L(\theta)\,d\nu_L(\theta)$ and $-\int_{[\theta_1,\overline{\theta}]} w'_R(\theta)\,d\nu_R(\theta)$ require no manipulation.

Hence, equating the coefficients of $u(\,\overline{\theta}\,)$, $w(\,\overline{\theta}\,)$, u'_L and u'_R to 0 in (1), we obtain:

$$\begin{aligned}
0 &= \overline{\theta}\,F(\,\overline{\theta}\,) - \overline{F}(\,\overline{\theta}\,) - \lambda\,\Lambda_u(\,\overline{\theta}\,), && (4)\\
0 &= F(\,\overline{\theta}\,) - \lambda\,\Lambda_w(\,\overline{\theta}\,), && (5)\\
0 &= -(\,\theta\,F - \overline{F}\,)\,d\theta + \tfrac{\theta}{\beta}\,F\,d\theta + \lambda\,\Lambda_u\,d\theta - \tfrac{\theta}{\beta}\,\lambda\,\Lambda_w\,d\theta + \tfrac{\theta}{\beta}\,d\nu_L, && (6)\\
0 &= -(\,\theta\,F - \overline{F}\,)\,d\theta + \tfrac{\theta}{\beta}\,F\,d\theta + \lambda\,\Lambda_u\,d\theta - \tfrac{\theta}{\beta}\,\lambda\,\Lambda_w\,d\theta + \tfrac{\theta}{\beta}\,d\nu_R. && (7)
\end{aligned}$$

Now, we certainly have $w'_L < 0$ on $[\,\underline{\theta},\theta_1\,]$. It therefore follows from constraint qualification (namely (2)) that $\nu_L = 0$. Equation (6) therefore implies that

$$\lambda\,(\theta\,\Lambda_w - \beta\,\Lambda_u) = \theta\,F - \beta\,(\theta\,F - \overline{F}) = (1-\beta)\,\theta\,F + \beta\,\overline{F} = \overline{G}$$

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

$$\begin{aligned}
\frac{\theta\,\Lambda_w(\theta)}{\theta-\underline{\theta}} &\to \underline{\theta}\,K'(w_0(\underline{\theta}))\,F'(\underline{\theta}+),\\
\frac{\beta\,\Lambda_u(\theta)}{\theta-\underline{\theta}} &\to \beta\,C'(u_0(\underline{\theta}))\,F'(\underline{\theta}+),\\
\frac{\overline{G}}{\theta-\underline{\theta}} &\to G(\underline{\theta}+) = (1-\beta)\,\underline{\theta}\,F'(\underline{\theta}+)
\end{aligned}$$

as $\theta \downarrow \underline{\theta}$. But, since $(u_0(\underline{\theta}), w_0(\underline{\theta}))$ is chosen freely from the ambient budget line by the $\underline{\theta}$ type, we must have

$$\frac{C'(u_0(\underline{\theta}))}{\underline{\theta}} = \frac{K'(w_0(\underline{\theta}))}{\beta}.$$

Hence, in fact,

$$\frac{\theta\,\Lambda_w(\theta) - \beta\,\Lambda_u(\theta)}{\theta-\underline{\theta}} \to 0.$$

On the other hand,

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

For any finite choice of λ, we therefore have the contradiction $0 = (1-\beta)\,\underline{\theta}\,F'(\underline{\theta}+) > 0$. This establishes that we cannot have $\theta_1 \in \big(\,\underline{\theta},\overline{\theta}\,\big)$.

Remark 5. *This is where we use the assumption $\beta < 1$.*

If $\theta_1 \in [\,\overline{\theta},\infty\,)$, then we can still derive equations (4, 5 and 6). In particular, we can still derive equation (6). We can therefore again derive a contradiction by essentially the same argument.

If $\theta_1 \in (0, \underline{\theta}]$, then we can still derive equations (4, 5 and 7). However, we can no longer derive equation (6). We therefore need new arguments. The first point to note is that, since $\theta_1 \leq \underline{\theta}$, all types $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$ choose the point that a hypothetical θ_1 type would choose from the ambient budget set. We therefore have

$$\Lambda_u(\overline{\theta}) = \int_{[\underline{\theta},\overline{\theta}]} C'(u_0(t))\, dF(t) = F(\overline{\theta})\, C'(u_0(\theta_1)), \tag{8}$$

$$\Lambda_w(\overline{\theta}) = \int_{[\underline{\theta},\overline{\theta}]} K'(w_0(t))\, dF(t) = F(\overline{\theta})\, K'(w_0(\theta_1)). \tag{9}$$

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

$$\frac{C'(u_0(\theta_1))}{\theta_1} = \frac{K'(w_0(\theta_1))}{\beta}.$$

Using (4) and (5), we therefore obtain

$$\frac{\overline{\theta}\, F(\overline{\theta}) - \overline{F}(\overline{\theta})}{F(\overline{\theta})} = \frac{\Lambda_u(\overline{\theta})}{\Lambda_w(\overline{\theta})} = \frac{C'(u_0(\theta_1))}{K'(w_0(\theta_1))} = \frac{\theta_1}{\beta}. \tag{10}$$

Hence

$$\begin{aligned}(\overline{\theta} - \theta_1)\, F(\overline{\theta}) &= \overline{\theta}\, F(\overline{\theta}) - \beta\left(\overline{\theta}\, F(\overline{\theta}) - \overline{F}(\overline{\theta})\right) \\ &= (1-\beta)\,\overline{\theta}\, F(\overline{\theta}) + \beta\, \overline{F}(\overline{\theta}) \\ &= \overline{G}(\overline{\theta}),\end{aligned} \tag{11}$$

where G and $\overline{G}$ are as above.

Remark 6. *Bearing in mind that $\theta_1 \leq \underline{\theta}$, so that $\overline{G}(\theta_1) = 0$, this equation can also be written*

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

or

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

That is, θ_1 satisfies the equation for the optimal cutoff in the problem without interpersonal transfers.

Remark 7. *This makes perfect sense: if $\theta_1 \leq \underline{\theta}$ then all θ types make the same choice, and the optimum with interpersonal transfers happens not to involve any transfers. It must therefore also be the optimum without interpersonal transfers, and must therefore satisfy the equation for the optimal cutoff in the problem without interpersonal transfers.*

However, we have not yet used equation (7). It follows from this equation that

$$d\nu_R = \tfrac{\beta}{\theta}\,(\,\theta\,F - \overline{F}\,)\,d\theta - F\,d\theta + \lambda\left(\Lambda_w - \tfrac{\beta}{\theta}\,\Lambda_u\right)d\theta.$$

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

$$\nu'_R = \tfrac{\beta}{\theta}\,(\,\theta\,F - \overline{F}\,) - F + \lambda\left(\Lambda_w - \tfrac{\beta}{\theta}\,\Lambda_u\right).$$

Furthermore:

$$\begin{aligned}\Lambda_u(\theta) &= \int_{[\underline{\theta},\theta]} C'(u_0(t))\,dF(t) = F(\theta)\,C'(u_0(\theta_1)) = \frac{F(\theta)}{F(\,\overline{\theta}\,)}\Lambda_u(\,\overline{\theta}\,)\\ &= \frac{F(\theta)}{F(\,\overline{\theta}\,)}\,\frac{\theta_1}{\beta}\,\Lambda_w(\,\overline{\theta}\,) = \frac{F(\theta)}{F(\,\overline{\theta}\,)}\,\frac{\theta_1}{\beta}\,\frac{F(\,\overline{\theta}\,)}{\lambda} = \frac{\theta_1}{\beta}\,\frac{F(\theta)}{\lambda}\end{aligned}$$

(where the last line follows from (10) and (5)); and

$$\begin{aligned}\Lambda_w(\theta) &= \int_{[\underline{\theta},\theta]} K'(w_0(t))\,dF(t) = F(\theta)\,K'(w_0(\theta_1)) = \frac{F(\theta)}{F(\,\overline{\theta}\,)}\Lambda_w(\,\overline{\theta}\,)\\ &= \frac{F(\theta)}{F(\,\overline{\theta}\,)}\,\frac{F(\,\overline{\theta}\,)}{\lambda} = \frac{F(\theta)}{\lambda}\end{aligned}$$

(where the last line follows from (5)). Hence

$$\lambda\,(\theta\,\Lambda_w - \beta\,\Lambda_u) = (\theta - \theta_1)\,F(\theta)$$

and

$$\begin{aligned}\theta\,\nu'_R &= \beta\,(\,\theta\,F - \overline{F}\,) - \theta\,F + (\theta - \theta_1)\,F\\ &= (\theta - \theta_1)\,F(\theta) - \overline{G}.\end{aligned}$$

Now, $F(\underline{\theta}) = \overline{G}(\underline{\theta}) = 0$. Hence $\underline{\theta}\,\nu'_R(\underline{\theta}) = 0$. Furthermore, we must have $\theta\,\nu'_R \geq 0$ on $\left(\underline{\theta},\overline{\theta}\right)$. Hence

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

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

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

Since $F'(\underline{\theta}+) > 0$, it follows that $\theta_1 \leq \beta\,\underline{\theta}$. Similarly, (11) implies that $(\overline{\theta} - \theta_1)\,F(\,\overline{\theta}\,) - \overline{G}(\,\overline{\theta}\,) = 0$. Hence $\overline{\theta}\,\nu'_R(\,\overline{\theta}\,) = 0$. Hence

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

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

$$(\theta\,\nu'_R)'(\overline{\theta}-) = (\beta\,\overline{\theta} - \theta_1)\,F'(\overline{\theta}-) \leq 0.$$

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

Remark 8. *This is where we use the assumption $\beta > 0$: we would not obtain a contradiction by combining two weak inequalities if we had $\beta = 0$.*

Remark 9. *More generally,*

$$\begin{aligned}(\theta\,\nu_R')' &= (\theta-\theta_1)\,F' + F - G\\ &= (\theta-\theta_1)\,F' - (1-\beta)\,\theta\,F'\\ &= (\beta\,\theta-\theta_1)\,F'.\end{aligned}$$

So, if we maintain the requirement that $\underline{\theta}\,\nu_R'(\underline{\theta}) = \overline{\theta}\,\nu_R'(\overline{\theta}) = 0$, but drop the requirement that $\theta\,\nu_R' \geq 0$ on $\left(\underline{\theta},\overline{\theta}\right)$, then we must conclude that $\theta_1 \in \left(\beta\,\underline{\theta},\beta\,\overline{\theta}\right)$ and $\theta\,\nu_R' < 0$ on $\left(\underline{\theta},\overline{\theta}\right)$. I.e. having $w' = 0$ is strictly suboptimal: $w' < 0$ would be better at all θ.

Theorem 3 Proof:
Expected utility of the population is:

$$Eu = \mu \int_{\underline{\theta}}^{\overline{\theta}} \left[\theta u(c_1) + v(c_2)\right] dF(\theta) + (1-\mu)\left[E(\theta)u(x+(1-\pi)y) + v(z)\right]$$

The balanced budget constraint is:

$$\mu \int_{\underline{\theta}}^{\overline{\theta}} \left[c_1 + c_2\right] dF(\theta) + (1-\mu)\left(w - \pi y\right) = 1$$

The Lagrangian is:

$$L(x,w,\Lambda|y) = \left\{ \begin{array}{c} \mu \int_{\underline{\theta}}^{\overline{\theta}} \left[\theta u(c_1) + v(c_2)\right] dF(\theta) + (1-\mu)\left[E(\theta)u(x+(1-\pi)y) + v(w-x-y)\right] \\ +\Lambda\left\{1 - \mu \int_{\underline{\theta}}^{\overline{\theta}} \left[c_1 + c_2\right] dF(\theta) - (1-\mu)\left(w - \pi y\right)\right\} \end{array} \right\}$$

Therefore by the Envelope Theorem (making use of continuity of the consumption functions):

$$\frac{\partial L}{\partial y}|_{y=0} = \left\{ \begin{array}{c} \mu \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial}{\partial y}\left[\theta u(c_1) + v(c_2)\right] dF(\theta) + (1-\mu)\left[\frac{E(\theta)}{1+\pi}u'(x) - v'(z)\right] \\ +\Lambda\left\{-\mu \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial}{\partial y}\left[c_1 + c_2\right] dF(\theta) + (1-\mu)\pi\right\} \end{array} \right\}$$

Recall that for $\theta \leq \theta_1$, consumption functions depend only on w, not on x or y. When $\theta_1 \leq \theta \leq \theta_2$, $c_1 = x$ and $c_2 = w - x$, so the functions do not depend on y. $\theta_2 \leq \theta \leq \theta_3$ is measureless when $y = 0$, so it can be ignored. Lastly, when $\theta > \theta_3$, $c_1 = x + (1-\pi)y$ and $c_2 = w - x - y$ (so that $c_1 + c_2 = w - \pi y$), yielding:

$$\frac{\partial L}{\partial y}|_{y=0} = \left\{ \begin{array}{c} \mu \int_{\theta_3}^{\overline{\theta}} \left[(1-\pi)\theta u'(x) - v'(z)\right] dF(\theta) + (1-\mu)\left[(1-\pi)E(\theta)u'(x) - v'(z)\right] \\ +\Lambda\left(\mu\pi \int_{\theta_3}^{\overline{\theta}} dF(\theta) + (1-\mu)\pi\right) \end{array} \right\}$$

Where here, $x + z = 1$.
We can recover Λ from the FOC for w (noting that $\frac{\partial(c_1+c_2)}{\partial w} = 1$):

$$0 = \frac{\partial L}{\partial w} = \frac{\partial Eu}{\partial w} - \Lambda$$

Giving the expected interpretation that Λ is the marginal utility of wealth.

$$\frac{\partial L}{\partial y}|_{y=0} = \left\{ \begin{array}{c} \mu \int_{\theta_3}^{\overline{\theta}} \left[(1-\pi)\theta u'(x) - v'(z)\right] dF(\theta) + (1-\mu)\left[(1-\pi)E(\theta)u'(x) - v'(z)\right] \\ +\frac{\partial Eu}{\partial w}\left(\mu\pi \int_{\theta_3}^{\overline{\theta}} dF(\theta) + (1-\mu)\pi\right) \end{array} \right\}$$

$$\frac{\partial L}{\partial y}|_{y=0} = \left\{ \begin{array}{c} \mu(1-\pi) \int_{\theta_3}^{\overline{\theta}} \left[\theta u'(x) - v'(z)\right] dF(\theta) + (1-\mu)(1-\pi)\left[E(\theta)u'(x) - v'(z)\right] \\ -\mu\pi \int_{\theta_3}^{\overline{\theta}} v'(z) dF(\theta) - (1-\mu)\pi v'(z) + \frac{\partial Eu}{\partial w}\left(\mu\pi \int_{\theta_3}^{\overline{\theta}} dF(\theta) + (1-\mu)\pi\right) \end{array} \right\}$$

$$\frac{\partial L}{\partial y}|_{y=0} = \left\{ \begin{array}{c} (1-\pi)\left(\mu \int_{\theta_3}^{\overline{\theta}} \left[\theta u'(x) - v'(z)\right] dF(\theta) + (1-\mu)\left[E(\theta)u'(x) - v'(z)\right]\right) \\ +\pi\left[\frac{\partial Eu}{\partial w} - v'(z)\right]\left(\mu \int_{\theta_3}^{\overline{\theta}} dF(\theta) + (1-\mu)\right) \end{array} \right\}$$

When $y = 0$, the Lagrangian collapses to the expected utility function:

$$Eu(x,w) = \mu \int_{\underline{\theta}}^{\overline{\theta}} \left[\theta u(c_1) + v(c_2)\right] dF(\theta) + (1-\mu)\left[E(\theta)u(x) + v(1-x)\right]$$

And the first order condition for optimality of x is given by:

$$0 = \frac{\partial Eu}{\partial x} = \mu \int_{\theta_1}^{\overline{\theta}} \left[\theta u'(x) - v'(z)\right] dF(\theta) + (1-\mu)\left[E(\theta)u'(x) - v'(z)\right]$$

Yielding:

$$\mu \int_{\theta_3}^{\overline{\theta}} \left[\theta u'(x) - v'(z)\right] dF(\theta) + (1-\mu)\left[E(\theta)u'(x) - v'(z)\right] = -\mu \int_{\theta_1}^{\theta_3} \left[\theta u'(x) - v'(z)\right] dF(\theta)$$

Which gives us:

$$\frac{\partial L}{\partial y}|_{y=0} = -(1-\pi)\mu \int_{\theta_1}^{\theta_3} \left[\theta u'(x) - v'(z)\right] dF(\theta) + \pi\left[\frac{\partial Eu}{\partial w} - v'(z)\right]\left(\mu \int_{\theta_3}^{\overline{\theta}} dF(\theta) + (1-\mu)\right)$$

When $y = 0$, we can write:

$$\frac{\partial Eu}{\partial w} = \mu\left(\int_{\underline{\theta}}^{\theta_1} \left[\theta u'(c_1)\frac{\partial c_1}{\partial w} + v'(c_2)\frac{\partial c_2}{\partial w}\right] dF(\theta) + \int_{\theta_1}^{\overline{\theta}} v'(z) dF(\theta)\right) + (1-\mu)v'(z)$$

Recalling that $\frac{\partial c_1}{\partial w} + \frac{\partial c_2}{\partial w} = 1$, and that $\theta u'(c_1) = v'(c_2)$ (from the first order condition for optimal consumption choice of the time-consistent $\beta = 1$ agent):

$$\frac{\partial Eu}{\partial w} = \mu\left(\int_{\underline{\theta}}^{\theta_1} v'(c_2) dF(\theta) + \int_{\theta_1}^{\overline{\theta}} v'(z) dF(\theta)\right) + (1-\mu)v'(z)$$

Recalling further that for $\theta < \theta_1$, $c_2 > z$, then $v'(c_2) < v'(z)$, yielding:

$$\frac{\partial Eu}{\partial w} < \mu\left(\int_{\underline{\theta}}^{\theta_1} v'(z) dF(\theta) + \int_{\theta_1}^{\overline{\theta}} v'(z) dF(\theta)\right) + (1-\mu)v'(z)$$

$$\frac{\partial Eu}{\partial w} < v'(z)$$

This implies:

$$\frac{\partial L}{\partial y}|_{y=0} < -(1-\pi)\mu \int_{\theta_1}^{\theta_3} \left[\theta u'(x) - v'(z)\right] dF(\theta) + \pi\left(v'(z) - v'(z)\right)\left(\mu \int_{\theta_3}^{\overline{\theta}} dF(\theta) + (1-\mu)\right)$$

$$\frac{\partial L}{\partial y}|_{y=0} < -(1-\pi)\mu u'(x)\int_{\theta_1}^{\theta_3}[\theta-\theta_1]\,dF(\theta)$$

$$\frac{\partial L}{\partial y}|_{y=0} < 0$$

Concluding the proof.