Online Appendices for "Which Early Withdrawal Penalty Attracts the Most Deposits to a Commitment Savings Account?"

John Beshears, Harvard University and NBER
James J. Choi, Yale University and NBER
Christopher Harris, University of Cambridge
David Laibson, Harvard University and NBER
Brigitte C. Madrian, Harvard University and NBER
Jung Sakong, University of Chicago

August 29th 2013
This draft: August 23rd 2017

A. Introduction

In this appendix we provide a complete analysis of the mechanism-design problem described in the main body of the paper.

B. Preliminaries

B.1. Functions of Bounded Variation. We begin by discussing the concept of bounded variation. This concept will be used to formulate our assumptions on the distribution function F in the subsection immediately following this one, namely Section B.2. More importantly, it plays an essential role in our proof of sufficiency in a much later section, namely Section P.

The simplest definition of a function of bounded variation is probably that given in the main text: a function $f : (0,\infty) \to \mathbb{R}$ is of bounded variation iff it is the difference of two bounded and non-decreasing functions $f^+, f^- : (0,\infty) \to \mathbb{R}$. This definition forms the starting point for the definition that we shall use. However, it needs to be developed into a form that is more convenient for the Lagrangean analysis below.

The first step is to collect the functions of bounded variation into equivalence classes. Intuitively speaking, two functions of bounded variation are equivalent iff they differ only at their points of discontinuity. This step is analogous to the first step in defining spaces of Lebesgue integrable functions. (In that case, one collects the Lebesgue integrable functions into equivalence classes. Two Lebesgue integrable functions are equivalent if they differ only on a set of measure 0.)

The second step is to place a norm on the resulting equivalence classes in such a way that the limit of a sequence of equivalence classes is again a suitable equivalence classes. (This step is analogous to the second step in defining spaces of Lebesgue integrable functions.) The main idea here is to note that, since f^+ and f^- are non-decreasing, they are effectively the distribution functions of a pair of non-negative bounded measures df^+ and df^-. Of course, neither df^+ nor df^- is unique. But their difference $df = df^+ - df^-$ is. The main component of the norm is therefore the total variation $\|df\|_{TV}$ of df. The other idea is to note that, while $\|df\|_{TV}$ effectively controls the derivative of f, it does not control the level of f. The remaining component of the norm can therefore be taken to be $|f_R(1)|$, where $f_R(1)$ is the limit of f from the right at 1.

The best way of understanding how these ideas work is to note that we can easily reconstruct f from df and $f_R(1)$. For all $\theta \in (0,1)$, we have

$$f_R(\theta) = f_R(1) - df((\theta, 1])$$

and

$$f_L(\theta) = f_R(1) - df([\theta, 1]) \,,$$

where $f_R(\theta)$ and $f_L(\theta)$ are the limits of f from the right and left at θ. And for all $\theta \in (1, \infty)$, we have

$$f_R(\theta) = f_R(1) + df((1, \theta])$$

and

$$f_L(\theta) = f_R(1) + df((1, \theta)) \,.$$

We also need to work with the space $\mathcal{BV}(\Theta, \mathbb{R})$ of functions of bounded variation on $\Theta = \left[\underline{\theta}, \overline{\theta}\right]$. By analogy with our discussion of the space $\mathcal{BV}((0, \infty), \mathbb{R})$, it should be clear that we can endow $\mathcal{BV}(\Theta, \mathbb{R})$ with the norm

$$\|f\|_{\mathcal{BV}} = |f_R(\theta_0)| + \|df\|_{TV} \,,$$

where θ_0 is a fixed element of $\left(\underline{\theta}, \overline{\theta}\right)$ and df is a bounded measure on Θ. There is, however, one surprise: a function $f \in \mathcal{BV}(\Theta, \mathbb{R})$ has a limit on the left at $\underline{\theta}$ and a limit on the right at $\overline{\theta}$. Indeed, we have

$$f_L(\underline{\theta}) = f_R(\theta_0) - df([\underline{\theta}, \theta_0])$$

and

$$f_R(\overline{\theta}) = f_R(\theta_0) + df\left((\theta_0, \overline{\theta}]\right).$$

To summarize, we denote the space of functions of bounded variation on $(0, \infty)$ by $\mathcal{BV}((0, \infty), \mathbb{R})$, and we denote the space of functions of bounded variation on $\Theta = \left[\underline{\theta}, \overline{\theta}\right]$ by $\mathcal{BV}(\Theta, \mathbb{R})$. Unless explicitly stated to the contrary, we shall always use the right-continuous representative of a function of bounded variation. We will usually denote this representative simply by f, but we will occasionally denote it by f_R for emphasis. We will denote the left-continuous representative of f by f_L.

B.2. Assumptions on F. We are now in a position to introduce our assumptions on the distribution function F of the taste shock θ. They are:

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

Here G is given by the formula $G(\theta) = (1 - \beta)\, \theta\, F'(\theta) + F(\theta)$. If A1 holds then G, like F and F', is a function of bounded variation on $(0, \infty)$.

B.3. The Support of F' is Connected. Fourth, we note that either $\beta = 1$, in which case the analysis is trivial, or $\beta < 1$, in which case the support of F' is connected.[1] More precisely, we have:

[1]Notice that F is a distribution function, not a distribution. Also, F' has a dual interpretation. It can be regarded as: either (i) the non-negative finite measure with distribution function F; or (ii) the density of that measure with respect to Lebesgue measure. By the same token, the support of F' has a dual interpretation. It can be regarded as: either (i) the support of the non-negative finite measure F'; or (ii) the support of the non-negative function of bounded variation F'. It makes no difference which of these two interpretations is adopted.

Proposition 1. *Suppose that $\beta < 1$ and that A1-A3 are satisfied. Then there exist $\underline{\kappa}, \overline{\kappa} \in [\underline{\theta}, \overline{\theta}]$ such that: (i) $\underline{\kappa} < \overline{\kappa}$; (ii) $F' > 0$ on $(\underline{\kappa}, \overline{\kappa})$; and (iii) $F' = 0$ on $(0, \infty) \setminus [\underline{\kappa}, \overline{\kappa}]$.*

In what follows we shall therefore take it that $\beta < 1$, and that the support of F' is $[\underline{\theta}, \overline{\theta}]$.

Proof. Note first that there exists $\kappa_1 \in (\underline{\theta}, \overline{\theta})$ such that $F'(\kappa_1) > 0$. Otherwise we would have $F' = 0$ everywhere on $(0, \infty)$, by right-continuity of F'. Next, there exists $\kappa_2 \in (\kappa_1, \overline{\theta})$ such that $F' > 0$ on (κ_1, κ_2), again by right-continuity of F'. Third, put $\underline{\kappa} = \inf \{\theta \mid F'(\theta) > 0\}$ and $\overline{\kappa} = \sup \{\theta \mid F'(\theta) > 0\}$. Then certainly $\underline{\theta} \leq \underline{\kappa} < \overline{\kappa} \leq \overline{\theta}$. Fourth, put $\alpha = \frac{2-\beta}{1-\beta}$. Then $G' \geq 0$ iff $(\theta^\alpha F')' \geq 0$ and $G' \leq 0$ iff $(\theta^\alpha F')' \leq 0$. There are therefore two possibilities. If $G' \geq 0$ at θ_M (i.e. $\Delta G(\theta_M) \geq 0$), then we must have $\theta^\alpha F' > 0$ on $(\underline{\kappa}, \theta_M]$ (because $(\theta^\alpha F')' \geq 0$ on this interval) and $\theta^\alpha F' > 0$ on $(\theta_M, \overline{\kappa})$ (because $(\theta^\alpha F')' \leq 0$ on this interval); and if $G' \leq 0$ at θ_M (i.e. $\Delta G(\theta_M) \leq 0$), then we must have $\theta^\alpha F' > 0$ on $(\underline{\kappa}, \theta_M)$ and $\theta^\alpha F' > 0$ on $[\theta_M, \overline{\kappa})$.[2] Either way, we see that: (i) $\theta^\alpha F' > 0$, and hence $F' > 0$, on $(\underline{\kappa}, \overline{\kappa})$; (ii) $\theta_M \leq \overline{\kappa}$, for otherwise we would have $F' > 0$ on the non-empty interval $(\overline{\kappa}, \theta_M)$, and this contradicts the choice of $\overline{\kappa}$; and (iii) $\theta_M \geq \underline{\kappa}$, for otherwise we would have $F' > 0$ on the non-empty interval $(\theta_M, \underline{\kappa})$, and this contradicts the choice of $\underline{\kappa}$. ■

B.4. Constraints on the Budget Set. Fifth, recall that self 0 chooses a subset B of the ambient action set A, and that self 1's choice of a consumption pair from B can therefore be described by a consumption curve $(c_1, c_2) : [\underline{\theta}, \overline{\theta}] \to B$. We consider three possible constraints on B, namely:

Constraint 1. B is a non-empty compact subset of A.

Constraint 2. The penalty for transferring consumption from period 2 to period 1 is no greater than π.[3]

[2] If $\theta_M \leq \underline{\kappa}$ then we take the intervals $(\underline{\kappa}, \theta_M)$ and $(\underline{\kappa}, \theta_M]$ to be empty. Similarly, if $\theta_M \geq \overline{\kappa}$, then we take the intervals $(\theta_M, \overline{\kappa})$ and $[\theta_M, \overline{\kappa})$ to be empty.

[3] In other words, for any given $(c_1, c_2) \in B$ and any $\Delta c_1 \in \left[0, \frac{1}{1+\pi} c_2\right]$, self 1 can increase her own consumption c_1 by Δc_1 at a cost of at most $(1 + \pi) \Delta c_1$ in terms of the consumption c_2 of self 2.

Constraint 3. The penalty for transferring consumption from period 1 to period 2 is no greater than π.[4]

Constraint 1 involves no loss of generality. Indeed, it must be possible for all possible types $\theta \in \Theta$ to find an optimum within B. This being the case, we can always take the closure of B without changing the outcome, since the utility function is continuous. Finally, since A itself is compact, so too is the closure of B. Constraint 2 is an essential part of the formulation of our problem. We wish to avoid extreme outcomes in which self 0 imposes an infinite penalty on self 1 for increasing her own consumption at the expense of self 2. Constraint 3 is simply the mirror image of Constraint 2. It eliminates extreme outcomes in which self 0 imposes an infinite penalty on self 1 for increasing the consumption of self 2 at her own expense.

Remark 2. *If we only impose Constraint 2, then the problem is one sided: Constraint 2 places a limit on the cost, in terms of c_2, of increasing c_1; but there is no corresponding limit on the cost, in terms of c_1, of increasing c_2. By imposing Constraint 3, we eliminate this asymmetry.*

Now suppose that B must satisfy all three constraints. Then B must take one of two forms: either

1. it consists of the single point $(0, 0)$; or
2. its frontier consists of a curve that begins at some $(0, c_2)$ such that $c_2 > 0$, slopes downwards with slope between $-(1 + \pi)$ and $-(1 + \pi)^{-1}$, and ends at some $(c_1, 0)$ such that $c_1 > 0$.

Self 0 will never choose the first option, since the optimal pooling point on the frontier of the ambient budget set A is preferable. (By the same token, self 0 will never choose a B, the frontier of which is close to $(0, 0)$.) But, if she chooses the second option, then the resulting consumption curve (c_1, c_2) will be interior. That is, we will have $c_1, c_2 > 0$ on Θ.[5]

[4] In other words, for any given $(c_1, c_2) \in B$ and any $\Delta c_2 \in \left[0, \frac{1}{1+\pi} c_1\right]$, self 1 can increase the consumption c_2 of self 2 by Δc_2 at a cost of at most $(1 + \pi) \Delta c_2$ in terms of her own consumption c_1.

[5] This follows from our assumption that $U_t'(0+) = \infty$.

The ideal approach to our problem would therefore be to impose all three constraints on B, and to solve the optimization problem of self 0 subject to these constraints. One could then verify ex post that Constraint 3 was not binding.[6]

In practice, we shall take a shortcut. Rather than working explicitly with Constraint 3, we shall instead replace it by the weaker requirement that consumption curves are interior. Our analysis could, of course, be reworked in such a way as to incorporate Constraint 3 explicitly. But, in practice, this would simply involve an additional notational burden.

Remark 3. *The situation would be very different if $\beta > 1$. In that case, it would be Constraint 2 that would not bind. We would therefore replace Constraint 2 by the weaker requirement that consumption curves are interior.*

B.5. Utility Curves. Suppose accordingly that we are given a B satisfying Constraints 1 and 2, and that the associated consumption curve is interior. Define a utility curve

$$(u_1, u_2) : \left[\underline{\theta}, \overline{\theta}\right] \to (U_1(0), U_1(\infty)) \times (U_2(0), U_2(\infty))$$

by the formula $(u_1, u_2)(\theta) = (U_1(c_1(\theta)), U_2(c_2(\theta)))$. Then (u_1, u_2) must satisfy the following conditions:

N1 $C_1(u_1(\theta)) + C_2(u_2(\theta)) \leq y$ for all $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$.

N2 u_1 is non-decreasing and u_2 is non-increasing.

N3 $\theta\, du_1 + \beta\, du_2 = 0$.

N4 $\beta\,(1+\pi)\, U_2'(C_2(u_2(\theta))) \geq \theta\, U_1'(C_1(u_1(\theta)))$.

Here: $C_t = U_t^{-1}$; and du_1 is a non-negative finite measure and du_2 is a non-positive finite measure.

Conversely, suppose that a utility curve (u_1, u_2) is interior, in the sense that it satisfies $u_1 > U_1(0)$ and $u_2 > U_2(0)$ on Θ, and that it satisfies Conditions N1-N4.

[6] It turns out that the slope of the optimal budget set is at most -1. So Constraint 3 certainly is not binding!

Define (c_1, c_2) by the formula $(c_1, c_2)(\theta) = \big(U_1^{-1}(u_1(\theta)), U_2^{-1}(u_2(\theta))\big)$. Then there exists a B with slope at least $-(1+\pi)$ such that (c_1, c_2) is the consumption curve arising from B. Moreover (c_1, c_2) is interior.

B.6. The CRRA Case. Suppose now that $U_1 = U_2 = U$ on $(0, \infty)$, and that U has constant relative risk aversion $\rho > 0$. Indeed, suppose for concreteness that U is given by the formula

$$U(c) = \left\{ \begin{array}{ll} \frac{c^{1-\rho}-1}{1-\rho} & \text{if} \quad \rho \neq 1 \\ \log(c) & \text{if} \quad \rho = 1 \end{array} \right\}.$$

Then N4 is equivalent to

N4′ $u_2(\theta) \leq -\frac{1}{\rho}\, a\Big(\frac{\theta}{(1+\pi)\,\beta}\Big) + b\Big(\frac{\theta}{(1+\pi)\,\beta}\Big)\, u_1(\theta)$,

where a and b are given by the formulae

$$a(z) = \left\{ \begin{array}{ll} \dfrac{z^{1-\frac{1}{\rho}}-1}{1-\frac{1}{\rho}} & \text{if} \quad \rho \neq 1 \\ \log(z) & \text{if} \quad \rho = 1 \end{array} \right\} \tag{1}$$

and

$$b(z) = z^{1-\frac{1}{\rho}}. \tag{2}$$

Remark 4. *It is obvious that N4 becomes weaker as π increases. Since N4′ is equivalent to N4, N4′ likewise becomes weaker as π increases.*

C. The Main Problem

Our strategy will be to study a relaxed version of the problem of self 0 in which we maximize self 0's expected utility $\int (\theta\, u(\theta) + w(\theta))\, dF(\theta)$ subject to N1, N3 and N4′, but not N2. Following Luenberger (1969, Sections 8.3 and 8.4, pp. 216-221), we shall need:

1. A vector space[7] X, which we take to be $\mathcal{C}(\Theta, \mathbb{R})^2$. Here: $\Theta = \big[\underline{\theta}\,, \overline{\theta}\,\big] \subset (0, \infty)$ is the space of types; and $\mathcal{C}(\Theta, \mathbb{R})$ is the space of continuous functions from Θ to $\mathbb{R}$.

[7] In our case X is actually a Banach space. For the Lagrangean analysis, we only need the fact that it is a vector space. When we later use calculus to find necessary and sufficient conditions for the maximization of the Lagrangean, we shall need the fact that it is a normed space. Cf. Luenberger (1969, Lemma 1, p. 227).

2. A convex set[8] $\Omega \subset X$, which we take to consist of all

$$(u, w) \in \Big(\mathcal{BV}(\Theta, \mathrm{ran}(U)) \cap \mathcal{C}(\Theta, \mathrm{ran}(U))\Big)^2$$

such that

$$\theta\, du + \beta\, dw = 0.$$

Here: $\mathrm{ran}(U)$ is the range of U;[9] $\mathcal{C}(\Theta, \mathrm{ran}(U))$ is the space of continuous functions from Θ to $\mathrm{ran}(U)$; $\mathcal{BV}(\Theta, \mathrm{ran}(U))$ is the space of all functions of bounded variation from Θ to $\mathrm{ran}(U)$; and du and dw are in general elements of the space $\mathcal{M}(\Theta, \mathbb{R})$ of finite Borel measures on Θ.

3. A concave function[10] $M : \Omega \to \mathbb{R}$ (the objective function), which we take to be given by the formula

$$M(u, w) = \int \Big(\theta\, u(\theta) + w(\theta)\Big)\, dF(\theta).$$

4. A normed space[11] Z, which we take to be $\mathcal{C}(\Theta, \mathbb{R})$.

5. A closed convex cone P in Z with vertex 0 and non-empty interior, which we take to be $\mathcal{C}(\Theta, [0, \infty))$. With this choice of P, $z_1 \geq z_2$ iff $z_1(\theta) \geq z_2(\theta)$ for all $\theta \in \Theta$ and $z_1 > z_2$ iff $z_1(\theta) > z_2(\theta)$ for all $\theta \in \Theta$. In other words, P is the positive cone of Z.

6. The space Z^* of continuous linear functionals on Z. Since $Z = \mathcal{C}(\Theta, \mathbb{R})$, Z^* can be identified with $\mathcal{M}(\Theta, \mathbb{R})$.

7. The positive cone P^* of Z^*. Since $P = \mathcal{C}(\Theta, [0, \infty))$, P^* can be identified with $\mathcal{M}(\Theta, [0, \infty))$ (i.e. the space of non-negative finite Borel measures on Θ).

[8] In our case Ω is actually a cone, the vertex of which is the constant mapping $\frac{1}{\rho-1}$ when $\rho \neq 1$ and the constant mapping 0 when $\rho = 1$.

[9] I.e. $\mathrm{ran}(U)$ is $(\frac{1}{\rho-1}, \infty)$ when $\rho < 1$, $(-\infty, \infty)$ when $\rho = 1$ and $(-\infty, \frac{1}{\rho-1})$ when and $\rho > 1$.

[10] In our case M is actually defined on the whole of X (and not just on Ω), and it is linear (and not just concave).

[11] In our case, Z is actually a Banach space, and not just a normed space.

8. Concave mappings $G_1, G_2 : \Omega \to Z$ (the constraint mappings),[12] which we take to be given by the formulae

$$(G_1(u, w))(\theta) = y - C(u(\theta)) - K(w(\theta))$$

and

$$(G_2(u, w))(\theta) = b\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big)\,u(\theta) - \frac{1}{\rho}\,a\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big) - w(\theta),$$

where $C = K = U^{-1} : \mathrm{ran}(U) \to (0, \infty)$, and a and b are given by the formulae (1) and (2).

Our problem is then to

$$\begin{array}{ll} \text{maximize} & M(x) \\ \text{subject to} & \left\{ \begin{array}{c} x \in \Omega \\ G_1(x) \geq 0 \\ G_2(x) \geq 0 \end{array} \right\}. \end{array} \tag{3}$$

D. Characterizing the Optimum

In our context, the Lagrangean is the mapping $L : \Omega \times Z^* \times Z^* \to \mathbb{R}$ given by the formula

$$L(x, \lambda_1, \lambda_2) = M(x) + \langle G_1(x), \lambda_1 \rangle + \langle G_2(x), \lambda_2 \rangle\,,$$

where $\langle G_i(x), \lambda_i \rangle$ denotes the real number obtained when the linear functional $\lambda_i \in Z^*$ is evaluated at the point $G_i(x) \in Z$.

In view of our assumptions, the maximum is achieved at $x_0 \in \Omega$ if and only if there exist $\lambda_1, \lambda_2 \in Z^*$ such that:

1. $L(x_0, \lambda_1, \lambda_2) \geq L(x, \lambda_1, \lambda_2)$ for all $x \in \Omega$;

2. $G_1(x) \geq 0$, $\lambda_1 \geq 0$ and
$$\langle G_1(x), \lambda_1 \rangle = 0; \tag{4}$$

[12] In our case G_1 is actually defined on $\Xi = \mathcal{C}(\Theta, \mathrm{ran}(U))^2$ (and not just on Ω). This will be useful when we later want to do calculus. Furthermore G_2 is defined on the whole of X (and not just on Ω), and it is linear (and not just concave).

3. $G_2(x) \geq 0$, $\lambda_2 \geq 0$ and

$$\langle G_2(x), \lambda_2 \rangle = 0. \tag{5}$$

In other words, there exists multipliers λ_1 and λ_2 such that: x_0 maximizes $L(\,\cdot\,, \lambda_1, \lambda_2)$ over Ω; complementary slackness holds for the first constraint; and complementary slackness holds for the second constraint.

Since P^* can be identified with $\mathcal{M}(\Theta, [0, \infty))$, we have the following explicit representations of $M(x)$, $\langle G_1(x), \lambda_1 \rangle$ and $\langle G_2(x), \lambda_2 \rangle$:

$$M(x) = \int \Big(\theta\, u(\theta) + w(\theta)\Big)\, dF(\theta), \tag{6}$$

$$\langle G_1(x), \lambda_1 \rangle = \int \Big(y - C(u(\theta)) - K(w(\theta))\Big)\, d\Lambda_1(\theta) \tag{7}$$

and

$$\langle G_2(x), \lambda_2 \rangle = \int \Big(b\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big)\, u(\theta) - \frac{1}{\rho}\, a\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big) - w(\theta)\Big)\, d\Lambda_2(\theta), \tag{8}$$

where Λ_1 and Λ_2 are the distribution functions of λ_1 and λ_2 respectively.

Remark 5. *In the interests of clarity and consistency, all integrals in this Appendix are Lebesgue-Stieltjes integrals, i.e. integrals with respect to functions of bounded variation.*

E. The Lagrangean is Fréchet Differentiable

It is immediate from the formulae (6), (7) and (8) that $L(x, \lambda_1, \lambda_2)$ is in fact well defined for all $x \in \Xi = \mathcal{C}(\Theta, \mathrm{ran}(U))^2$. Let us consider accordingly any $x_0 = (u_0, w_0) \in \Xi$ and any $x_1 = (u_1, w_1) \in X$. Because Ξ is open, $x_0 + \varepsilon\, x_1 \in \Xi$ for all $\varepsilon > 0$ sufficiently small. Furthermore, it can be verified that the directional derivative $\nabla_{x_1} L(x_0, \lambda_1, \lambda_2)$ of L at x_0 in the direction x_1 takes the form

$$\int \Big(\theta\, u_1 + w_1\Big)\, dF - \int \Big(C'(u_0)\, u_1 + K'(w_0)\, w_1\Big)\, d\Lambda_1 + \int \Big(b\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big)\, u_1 - w_1\Big)\, d\Lambda_2. \tag{9}$$

This is easily seen to define a continuous linear functional

$$\nabla L(x_0, \lambda_1, \lambda_2) : x_1 \mapsto \nabla_{x_1} L(x_0, \lambda_1, \lambda_2)$$

on X. That is, $L(\,\cdot\,,\lambda_1,\lambda_2)$ is Gâteaux differentiable at x_0 with gradient $\nabla L(x_0,\lambda_1,\lambda_2) \in X^*$. Finally, $\nabla L(\,\cdot\,,\lambda_1,\lambda_2) : \Xi \to X^*$ can be shown to be continuous. It follows that $L(\,\cdot\,,\lambda_1,\lambda_2)$ is Fréchet differentiable on Ξ.

F. Maximizing the Lagrangean

Since $L(\,\cdot\,,\lambda_1,\lambda_2)$ is convex and Fréchet differentiable on Ξ, and since Ω is convex, the maximum of $L(\,\cdot\,,\lambda_1,\lambda_2)$ over Ω is achieved at $x_0 \in \Omega$ iff

$$\nabla_{x-x_0} L(x_0,\lambda_1,\lambda_2) \le 0$$

for all $x \in \Omega$. In this section we shall identify the restrictions that this places on λ_1 and λ_2.

To this end, put

$$Y = \Big(\mathcal{BV}(\Theta,\mathbb{R}) \cap \mathcal{C}(\Theta,\mathbb{R})\Big) \times \mathbb{R};$$

and consider the affine transformation

$$S : Y \to \Big(\mathcal{BV}(\Theta,\mathbb{R}) \cap \mathcal{C}(\Theta,\mathbb{R})\Big)^2$$

that maps $y = (u,r)$ to $x = (u_0+u, w_0+w)$, where w is the unique solution of the equation $\theta\,du + \beta\,dw = 0$ with boundary condition $w(\,\overline{\theta}\,) = r$.

For any $y \in Y$, we have

$$\begin{aligned}
\nabla_{S(y)-x_0} L(x_0,\lambda_1,\lambda_2) &= \int \Big(\theta\,u + w\Big)\,dF - \int \Big(C'(u_0)\,u + K'(w_0)\,w\Big)\,d\Lambda_1 \\
&\qquad + \int \Big(b\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big)\,u - w\Big)\,d\Lambda_2 \\
&= \int \Big(\theta\,u + w\Big)\,dF - \int \Big(\tfrac{C'(u_0)}{K'(w_0)}\,u + w\Big)\,d\widetilde{\Lambda}_1 \\
&\qquad + \int \Big(b\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big)\,u - w\Big)\,d\Lambda_2
\end{aligned}$$

(where $d\widetilde{\Lambda}_1 = K'(w_0)\,d\Lambda_1$). Furthermore, integrating by parts, we have

$$\int w\,dF = [\,w\,F\,]_{\underline{\theta}-}^{\overline{\theta}} - \int F\,dw = w(\,\overline{\theta}\,)\,F(\,\overline{\theta}\,) + \int F\,\frac{\theta}{\beta}\,du$$

(because $F(\underline{\theta}-) = 0$ and $dw = -\frac{\theta}{\beta}\, du$)

$$= r\, F(\overline{\theta}\,) + \int F\, \frac{\theta}{\beta}\, du.$$

Moreover

$$\begin{aligned} \int F\, \theta\, du &= \int F\, (\theta\, du + u\, d\theta) - \int F\, u\, d\theta \\ &= [\,\theta\, u\, F\,]_{\underline{\theta}-}^{\overline{\theta}} - \int \theta\, u\, dF - \int F\, u\, d\theta \\ &= \overline{\theta}\, u(\overline{\theta}\,)\, F(\overline{\theta}\,) - \int \theta\, u\, dF - \int F\, u\, d\theta \end{aligned}$$

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

$$\int w\, dF = \left(\frac{\overline{\theta}}{\beta} u(\overline{\theta}\,) + r\right) F(\overline{\theta}\,) - \frac{1}{\beta} \int u\, (\theta\, dF + F\, d\theta).$$

Similarly,

$$\int w\, d\widetilde{\Lambda}_1 = \left(\frac{\overline{\theta}}{\beta} u(\overline{\theta}\,) + r\right) \widetilde{\Lambda}_1(\overline{\theta}\,) - \frac{1}{\beta} \int u\, (\theta\, d\widetilde{\Lambda}_1 + \widetilde{\Lambda}_1\, d\theta)$$

and

$$\int w\, d\Lambda_2 = \left(\frac{\overline{\theta}}{\beta} u(\overline{\theta}\,) + r\right) \Lambda_2(\overline{\theta}\,) - \frac{1}{\beta} \int u\, (\theta\, d\Lambda_2 + \Lambda_2\, d\theta).$$

Overall, then,

$$\begin{aligned} \nabla_{S(y)-x_0} L(x_0, \lambda_1, \lambda_2) = &\left(\frac{\overline{\theta}}{\beta} u(\overline{\theta}\,) + r\right) \Big(F(\overline{\theta}\,) - \widetilde{\Lambda}_1(\overline{\theta}\,) - \Lambda_2(\overline{\theta}\,)\Big) \\ &- \frac{1}{\beta} \int u\, \Big((1-\beta)\, \theta\, dF + F\, d\theta\Big) \\ &+ \frac{1}{\beta} \int u\, \Big(\Big(\theta - \beta\, \tfrac{C'(u_0)}{K'(w_0)}\Big)\, d\widetilde{\Lambda}_1 + \widetilde{\Lambda}_1\, d\theta\Big) \\ &+ \frac{1}{\beta} \int u\, \Big(\Big(\theta + \beta\, b\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big)\Big)\, d\Lambda_2 + \Lambda_2\, d\theta\Big). \end{aligned}$$

Next, it is easy to see that the mapping

$$y \mapsto \nabla_{S(y)-x_0} L(x_0, \lambda_1, \lambda_2)$$

defines a continuous linear functional on Y. Since it does not depend on the derivatives of y, this functional extends uniquely to a continuous linear functional

$$y^* : \mathcal{C}(\Theta, \mathbb{R}) \times \mathbb{R} \to \mathbb{R}.$$

Indeed, we have

$$y^* = (u^*, r^*) \in \mathcal{M}(\Theta, \mathbb{R}) \times \mathbb{R} = \left(\mathcal{C}(\Theta, \mathbb{R}) \times \mathbb{R}\right)^*,$$

where

$$\begin{aligned} du^* = &-\frac{1}{\beta}\Big((1-\beta)\,\theta\,dF + F\,d\theta\Big) + \frac{1}{\beta}\Big(\Big(\theta - \beta\,\tfrac{C'(u_0)}{K'(w_0)}\Big)\,d\widetilde{\Lambda}_1 + \widetilde{\Lambda}_1\,d\theta\Big) \\ &+ \frac{1}{\beta}\Big(\Big(\theta + \beta\,b\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big)\Big)\,d\Lambda_2 + \Lambda_2\,d\theta\Big) + \frac{\overline{\theta}}{\beta}\Big(F(\overline{\theta}\,) - \widetilde{\Lambda}_1(\overline{\theta}\,) - \Lambda_2(\overline{\theta}\,)\Big)\,dI \end{aligned} \tag{10}$$

and

$$r^* = F(\overline{\theta}\,) - \widetilde{\Lambda}_1(\overline{\theta}\,) - \Lambda_2(\overline{\theta}\,)$$

and I is the distribution function of the unit mass at $\overline{\theta}$.

Remark 6. *In reading (10), note that F, $\widetilde{\Lambda}_1$, Λ_2 and θ are functions of bounded variation (with θ being simply the identity function). Hence dF, $d\widetilde{\Lambda}_1$, $d\Lambda_2$ and $d\theta$ are measures, and the equation as a whole holds in terms of measures..*

Finally, it is easy to see that there exists $\varepsilon > 0$ such that $S(y) \in \Omega$ for all $y \in Y \cap B_\varepsilon(0)$, where $B_\varepsilon(0)$ is the open ball in $\mathcal{C}(\Theta, \mathbb{R}) \times \mathbb{R}$ with radius ε and centre 0. It follows that $\langle y, y^* \rangle \leq 0$ for all $y \in Y \cap B_\varepsilon(0)$. But $Y \cap B_\varepsilon(0)$ is dense in $B_\varepsilon(0)$.

Hence $\langle y, y^* \rangle \leq 0$ for all $y \in B_\varepsilon(0)$. Hence $y^* = 0$. In other words, we have

$$\begin{aligned} 0 = &-\frac{1}{\beta}\Big((1-\beta)\,\theta\,dF + F\,d\theta\Big) + \frac{1}{\beta}\Big(\Big(\theta - \beta\,\tfrac{C'(u_0)}{K'(w_0)}\Big)\,d\widetilde{\Lambda}_1 + \widetilde{\Lambda}_1\,d\theta\Big) \\ &+ \frac{1}{\beta}\Big(\Big(\theta + \beta\,b\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big)\Big)\,d\Lambda_2 + \Lambda_2\,d\theta\Big) + \frac{\overline{\theta}}{\beta}\Big(F(\overline{\theta}\,) - \widetilde{\Lambda}_1(\overline{\theta}\,) - \Lambda_2(\overline{\theta}\,)\Big)\,dI \end{aligned} \tag{11}$$

and

$$0 = F(\overline{\theta}\,) - \widetilde{\Lambda}_1(\overline{\theta}\,) - \Lambda_2(\overline{\theta}\,). \tag{12}$$

Taking advantage of (12), (11) simplifies to

$$\begin{aligned} 0 \;\; = \;\; &-G\,d\theta + \Big(\Big(\theta - \beta\,\tfrac{C'(u_0)}{K'(w_0)}\Big)\,d\widetilde{\Lambda}_1 + \widetilde{\Lambda}_1\,d\theta\Big) \\ &+ \Big(\Big(\theta + \beta\,b\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big)\Big)\,d\Lambda_2 + \Lambda_2\,d\theta\Big)\,, \end{aligned} \tag{13}$$

where G is given by the formula $G(\theta) = (1-\beta)\,\theta\,F'(\theta) + F(\theta)$.

G. A One-Parameter Family of Utility Curves

We shall consider a family of utility curves depending on the single parameter $\theta_1 \in \big(0, \overline{\theta}\,\big]$. For each θ_1, we begin by finding the point $(c^*(\theta_1), k^*(\theta_1))$ that would be chosen by a self 1 of type θ_1 from the ambient budget set A. The utility curve corresponding to θ_1 is then the utility curve associated with a two-part budget set with slopes of -1 and $-(1+\pi)$ to the left and right of a kink at $(c^*(\theta_1), k^*(\theta_1))$. Notice that we specifically allow for the possibility that $\theta_1 < \underline{\theta}$.

Put $\theta_2 = (1+\pi)\,\theta_1$. Knife-edge cases apart, there are then five possible cases arising from the relative positions of the two non-trivial intervals $\big[\,\underline{\theta}\,, \overline{\theta}\,\big]$ and $[\theta_1, \theta_2]$:

Case 1 $\big[\,\underline{\theta}\,, \overline{\theta}\,\big]$ contains $[\theta_1, \theta_2]$;

Case 2 $[\theta_1, \theta_2]$ contains $\big[\,\underline{\theta}\,, \overline{\theta}\,\big]$;

Case 3 the two intervals overlap, with $[\theta_1, \theta_2]$ lying to the left and $\big[\,\underline{\theta}\,, \overline{\theta}\,\big]$ to the right;

Case 4 the two intervals overlap, with $\big[\,\underline{\theta}\,, \overline{\theta}\,\big]$ lying to the left and $[\theta_1, \theta_2]$ to the right;

Case 5 $[\theta_1, \theta_2]$ lies entirely to the left of $\left[\underline{\theta}\,, \overline{\theta}\,\right]$.

(The case in which $\left[\underline{\theta}\,, \overline{\theta}\,\right]$ lies entirely to the left of $[\theta_1, \theta_2]$ cannot occur, because we are confining θ_1 to the interval $\left(0, \overline{\theta}\,\right]$.)

G.1. Case 1. If the utility curve corresponding to θ_1 is to be an optimum, then the associated multipliers λ_1 and λ_2 must satisfy the three necessary conditions (4), (5) and (13) (i.e. complementary slackness for the first constraint, complementary slackness for the second constraint and the measure equation). In this section, we show that these three necessary conditions tie down λ_1 and λ_2 uniquely. The fourth necessary condition, namely the boundary condition (12), is not needed at this stage. (It will be used below to tie down θ_1.)

By construction, the maximum-penalty constraint is strictly slack on $[\underline{\theta}, \theta_2)$ and the budget constraint is strictly slack on $\left(\theta_2, \overline{\theta}\,\right]$. Hence $d\Lambda_2 = 0$ on the former interval and $d\widetilde{\Lambda}_1 = 0$ on the latter. Furthermore (13) implies that

$$(1-\beta)\,\theta_2\,\Delta F(\theta_2) = (\theta_2 - \theta_1)\,\Delta\widetilde{\Lambda}_1(\theta_2) + \left(\theta_2 + \beta\, b\Big(\tfrac{\theta_2}{(1+\pi)\,\beta}\Big)\right)\Delta\Lambda_2(\theta_2),$$

where $\Delta F(\theta_2)$, $\Delta\widetilde{\Lambda}_1(\theta_2)$ and $\Delta\Lambda_2(\theta_2)$ are the atoms of dF, $d\widetilde{\Lambda}_1$ and $d\Lambda_2$ at θ_2. But Assumption A1 implies that $\Delta F(\theta_2) = 0$. Since all the terms on the right-hand side of the equation are non-negative, it follows that $\Delta\widetilde{\Lambda}_1(\theta_2) = \Delta\Lambda_2(\theta_2) = 0$. Hence $\Delta\Lambda_2(\theta_2) = 0$ (and therefore $d\Lambda_2 = 0$ on $[\underline{\theta}, \theta_2]$) and $\Delta\widetilde{\Lambda}_1(\theta_2) = 0$ (and therefore $d\widetilde{\Lambda}_1 = 0$ on $\left[\theta_2, \overline{\theta}\,\right]$).

Now let us consider the three intervals $[\underline{\theta}, \theta_1]$, $[\theta_1, \theta_2]$ and $\left[\theta_2, \overline{\theta}\,\right]$ in turn. On $[\underline{\theta}, \theta_1]$, we have $\frac{C'(u_0)}{K'(w_0)} = \frac{\theta}{\beta}$, $\Lambda_2 = 0$ and $d\Lambda_2 = 0$. Hence (13) becomes

$$0 = -G\,d\theta + \widetilde{\Lambda}_1\,d\theta.$$

It follows that $\widetilde{\Lambda}_1 = G$ almost everywhere with respect to Lebesgue measure $d\theta$. Since both $\widetilde{\Lambda}_1$ and G are functions of bounded variation, it then follows (bearing in mind the convention that functions of bounded variation are right continuous) that $\widetilde{\Lambda}_1 = G$ everywhere on $[\underline{\theta}, \theta_1)$, and hence that $\widetilde{\Lambda}_1(\theta_1-) = G(\theta_1-)$.

On $[\theta_1, \theta_2]$, we have $\frac{C'(u_0)}{K'(w_0)} = \frac{\theta_1}{\beta}$, $\Lambda_2 = 0$ and $d\Lambda_2 = 0$. Hence (13) becomes

$$0 = -G\, d\theta + (\theta - \theta_1)\, d\widetilde{\Lambda}_1 + \widetilde{\Lambda}_1\, d\theta.$$

This implies that $\widetilde{\Lambda}_1$ takes the form

$$\widetilde{\Lambda}_1(\theta) = \frac{1}{\theta - \theta_1} \int_{\theta_1}^{\theta} G(t)\, dt$$

for all $\theta \in (\theta_1, \theta_2)$, that $\widetilde{\Lambda}_1(\theta_1) = G(\theta_1)$ (since $\widetilde{\Lambda}_1$ and G are right continuous) and that $\widetilde{\Lambda}_1(\theta_2) = \widetilde{\Lambda}_1(\theta_2-) = \frac{1}{\theta_2 - \theta_1} \int_{\theta_1}^{\theta_2} G(\theta)\, d\theta$ (since $\widetilde{\Lambda}_1$ cannot have a jump at θ_2).

On $\left[\theta_2, \overline{\theta}\right]$, we have $d\widetilde{\Lambda}_1 = 0$. Hence (13) becomes

$$0 = -G\, d\theta + \widetilde{\Lambda}_1\, d\theta + \Big(\theta + \beta\, b\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big)\Big)\, d\Lambda_2 + \Lambda_2\, d\theta.$$

Furthermore, we have the boundary condition $\Lambda_2(\theta_2) = 0$ (since Λ_2 cannot have a jump at θ_2). Putting $\widetilde{\Lambda}_2 = \Lambda_2 + \widetilde{\Lambda}_1(\theta_2)$, this equation simplifies slightly to

$$0 = -G\, d\theta + \Big(\theta + \beta\, b\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big)\Big)\, d\widetilde{\Lambda}_2 + \widetilde{\Lambda}_2\, d\theta,$$

with boundary condition $\widetilde{\Lambda}_2(\theta_2) = \widetilde{\Lambda}_1(\theta_2)$.

G.2. Cases 2 to 5. In order to handle the remaining cases, we need a unified construction. (This construction includes Case 1 too.) It is more convenient to work in terms of the distribution function $\Psi = \Psi(\,\cdot\,; \theta_1)$ of the total multiplier $d\widetilde{\Lambda}_1 + d\Lambda_2$, and to view this as a function on $\left[0, \overline{\theta}\right]$. The construction is then very simple. For all $\theta_1 \in \left(0, \overline{\theta}\right]$:

1. put $\Psi = G$ on $[0, \theta_1]$;
2. if $\theta_1 < \overline{\theta}$ (so that Ψ is not yet defined on the whole of $\left[0, \overline{\theta}\right]$), then let Ψ be the unique bounded solution of the o.d.e.

$$0 = -G + (\theta - \theta_1)\, \Psi' + \Psi$$

on $\big(\theta_1, \theta_2 \wedge \overline{\theta}\,\big]$, i.e. put

$$\Psi(\theta) = \frac{1}{\theta - \theta_1} \int_{\theta_1}^{\theta} G(t)\, dt;$$

3. if $\theta_2 < \overline{\theta}$ (so that Ψ is still not defined on the whole of $\big[\,0, \overline{\theta}\,\big]$), then let Ψ be the unique solution of the o.d.e.

$$0 = -G + \Big(\theta + \beta\, b\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big)\Big)\, \Psi' + \Psi$$

on $\big(\theta_2, \overline{\theta}\,\big]$ with boundary condition

$$\Psi(\theta_2) = \frac{1}{\theta_2 - \theta_1} \int_{\theta_1}^{\theta_2} G(t)\, dt.$$

Then, using arguments similar to those of the preceding section, it is easy to verify that the three necessary conditions (4), (5) and (13) together imply that, for all $\theta_1 \in \big(\,0, \overline{\theta}\,\big]$, Ψ must take the given form.

Remark 7. *We can easily extend the construction of Ψ to include the case $\theta_1 = 0$. Indeed, in line with the construction above, we can let $\Psi(\,\cdot\,; 0)$ be the unique bounded solution of the o.d.e.*

$$0 = -G + \Big(\theta + \beta\, b\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big)\Big)\, \Psi' + \Psi$$

on $\big(\,0\,, \overline{\theta}\,\big]$.

Remark 8. *With this definition of $\Psi(\,\cdot\,; 0)$, $\Psi(\,\cdot\,; \theta_1)$ is independent of θ_1 for $\theta_1 \in$* $\big[\,0, \frac{1}{1+\pi}\,\underline{\theta}\,\big]$.

H. Existence of an Optimum

For all $\theta_1 \in \big(\,0, \overline{\theta}\,\big]$, we have shown that there exists a unique $\Psi = \Psi(\,\cdot\,; \theta_1)$ satisfying the two complementary slackness conditions (4) and (5) and the measure equation (13). This $\Psi(\,\cdot\,; \theta_1)$ does not in general satisfy the boundary condition (12). The purpose of the current section is to establish that there is at least one choice of θ_1 for which (12) is satisfied.

For all $\theta_1 \in \big(0, \overline{\theta}\,\big]$, put $\psi(\theta_1) = \Psi\big(\,\overline{\theta}\,;\theta_1\big)$. Now consider G. We certainly have $G = 0$ on $(\,0, \underline{\theta}\,)$ and $G = F(\,\overline{\theta}\,)$ on $\big[\,\overline{\theta}\,,\infty\,\big)$. Furthermore Assumption A3 tells us that G is first increasing (on $(0, \theta_M)$) and then decreasing (on (θ_M, ∞)). It is therefore obvious that there exists $\underline{\theta}_F \in \big[\,\underline{\theta}\,, \overline{\theta}\,\big]$ such that $G < F(\,\overline{\theta}\,)$ on $(\,0, \underline{\theta}_F)$ and $G \geq F(\,\overline{\theta}\,)$ on $\big[\,\underline{\theta}_F, \overline{\theta}\,\big]$. Our first lemma sharpens this observation.

Lemma 9. *There exists* $\overline{\theta}_F \in \big[\,\underline{\theta}\,, \overline{\theta}\,\big)$ *such that* $G \leq F(\,\overline{\theta}\,)$ *on* $\big(0, \overline{\theta}_F\big)$ *and* $G > F(\,\overline{\theta}\,)$ *on* $\big(\,\overline{\theta}_F, \overline{\theta}\,\big)$.

Proof. Suppose first that $\theta_M < \overline{\theta}$. Suppose further that there exists $\xi_0 \in \big(\theta_M, \overline{\theta}\,\big)$ such that $G(\xi_0) = G(\overline{\theta}\,)$. Since $G' \leq 0$ on $(\theta_M, \infty) \supset \big(\xi_0, \overline{\theta}\,\big]$, it follows that $G' = 0$ on $\big(\xi_0, \overline{\theta}\,\big]$. We also know that $G = F(\,\overline{\theta}\,)$ on $\big[\,\overline{\theta}\,,\infty\big)$, and therefore that $G' = 0$ on $\big(\,\overline{\theta}\,,\infty\big)$. Overall, then, $G' = 0$ on (ξ_0, ∞). Hence $\theta^\alpha\, F'(\theta)$ is constant on (ξ_0, ∞), where $\alpha = \frac{2-\beta}{1-\beta}$. (Cf. the proof of Proposition 1.) Hence $F' = 0$ on (ξ_0, ∞). But this contradicts the fact that $\overline{\theta}$ is the maximum of the support of F. We may therefore conclude that $G > G(\,\overline{\theta}\,) = F(\,\overline{\theta}\,)$ on $\big(\theta_M, \overline{\theta}\,\big)$.

Suppose second that $\theta_M = \overline{\theta}$. Then $G' \geq 0$ on $(0, \overline{\theta}\,)$. Hence $\theta^\alpha\, F'(\theta)$ is non-decreasing on $(0, \overline{\theta}\,)$. In particular, if we fix $\xi_1 \in \big(\,\underline{\theta}\,, \overline{\theta}\,\big)$, then we will have $\theta^\alpha\, F'(\theta) \geq \xi_1^\alpha\, F'(\xi_1)$ for all $\theta \in \big(\,\xi_1\,, \overline{\theta}\,\big)$. Letting $\theta \uparrow \overline{\theta}$, it follows that $\overline{\theta}^{\,\alpha}\, F'(\overline{\theta}-) \geq \xi_1^\alpha\, F'(\xi_1)$. But $F'(\xi_1) > 0$, since ξ_1 lies in the interior of the support of F. Hence $F'(\,\overline{\theta}-) > 0$. Hence $G(\,\overline{\theta}-) = (1-\beta)\,\overline{\theta}\, F'(\,\overline{\theta}-) + F(\,\overline{\theta}\,) > F(\,\overline{\theta}\,)$. Hence there exists $\varepsilon > 0$ such that $G > F(\,\overline{\theta}\,)$ on $\big(\,\overline{\theta} - \varepsilon, \overline{\theta}\,\big)$.

Overall, then, we have the following picture: $G = 0$ on $(0, \underline{\theta}\,)$; there exists $\xi_2 \in \big[\,\underline{\theta}\,, \overline{\theta}\,\big)$ such that $G > F(\,\overline{\theta}\,)$ on $\big(\,\xi_2, \overline{\theta}\,\big)$; and $G = F(\,\overline{\theta}\,)$ on $\big[\,\overline{\theta}\,,\infty\big)$. Furthermore Assumption A3 ensures that $\big\{\theta \mid G(\theta) > F(\,\overline{\theta}\,)\big\}$ is an interval. We may therefore put $\overline{\theta}_F = \inf \big\{\theta \mid \theta \in \big[\,\underline{\theta}\,, \overline{\theta}\,\big)\,, G(\theta) > F(\,\overline{\theta}\,)\big\}$. ■

Now, it follows from the construction of Ψ given in Section G.2 that $\Psi(\,\overline{\theta}\,;\theta_1)$ is a convex combination of the values of G on the interval $\big(\,\theta_1, \overline{\theta}\,\big)$. Combining this observation with Lemma 9, we obtain:

Lemma 10. $\psi > F(\,\overline{\theta}\,)$ *on* $\big[\,\overline{\theta}_F\,, \overline{\theta}\,\big)$. ■

We also have:

Lemma 11. *$\psi < F(\overline{\theta})$ on $\left(0, \frac{1}{1+\pi}\,\underline{\theta}\right]$.*

Proof. We begin by noting that $\Psi = \Psi(\,\cdot\,;\theta_1)$ is independent of θ_1 for $\theta_1 \in \left[0, \frac{1}{1+\pi}\,\underline{\theta}\right]$. It is therefore the unique bounded solution of the o.d.e.

$$0 = -G + \left(\theta + \beta\, b\!\left(\tfrac{\theta}{(1+\pi)\,\beta}\right)\right) \Psi' + \Psi \tag{14}$$

on $\left(0\,, \overline{\theta}\right]$. We compare Ψ with the function Φ which is the unique bounded solution of the o.d.e.

$$0 = -G + \theta\,\Phi' + \Phi \tag{15}$$

on $\left(0\,, \overline{\theta}\right]$. Now

$$\Phi(\theta) = \frac{1}{\theta}\int_0^\theta G(t)\,dt = (1-\beta)\,F(\theta) + \beta\,\frac{1}{\theta}\int_0^\theta F(t)\,dt.$$

Hence: $\Phi = 0$ on $\left(0\,, \overline{\theta}\right]$; and $0 < \Phi < F$ on $\left(\underline{\theta}\,, \overline{\theta}\right]$. Hence

$$\Phi' = \frac{G - \Phi}{\theta} \geq \frac{F - \Phi}{\theta} \geq 0$$

on $\left(0\,, \overline{\theta}\right]$, with strict inequality on $\left(\underline{\theta}\,, \overline{\theta}\right]$. Furthermore, Φ is a supersolution of the equation for Ψ. Indeed, we have

$$\begin{aligned} & -G + \left(\beta\, b\!\left(\tfrac{\theta}{(1+\pi)\,\beta}\right) + \theta\right) \Phi' + \Phi \\ = \quad & -G + \beta\, b\!\left(\tfrac{\theta}{(1+\pi)\,\beta}\right) \Phi' + \theta\,\Phi' + \Phi \end{aligned}$$

(on rearranging)

$$= \beta\, b\!\left(\tfrac{\theta}{(1+\pi)\,\beta}\right) \Phi'$$

(since Φ satisfies (15))

$$\geq 0$$

on $\left(0\,, \overline{\theta}\right]$, with strict inequality on $\left(\underline{\theta}\,, \overline{\theta}\right]$. That is, Φ is a supersolution of the equation for Ψ, and a strict supersolution on $\left(\underline{\theta}\,, \overline{\theta}\right]$. Hence $\Phi > \Psi$ on $\left(\underline{\theta}\,, \overline{\theta}\right]$. In particular, $\psi(0) = \Psi(\overline{\theta}\,) < \Phi(\overline{\theta}\,) < F(\overline{\theta}\,)$. The general case now follows on noting that $\Psi(\,\cdot\,;\theta_1) = \Psi(\,\cdot\,;0)$ for all $\theta_1 \in \big(0, \frac{1}{1+\pi}\,\underline{\theta}\,\big]$. Cf. the remark at the end of Section G.2. ■

Since ψ is continuous, we can combine Lemmas 10 and 11 to obtain:

Proposition 12. *There exists $\theta_1 \in \big(\frac{1}{1+\pi}\,\underline{\theta}\,, \overline{\theta}_F\big)$ such that $\psi(\theta_1) = F(\overline{\theta}\,)$.* ■

That is, there exists $\theta_1 \in \big(\frac{1}{1+\pi}\,\underline{\theta}, \overline{\theta}_F\big)$ such that equation (12) is satisfied. However, we still need to verify that the multipliers associated with any such θ_1 are non-negative.

I. Non-Negativity of the Multiplier

In this section we show that, if $\theta_1 \in \big(0, \overline{\theta}_F\big)$ is such that $\psi(\theta_1) \leq F(\overline{\theta}\,)$, then $\Psi = \Psi(\,\cdot\,;\theta_1)$ is non-decreasing on $\big[\,0, \overline{\theta}\,\big]$. We treat the intervals $[0, \theta_1]$, $\big(\theta_1, \overline{\theta}_F\big)$ and $\big[\,\overline{\theta}_F, \overline{\theta}\,\big]$ in turn. We begin with a simple lemma.

Lemma 13. $\overline{\theta}_F \leq \theta_M$.

Proof. In the light of Lemma 9, $\sup G > F(\overline{\theta}\,)$. Moreover it follows from the definition of θ_M that $\sup G = \max\{G_L(\theta_M), G(\theta_M)\}$. There are therefore two possibilities. Either $G_L(\theta_M) > F(\overline{\theta}\,)$, in which case there is a left neighbourhood of θ_M on which $G > F(\overline{\theta}\,)$, and therefore $\overline{\theta}_F < \theta_M$. Or $G(\theta_M) > F(\overline{\theta}\,)$, in which case necessarily $\overline{\theta}_F \leq \theta_M$. ■

Lemma 14. $G' \geq 0$ *on* $\big[\,0, \overline{\theta}_F\,\big]$.

Proof. From Lemma 13 we know that $\big[\,0, \overline{\theta}_F\,\big) \subset [\,0, \theta_M\,)$. Hence $G' \geq 0$ on $\big[\,0, \overline{\theta}_F\,\big)$. However, $G \leq F(\overline{\theta}\,)$ on $\big[\,0, \overline{\theta}_F\,\big)$ and $G > F(\overline{\theta}\,)$ on $\big(\,\overline{\theta}_F, \overline{\theta}\,\big)$. Hence $\Delta G(\,\overline{\theta}_F) \geq 0$. Hence $G' \geq 0$ on $\big[\,0, \overline{\theta}_F\,\big]$. ■

Proposition 15. *Suppose that $\psi(\theta_1) \leq F(\overline{\theta}\,)$. Then $\Psi' \geq 0$ on $[0, \theta_1]$.*

Proof. Since $\psi(\theta_1) \leq F(\overline{\theta}\,)$, Lemma 10 implies that $\theta_1 < \overline{\theta}_F$. Hence $[0, \theta_1] \subset \big[0, \overline{\theta}_F\big)$. But $\Psi = G$ on $[0, \theta_1]$ by construction of Ψ, and Lemma 14 tells us that $G' \geq 0$ on $\big[0, \overline{\theta}_F\big]$. It follows that $\Psi' \geq 0$ on $[0, \theta_1]$. ■

In order to show that $\Psi' \geq 0$ on $\left(\theta_1, \overline{\theta}\,\right]$, we use the fact that Ψ solves

$$0 = -G + (\theta - \theta_1)\,\Psi' + \Psi \tag{16}$$

on $\left(\theta_1, \theta_2 \wedge \overline{\theta}\,\right]$ and

$$0 = -G + \left(\theta + \beta\, b\!\left(\tfrac{\theta}{(1+\pi)\,\beta}\right)\right)\Psi' + \Psi \tag{17}$$

on $\left(\theta_2 \wedge \overline{\theta}\,, \overline{\theta}\,\right]$. We also make use of the corresponding o.d.e. for θ, namely

$$\dot{\theta} = -\left(\theta - \theta_1\right) \tag{18}$$

on $\left(\theta_1, \theta_2 \wedge \overline{\theta}\,\right]$ and

$$\dot{\theta} = -\left(\theta + \beta\, b\!\left(\tfrac{\theta}{(1+\pi)\,\beta}\right)\right) \tag{19}$$

on $\left(\theta_2 \wedge \overline{\theta}\,, \overline{\theta}\,\right]$. Specifically, for all $g, h \in \left(\theta_1, \overline{\theta}\,\right]$ such that $h < g$, let $T(h,g)$ denote the time at which the solution of the o.d.e. (18-19) for θ starting from g hits h, and put $S(h,g) = \exp(-T(h,g))$. Notice that $S(\cdot, g)$ increases from 0 at θ_1 to 1 at g.

Lemma 16. *Suppose that $\psi(\theta_1) \leq F(\overline{\theta}\,)$. Then $\Psi \leq G$ on (θ_1, θ_M).*

Proof. Since $\psi(\theta_1) \leq F(\overline{\theta}\,)$, Lemma 10 implies that $\theta_1 < \overline{\theta}_F$. Furthermore Lemma 13 tells us that $\overline{\theta}_F \leq \theta_M$. For all $g \in (\theta_1, \theta_M)$, we therefore have

$$\Psi(g) = \int_{\theta_1}^{g} \tfrac{\partial S}{\partial h}(h,g)\, G(h)\, dh \leq \int_{\theta_1}^{g} \tfrac{\partial S}{\partial h}(h,g)\, G(g-)\, dh$$

(with equality iff $G(g-) = G(\theta_1)$)

$$= G(g-).$$

Taking limits from the right (and using the continuity of Ψ and the right continuity of G) then yields $\Psi \leq G$. ■

Lemma 17. *The sign of Ψ' coincides with that of $G - \Psi$ on $\left(\theta_1, \overline{\theta}\,\right]$.*

Proof. We have

$$\Psi' = \frac{G - \Psi}{\theta - \theta_1}$$

on $\big(\theta_1, \theta_2 \wedge \overline{\theta}\big]$ (from equation (16)) and

$$\Psi' = \frac{G - \Psi}{\theta + \beta\, b\Big(\frac{\theta}{(1+\pi)\,\beta}\Big)}$$

on $\big(\theta_2 \wedge \overline{\theta}\,, \overline{\theta}\big]$ (from equation (17)). Bearing in mind that we have $\theta - \theta_1 > 0$ on $\big(\theta_1, \theta_2 \wedge \overline{\theta}\big]$, it follows that the sign of Ψ' coincides with that of $G - \Psi$ on $\big(\theta_1, \theta_2 \wedge \overline{\theta}\big] \cup \big(\theta_2 \wedge \overline{\theta}\,, \overline{\theta}\big] = \big(\theta_1, \overline{\theta}\,\big]$. ■

Proposition 18. *Suppose that* $\psi(\theta_1) \leq F(\overline{\theta}\,)$. *Then* $\Psi' \geq 0$ *on* $\big(\theta_1, \overline{\theta}_F\big)$.

Proof. From Lemma 16 we know that $\Psi \leq G$ on (θ_1, θ_M) and therefore on $\big(\theta_1, \overline{\theta}_F\big) \subset (\theta_1, \theta_M)$. Lemma 17 then implies that $\Psi' \geq 0$ there. ■

Proposition 19. *Suppose that* $\psi(\theta_1) \leq F(\overline{\theta}\,)$. *Then* $\Psi' > 0$ *on* $\big[\,\overline{\theta}_F, \overline{\theta}\,\big)$.

Proof. For all $\theta \in \big[\,\overline{\theta}_F, \overline{\theta}\,\big)$, we have

$$\Psi(\overline{\theta}\,) = S(\theta, \overline{\theta}\,)\, \Psi(\theta) + \int_{\theta}^{\overline{\theta}} \tfrac{\partial S}{\partial h}(h, \overline{\theta}\,)\, G(h)\, dh. \tag{20}$$

Since

$$\int_{\theta}^{\overline{\theta}} \tfrac{\partial S}{\partial h}(h, \overline{\theta}\,)\, dh = 1 - S(\theta, \overline{\theta}\,),$$

this means that $\Psi(\overline{\theta}\,)$ is a convex combination of $\Psi(\theta)$ and the values of G on $\big(\theta, \overline{\theta}\,\big]$. But $\Psi(\overline{\theta}\,) = \psi(\theta_1) \leq F(\overline{\theta}\,)$ and $G > F(\overline{\theta}\,)$ on $\big(\theta, \overline{\theta}\,\big)$. So we must have $\Psi(\theta) < F(\overline{\theta}\,)$. On the other hand, $G(\theta) \geq F(\overline{\theta}\,)$ since $\theta \geq \overline{\theta}_F$. Lemma 17 therefore implies that $\Psi'(\theta) > 0$. ■

J. Uniqueness of the Optimum

At this point we have established that the utility curve associated with θ_1 solves the main maximization problem (3) iff $\psi(\theta_1) = F(\overline{\theta}\,)$. Furthermore $\psi < F(\overline{\theta}\,)$ on $\big[\,0, \frac{1}{1+\pi}\,\underline{\theta}\,\big]$ and $\psi > F(\overline{\theta}\,)$ on $\big[\,\overline{\theta}_F\,, \overline{\theta}\,\big)$. Hence there exists $\theta_1 \in \big(\frac{1}{1+\pi}\,\underline{\theta}\,, \overline{\theta}_F\big)$ such that $\psi(\theta_1) = F(\overline{\theta}\,)$. In the current section, we show that the set of θ_1 for which $\psi(\theta_1) = F(\overline{\theta}\,)$ is an interval. Furthermore, if we strengthen Assumption A3 by requiring that G is strictly increasing to the left of its peak, then $\psi' > 0$ over the relevant range. It then follows that there is a unique θ_1 for which $\psi(\theta_1) = F(\overline{\theta})$. This result is limited: it shows that – under the strengthened version of A3 – there is exactly one solution to problem (3) within our one-parameter family of utility curves; but it does not show that that there is exactly one solution to problem (3) in Ω. It is, however, very suggestive.

The main idea of the proof is to find an explicit formula for ψ', and then use this formula to determine the sign of ψ'. Of course, the formula depends on whether $\theta_1 < \frac{1}{1+\pi}\overline{\theta}$ or $\theta_1 > \frac{1}{1+\pi}\overline{\theta}$. In the former case: $\theta_2 = (1+\pi)\,\theta_1 < \overline{\theta}$; the maximum-penalty constraint is strictly binding; and the types in the range $\left[\theta_2, \overline{\theta}\,\right]$ will choose to incur the consumption penalty. In the latter case: $\theta_2 = (1+\pi)\,\theta_1 > \overline{\theta}$; the maximum-penalty constraint is strictly slack; and no type will choose to incur the consumption penalty.

In order to state the formula for ψ', it will be helpful to introduce the functions ϕ, χ, ζ and η given by the formulae

$$\phi(\theta_1) = \frac{1}{\theta_2 - \theta_1} \int_{\theta_1}^{\theta_2} G(\theta)\, d\theta \tag{21}$$

for all $\theta_1 \in (0, \infty)$ (where we have suppressed the dependence of θ_2 on θ_1),

$$\chi(\theta_1) = \frac{1}{\overline{\theta} - \theta_1} \int_{\theta_1}^{\overline{\theta}} G(\theta)\, d\theta \tag{22}$$

for all $\theta_1 \in \left(0, \overline{\theta}\,\right)$,

$$\zeta(\theta_1) = \frac{\frac{\beta}{1+\pi}\, b\!\left(\frac{\theta_1}{\beta}\right)}{\theta_1\left(\theta_1 + \frac{\beta}{1+\pi}\, b\!\left(\frac{\theta_1}{\beta}\right)\right)}\left(G(\theta_2) - \phi(\theta_1)\right) + \frac{1}{\pi\,\theta_1}\left(G(\theta_2) - G(\theta_1)\right) \tag{23}$$

for all $\theta_1 \in (0, \infty)$ (where we have suppressed the dependence of θ_2 on θ_1) and

$$\eta(\theta_1) = \frac{\chi(\theta_1) - G(\theta_1)}{\overline{\theta} - \theta_1} \tag{24}$$

for all $\theta_1 \in \big(0, \overline{\theta}\,\big)$.

Lemma 20. *$\psi'(\theta_1) = S(\theta_2, \overline{\theta}\,)\,\zeta(\theta_1)$ for $\theta_1 \in \big(0, \frac{1}{1+\pi}\,\overline{\theta}\,\big)$.*

Proof. Equation (20) gives

$$\psi(\theta_1) = \int_{\theta_2}^{\overline{\theta}} \tfrac{\partial S}{\partial h}(h, \overline{\theta}\,)\, G(h)\, dh + S(\theta_2, \overline{\theta}\,)\,\phi(\theta_1),$$

where we have used the fact that $\Psi(\,\overline{\theta}\,; \theta_1) = \psi(\theta_1)$ and $\Psi(\theta_2; \theta_1) = \phi(\theta_1)$. Hence

$$\begin{aligned}\psi' &= -\tfrac{\partial S}{\partial h}(\theta_2, \overline{\theta}\,)\, G(\theta_2)\,\theta_2' + \tfrac{\partial S}{\partial h}(\theta_2, \overline{\theta}\,)\,\theta_2'\,\phi + S(\theta_2, \overline{\theta}\,)\,\phi' \\ &= \exp(-T(\theta_2, \overline{\theta}\,))\left(\tfrac{\partial T}{\partial h}(\theta_2, \overline{\theta}\,)\,\theta_2'\,(G(\theta_2) - \phi) + \phi'\right),\end{aligned}$$

where we have suppressed the dependence of ϕ and ψ on θ_1, and where θ_2' and ϕ' denote the derivatives of θ_2 and ϕ with respect to θ_1. Furthermore

$$\phi = \frac{1}{\theta_2 - \theta_1} \int_{\theta_1}^{\theta_2} G(t)\, dt$$

and

$$T(\theta_2, \overline{\theta}\,) = \int_{\theta_2}^{\overline{\theta}} \frac{dt}{t + \beta\, b\Big(\frac{t}{(1+\pi)\,\beta}\Big)}.$$

Hence

$$\begin{aligned}\phi' &= -\frac{\theta_2' - 1}{(\theta_2 - \theta_1)^2} \int_{\theta_1}^{\theta_2} G(t)\, dt + \frac{1}{\theta_2 - \theta_1}\,(G(\theta_2)\,\theta_2' - G(\theta_1)) \\ &= \frac{1}{\theta_2 - \theta_1}\,(-\pi\,\phi + ((1+\pi)\,G(\theta_2) - G(\theta_1))) \\ &= \frac{1}{\theta_1}\,(G(\theta_2) - \phi) + \frac{1}{\pi\,\theta_1}\,(G(\theta_2) - G(\theta_1))\end{aligned}$$

and

$$\tfrac{\partial T}{\partial h}(\theta_2, \overline{\theta}\,) = -\frac{1}{\theta_2 + \beta\, b\Big(\frac{\theta_2}{(1+\pi)\,\beta}\Big)}. \tag{25}$$

Overall, then,

$$\exp(T(\theta_2,\overline{\theta}\,))\,\psi' = \tfrac{\partial T}{\partial h}(\theta_2,\overline{\theta}\,)\,\theta_2'\,(G(\theta_2)-\phi)+\phi'$$

$$= -\frac{1+\pi}{\theta_2+\beta\, b\Big(\frac{\theta_2}{(1+\pi)\,\beta}\Big)}\,(G(\theta_2)-\phi)+\frac{1}{\theta_1}(G(\theta_2)-\phi)+\frac{1}{\pi\,\theta_1}(G(\theta_2)-G(\theta_1))$$

$$= \left(\frac{1}{\theta_1}-\frac{1}{\theta_1+\frac{\beta}{1+\pi}\,b\Big(\frac{\theta_1}{\beta}\Big)}\right)(G(\theta_2)-\phi)+\frac{1}{\pi\,\theta_1}(G(\theta_2)-G(\theta_1))$$

(collecting terms in $(G(\theta_2)-\phi)$ and $(G(\theta_2)-G(\theta_1))$, and using the fact that $\theta_2 = (1+\pi)\,\theta_1$)

$$= \frac{\frac{\beta}{1+\pi}\,b\Big(\frac{\theta_1}{\beta}\Big)}{\theta_1\Big(\theta_1+\frac{\beta}{1+\pi}\,b\Big(\frac{\theta_1}{\beta}\Big)\Big)}\,(G(\theta_2)-\phi)+\frac{1}{\pi\,\theta_1}\,(G(\theta_2)-G(\theta_1)).$$

Making ψ' the subject of this equation, we obtain the required result. ■

The second of the two formulae for ψ' is given by the following lemma.

Lemma 21. *$\psi' = \eta$ on $\big[\frac{1}{1+\pi}\,\overline{\theta}\,,\overline{\theta}\,\big)$.*

Proof. We have $\psi = \chi$ on $\big[\frac{1}{1+\pi}\,\overline{\theta}\,,\overline{\theta}\,\big)$. Moreover it is easy to check that

$$\chi'(\theta_1) = \frac{\chi(\theta_1)-G(\theta_1)}{\overline{\theta}-\theta_1}$$

on $\big(0,\overline{\theta}\,\big)$. ■

There are now two main cases to consider. The more general of the two main cases occurs when $\frac{1}{1+\pi}\,\overline{\theta} < \overline{\theta}_F$. In this case, there are three main subcases to consider:

Subcase 1 $\theta_1 \in \big(\frac{1}{1+\pi}\,\underline{\theta}\,,\frac{1}{1+\pi}\,\overline{\theta}_F\big]$. In this subcase, the maximum-penalty constraint is strictly binding in the sense that $\theta_2 < \overline{\theta}$. I.e. all the types in the non-trivial range of $\big[\,\theta_2,\overline{\theta}\,\big]$ choose to make an early withdrawal from the penalty account.

Subcase 2 $\theta_1 \in \big[\frac{1}{1+\pi}\,\overline{\theta}_F,\frac{1}{1+\pi}\,\overline{\theta}\,\big]$. In this subcase, the maximum-penalty constraint is weakly binding in the sense that $\theta_2 \le \overline{\theta}$.

Subcase 3 $\theta_1 \in \big[\frac{1}{1+\pi}\,\overline{\theta}\,, \overline{\theta}_F\big)$. In this subcase, the maximum-penalty constraint is weakly slack in the sense that $\theta_2 \geq \overline{\theta}$.

The less general of the two main cases occurs when $\frac{1}{1+\pi}\,\overline{\theta} \geq \overline{\theta}_F$. In this case, the third subcase does not arise.

We deal with both of the two main cases simultaneously. The first subcase is settled by the following lemma.

Lemma 22. *Suppose that* $\theta_1 \in \big(0, \frac{1}{1+\pi}\,\overline{\theta}_F\big]$. *Then* $\psi'(\theta_1) \geq 0$.

{

Since $\theta_1 \leq \frac{1}{1+\pi}\,\overline{\theta}_F$, we certainly have $\theta_1 < \frac{1}{1+\pi}\,\overline{\theta}$. Hence the formula $\psi'(\theta_1) = S(\theta_2, \overline{\theta}\,)\,\zeta(\theta_1)$ applies.

The inequality presumably concerns the right-continuous version of ψ', even at $\frac{1}{1+\pi}\,\overline{\theta}_F$. The older statement "$G(\theta_2) \geq \phi(\theta_1)$ (with equality iff $G_L(\theta_2) = G(\theta_1)$)" cannot therefore make sense: there could be a jump in G at θ_2.

}

Proof. The proof relies on the formula $\psi'(\theta_1) = S(\theta_2, \overline{\theta}\,)\,\zeta(\theta_1)$ given in Lemma 20. This formula is valid for $\theta_1 \in \big(0, \frac{1}{1+\pi}\,\overline{\theta}\,\big) \supset \big(0, \frac{1}{1+\pi}\,\overline{\theta}_F\big]$.

We have $[\theta_1, \theta_2] \subset \big(0, \overline{\theta}_F\big]$ and therefore $G' \geq 0$ on $[\theta_1, \theta_2]$. Hence $G(\theta_2) \geq \phi(\theta_1)$ and $G(\theta_2) \geq G(\theta_1)$. It then follows from formula (23) that $\zeta(\theta_1) \geq 0$ (with equality iff $G(\theta_2) = G(\theta_1)$), and thence that $\psi'(\theta_1) \geq 0$ (with equality iff $G(\theta_2) = G(\theta_1)$). ■

We now turn to the second subcase.

Lemma 23. *Suppose that* $\theta_1 \in \big[\frac{1}{1+\pi}\,\overline{\theta}_F, \frac{1}{1+\pi}\,\overline{\theta}\,\big]$ *and that* $\psi(\theta_1) \leq F(\overline{\theta}\,)$. *Then* $G(\theta_2) - \phi(\theta_1) \geq 0$.

Proof. We have $\theta_2 \in \big[\,\overline{\theta}_F, \overline{\theta}\,\big]$ and therefore

$$G(\theta_2) \geq F(\overline{\theta}\,)$$

(with strict inequality if $\theta_1 \in \big(\frac{1}{1+\pi}\,\overline{\theta}_F, \frac{1}{1+\pi}\,\overline{\theta}\,\big)$, because then $\theta_2 \in \big(\,\overline{\theta}_F, \overline{\theta}\,\big)$)

$$\geq \psi(\theta_1) = \Psi(\overline{\theta}\,; \theta_1)$$

(by assumption and by definition of ψ respectively)

$$\geq \Psi(\theta_2; \theta_1)$$

(with strict inequality if $\theta_1 \in \left[\frac{1}{1+\pi}\,\overline{\theta}_F, \frac{1}{1+\pi}\,\overline{\theta}\,\right)$, because then $\theta_2 < \overline{\theta}$)

$$= \phi(\theta_1)$$

(by definition of ϕ). ■

Lemma 24. *Suppose that* $\theta_1 \in \left[\frac{1}{1+\pi}\,\overline{\theta}_F, \frac{1}{1+\pi}\,\overline{\theta}\,\right]$ *and that* $\psi(\theta_1) \leq F(\,\overline{\theta}\,)$. *Then* $G(\theta_2) - G(\theta_1) > 0$.

Proof. We have

$$G(\theta_2) \geq F(\,\overline{\theta}\,)$$

(with strict inequality if $\theta_1 \in \left(\frac{1}{1+\pi}\,\overline{\theta}_F, \frac{1}{1+\pi}\,\overline{\theta}\,\right)$)

$$\geq \psi(\theta_1) = \Psi(\,\overline{\theta}\,; \theta_1) > \Psi(\theta_1; \theta_1) = G(\theta_1)$$

(by assumption, by definition of ψ, because $\theta_1 < \overline{\theta}$ and by construction of Ψ respectively). ■

Combining Lemmas 23 and 24, we obtain the following result about the right-hand derivative of ψ.

Lemma 25. *Suppose that* $\theta_1 \in \left[\frac{1}{1+\pi}\,\overline{\theta}_F, \frac{1}{1+\pi}\,\overline{\theta}\,\right)$ *and that* $\psi(\theta_1) \leq F(\,\overline{\theta}\,)$. *Then* $\psi'(\theta_1) > 0$.

Proof. The proof again relies on the formula $\psi'(\theta_1) = S(\theta_2, \overline{\theta}\,)\,\zeta(\theta_1)$ given in Lemma 20. In view of this formula, $\psi'(\theta_1) > 0$ if $G(\theta_2) \geq \phi(\theta_1)$ and $G(\theta_2) \geq G(\theta_1)$ with at least one strict inequality. But Lemmas 23 and 24 show that $G(\theta_2) \geq \phi(\theta_1)$ and $G(\theta_2) > G(\theta_1)$ respectively. ■

We also need the corresponding result about the left-hand derivative of ψ.

Lemma 26. *Suppose that* $\theta_1 \in \left(\frac{1}{1+\pi}\,\overline{\theta}_F, \frac{1}{1+\pi}\,\overline{\theta}\,\right]$ *and that* $\psi(\theta_1) \leq F(\,\overline{\theta}\,)$. *Then* $\psi'_L(\theta_1) > 0$.

Proof. The proof parallels that of Lemma 25, with minor changes. First of all, bearing in mind that ϕ is continuous, we have

$$\zeta_L(\theta_1) = \frac{\frac{\beta}{1+\pi}\, b\!\left(\frac{\theta_1}{\beta}\right)}{\theta_1\left(\theta_1 + \frac{\beta}{1+\pi}\, b\!\left(\frac{\theta_1}{\beta}\right)\right)}\,(G_L(\theta_2) - \phi(\theta_1)) + \frac{1}{\pi\,\theta_1}(G_L(\theta_2) - G_L(\theta_1)) \tag{26}$$

for all $\theta_1 \in (0, \infty)$. As in Lemma 20, we then have $\psi'_L(\theta_1) = S(\theta_2, \overline{\theta}\,)\,\zeta_L(\theta_1)$ for $\theta_1 \in \left(0, \frac{1}{1+\pi}\,\overline{\theta}\,\right]$. Next, just as the single peakedness of G implies that $G > F(\overline{\theta}\,)$ on $\left(\overline{\theta}_F, \overline{\theta}\right)$, so it also implies that $G_L > F(\overline{\theta}\,)$ on $\left(\overline{\theta}_F, \overline{\theta}\right)$. Arguing as in Lemma 23, we can therefore show that $G_L(\theta_2) - \phi(\theta_1) \geq 0$ for $\theta_1 \in \left(\frac{1}{1+\pi}\,\overline{\theta}_F, \frac{1}{1+\pi}\,\overline{\theta}\,\right]$, with strict inequality if $\theta_1 < \frac{1}{1+\pi}\,\overline{\theta}$. (We cannot however extend this to $\theta_1 \in \left[\frac{1}{1+\pi}\,\overline{\theta}_F, \frac{1}{1+\pi}\,\overline{\theta}\,\right]$, since we cannot deduce from the fact that $G_L > F(\overline{\theta}\,)$ on $\left(\overline{\theta}_F, \overline{\theta}\right)$ that $G_L \geq F(\overline{\theta}\,)$ at $\overline{\theta}_F$.) Next, as in Lemma 24, we have $G_L(\theta_2) - G_L(\theta_1) > 0$ on $\left(\frac{1}{1+\pi}\,\overline{\theta}_F, \frac{1}{1+\pi}\,\overline{\theta}\,\right]$. Indeed, as in the proof of that lemma, we can show that

$$G_L(\theta_2) \geq F(\overline{\theta}\,)$$

(with strict inequality if $\theta_1 \in \left(\frac{1}{1+\pi}\,\overline{\theta}_F, \frac{1}{1+\pi}\overline{\theta}\right)$)

$$\geq \psi(\theta_1) = \Psi(\overline{\theta}\,; \theta_1) > \Psi(\theta_1; \theta_1) = G(\theta_1).$$

In particular, $G(\theta_1) < F(\overline{\theta}\,)$; and therefore $\theta_1 < \overline{\theta}_F \leq \theta_M$; and therefore $G_L(\theta_1) \leq G(\theta_1)$. Finally, applying (26) yields the required result. ■

Next, we prove a lemma that will be needed for the third subcase.

Lemma 27. *$\chi > G$ on $\left[0, \overline{\theta}_F\right)$.*

Proof. For all $\theta_1 \in \left[0, \overline{\theta}_F\right)$, we have

$$\chi(\theta_1) = \frac{1}{\overline{\theta} - \theta_1}\int_{\theta_1}^{\overline{\theta}} G(\theta)\,d\theta = \frac{1}{\overline{\theta} - \theta_1}\int_{\theta_1}^{\overline{\theta}_F} G(\theta)\,d\theta + \frac{1}{\overline{\theta} - \theta_1}\int_{\overline{\theta}_F}^{\overline{\theta}} G(\theta)\,d\theta.$$

Moreover:

$$\int_{\theta_1}^{\overline{\theta}_F} G(\theta)\,d\theta \geq (\,\overline{\theta}_F - \theta_1)\,G(\theta_1),$$

since $G' \geq 0$ on $\left[0, \overline{\theta}_F\right]$ by Lemma 14; and

$$\int_{\overline{\theta}_F}^{\overline{\theta}} G(\theta)\, d\theta > (\,\overline{\theta} - \overline{\theta}_F)\, F(\,\overline{\theta}\,) \geq (\,\overline{\theta} - \overline{\theta}_F)\, G(\theta_1),$$

since $G > F(\,\overline{\theta}\,)$ on $\left(\overline{\theta}_F, \overline{\theta}\right)$ and (by Lemma 9) $G \leq F(\,\overline{\theta}\,)$ on $\left(0, \overline{\theta}_F\right)$. Hence

$$\chi(\theta_1) > \frac{\overline{\theta}_F - \theta_1}{\overline{\theta} - \theta_1}\, G(\theta_1) + \frac{\overline{\theta} - \overline{\theta}_F}{\overline{\theta} - \theta_1}\, G(\theta_1) = G(\theta_1),$$

as required. ■

We can now deal with the third subcase, which arises only in the first scenario.

Lemma 28. *Suppose that* $\frac{1}{1+\pi}\,\overline{\theta} < \overline{\theta}_F$ *– i.e. that we are in the first scenario – and that* $\theta_1 \in \left[\frac{1}{1+\pi}\,\overline{\theta}, \overline{\theta}_F\right)$. *Then* $\psi'(\theta_1) > 0$.

Proof. Since $\theta_1 \geq \frac{1}{1+\pi}\,\overline{\theta}$, we may apply Lemma 21 to obtain

$$\psi'(\theta_1) = \frac{\chi(\theta_1) - G(\theta_1)}{\overline{\theta} - \theta_1}.$$

Since $\theta_1 < \overline{\theta}_F$, we may apply Lemma 27 to obtain $\chi(\theta_1) - G(\theta_1) > 0$. The result follows. ■

Combining Lemmas 22, 25, 26 and 28, we obtain:

Proposition 29. *The set of* $\theta_1 \in \left(0, \overline{\theta}_F\right)$ *such that* $\psi(\theta_1) = F(\,\overline{\theta}\,)$ *is a closed interval.* ■

The idea behind the proof of the proposition is straightforward. We know from Proposition 12 that all solutions to the equation $\psi = F(\,\overline{\theta}\,)$ lie in $\left(\frac{1}{1+\pi}\,\underline{\theta}\,, \overline{\theta}_F\right)$. Hence, to prove the proposition, we need only show that $\psi' \geq 0$ on this interval. Furthermore this is what Lemma 22 (for the interval $\left(\frac{1}{1+\pi}\,\underline{\theta}\,, \frac{1}{1+\pi}\,\overline{\theta}_F\right]$), Lemmas 25 and 26 (for the interval $\left(\frac{1}{1+\pi}\,\overline{\theta}_F, \frac{1}{1+\pi}\,\overline{\theta}\,\right)$) and Lemma 28 (for the interval $\left[\frac{1}{1+\pi}\,\overline{\theta}, \overline{\theta}_F\right)$) seem to tell us. The only complication is that Lemmas 25 and 26 both require the side condition $\psi \leq F(\,\overline{\theta}\,)$. However, they make up for this by providing strict rather than weak inequalities. The proof of the Proposition does therefore go through.

Indeed, we actually have $\psi' > 0$ on the interval $\left(\frac{1}{1+\pi}\,\overline{\theta}_F, \overline{\theta}_F\right)$. Hence, the only way in which non-uniqueness can occur at all is if there is a non-trivial interval, contained in $\left(\frac{1}{1+\pi}\,\underline{\theta}\,, \frac{1}{1+\pi}\,\overline{\theta}_F\right]$, on which $\psi = F(\,\overline{\theta}\,)$. Unfortunately, it is possible to construct an example in which precisely this form of non-uniqueness occurs. The spirit of the example is that there exist $\theta_3, \theta_4 \in \left[\,\underline{\theta}\,, \overline{\theta}_F\right)$ such that: (i) $\theta_4 > (1+\pi)\,\theta_3$ (i.e. it is possible that the entire interval of types associated with the kink lies within $[\theta_3, \theta_4)$); and (ii) $G_L(\theta_4) = G(\theta_3)$ (i.e. G is constant on $[\theta_3, \theta_4)$). It then follows that, if there exists $\theta_1 \in \left[\theta_3, \frac{1}{1+\pi}\,\theta_4\right]$ such that $\psi(\theta_1) = F(\,\overline{\theta}\,)$, then $\psi(\theta_1) = F(\,\overline{\theta}\,)$ for all $\theta_1 \in \left[\theta_3, \frac{1}{1+\pi}\,\theta_4\right]$.

There are two ways to eliminate this possibility. The first way is to ensure that G cannot have a "flat" of the type envisaged. The following assumption is more than sufficient to ensure this:

A4 G is strictly increasing on $[\,\underline{\theta}, \theta_M)$.[13]

We then have:

Proposition 30. *Suppose that Assumptions A1-A4 hold. Then there is a unique* $\theta_1 \in \left(0, \overline{\theta}_F\right)$ *such that* $\psi(\theta_1) = F(\,\overline{\theta}\,)$. ■

Working with Assumption A4 certainly simplifies our comparative statics. See Sections K-N below. However, we can still obtain satisfactory comparative-statics results without it. See Section O below.

The second way to eliminate the possibility of non-uniqueness is ensure that G cannot have a long enough flat:

Proposition 31. *Suppose that Assumptions A1-A3 hold, and that* $\pi > \frac{\overline{\theta}-\underline{\theta}}{\underline{\theta}}$. *Then there is a unique* $\theta_1 \in \left(0, \overline{\theta}_F\right)$ *such that* $\psi(\theta_1) = F(\,\overline{\theta}\,)$. ■

In particular, if $\pi = \infty$, then we there is a unique optimum within our one-parameter family of candidate optima.

[13] Notice that G is identically 0 on $(0, \underline{\theta}\,)$. It does not therefore make sense to require that G is strictly increasing on $(0, \theta_M)$.

K. Comparative Statics with A4

The analysis of Sections H-J shows that, for all $\pi \in [0, \infty)$, the set of solutions of the equation

$$\psi(\theta_1, \pi) = F(\,\overline{\theta}\,) \tag{27}$$

is a non-empty interval. We denote this interval by $\tau(\pi) = [\,\underline{\tau}(\pi), \overline{\tau}(\pi)\,]$. The purpose of the current section is to investigate the dependence of τ on π.

In order to simplify the exposition, it will be helpful to assume for the time being that A4 holds. This ensures that the interval $\tau(\pi)$ collapses to a single point, which we shall denote by $\tau_1(\pi)$. It also ensures that $\frac{\partial \psi}{\partial \theta_1}(\tau_1(\pi), \pi) > 0$.

If we assume further that all the functions involved are sufficiently smooth, then we can apply the implicit function to the equation

$$\psi(\tau_1(\pi), \pi) = F(\,\overline{\theta}\,)$$

to conclude that

$$\tau_1' = -\frac{\frac{\partial \psi}{\partial \pi}}{\frac{\partial \psi}{\partial \theta_1}}. \tag{28}$$

In particular: τ_1 will be increasing (decreasing) in π iff $\frac{\partial \psi}{\partial \pi} < 0$ $\left(\frac{\partial \psi}{\partial \pi} > 0\right)$; and the allocation to the illiquid account will be increasing (decreasing) in π iff $\frac{\partial \psi}{\partial \pi} > 0$ $\left(\frac{\partial \psi}{\partial \pi} < 0\right)$.

Motivated by these observations, we look first at the case in which the maximum-penalty constraint is strictly binding. More precisely, we put $\tau_2(\pi) = (1 + \pi)\,\tau_1(\pi)$, and we consider the case in which $\tau_2(\pi) < \overline{\theta}$. In other words, there is a non-trivial interval of types $\left(\tau_2(\pi), \overline{\theta}\,\right)$ who choose to consume out of the illiquid account and therefore pay the penalty for doing so. In this case we begin by finding explicit formulae for $\frac{\partial \psi}{\partial \pi}$ and $\frac{\partial \psi}{\partial \theta_1}$. We then go on to find conditions under which $\frac{\partial \psi}{\partial \pi} > 0$ and $\frac{\partial \psi}{\partial \theta_1} > 0$, thereby ensuring that $\tau_1' < 0$ (and hence that the allocation to the illiquid account will be strictly increasing in π).

We look second at the case in which the maximum-penalty constraint is strictly slack. More precisely, we consider the case in which $\tau_2(\pi) > \overline{\theta}$. In other words, even the highest type is not tempted to consume out of the illiquid account. In this case we again begin by finding explicit formulae for $\frac{\partial \psi}{\partial \pi}$ and $\frac{\partial \psi}{\partial \theta_1}$. It turns out that $\frac{\partial \psi}{\partial \pi} = 0$.

We therefore concentrate on finding conditions under which $\frac{\partial\psi}{\partial\theta_1} > 0$, thereby ensuring that $\tau_1' = 0$ (and hence that the allocation to the illiquid account does not change with π).

We look third at the intermediate case in which $\tau_2(\pi) = \overline{\theta}$. This case is important because it is $\tau_2(\pi)$ that determines whether we are in the strictly binding case $\tau_2(\pi) < \overline{\theta}$ or the strictly slack case $\tau_2(\pi) > \overline{\theta}$. Our analysis of the comparative statics of our problem is not therefore complete until we have understood how the transition between these two cases occurs.

L. The Strictly Binding Case

In this section we focus on the set V of (θ_1, π) such that

1. $\theta_1 \in \left(0, \overline{\theta}\right)$,
2. $\pi \in (0, \infty)$ and
3. $\theta_2 = (1+\pi)\,\theta_1 < \overline{\theta}$.

In other words, we do not impose the requirement that $\theta_1 = \tau_1(\pi)$ (i.e. that θ_1 be optimal for the given π), but we do require that the maximum-penalty constraint is binding (in the sense that types in the non-empty interval $\left(\theta_2, \overline{\theta}\right)$ are choosing to pay the penalty).

L.1. The formula for $\frac{\partial\psi}{\partial\pi}$. Consider the o.d.e.

$$\dot{\theta} = -\left(\theta + \beta\, b\left(\tfrac{\theta}{(1+\pi)\,\beta}\right)\right) \tag{29}$$

on $\left[\theta_2, \overline{\theta}\right]$, with initial condition $\theta(0) = \overline{\theta}$. Let $T(h;\pi)$ denote the first hitting time of $h \in \left[\theta_2, \overline{\theta}\right]$, and put $S(h;\pi) = \exp(-T(h;\pi))$. Then the formula for $\frac{\partial\psi}{\partial\pi}$ is given by the following proposition.

Proposition 32. *Suppose that $\theta_2 < \overline{\theta}$. Then*

$$\begin{aligned}\frac{\partial\psi}{\partial\pi}(\theta_1,\pi) = {} & \left(\frac{1}{\pi}\,S(\theta_2,\pi) - \theta_1\,\frac{\partial S}{\partial h}(\theta_2,\pi) - \frac{\partial S}{\partial\pi}(\theta_2,\pi)\right)(G(\theta_2) - \phi(\theta_1,\pi)) \\ & - \int_{\left(\theta_2,\overline{\theta}\right]} \frac{\partial S}{\partial\pi}(h,\pi)\,dG(h).\end{aligned}$$

Proof. Equation (20) can be written

$$\psi(\theta_1, \pi) = \int_{[\theta_2, \overline{\theta}]} \frac{\partial S}{\partial h}(h, \pi)\, G(h)\, dh + S(\theta_2, \pi)\, \phi(\theta_1, \pi).$$

Hence

$$\begin{aligned} \frac{\partial \psi}{\partial \pi} &= \int_{[\theta_2, \overline{\theta}]} \frac{\partial^2 S}{\partial h \partial \pi}(h, \pi)\, G(h)\, dh - \frac{\partial S}{\partial h}(\theta_2, \pi)\, G(\theta_2)\, \frac{\partial \theta_2}{\partial \pi} \\ &\quad + \left(\frac{\partial S}{\partial h}(\theta_2, \pi)\, \frac{\partial \theta_2}{\partial \pi} + \frac{\partial S}{\partial \pi}(\theta_2, \pi) \right) \phi + S(\theta_2, \pi)\, \frac{\partial \phi}{\partial \pi}, \end{aligned} \tag{30}$$

where we have suppressed the dependence of ψ and ϕ on θ_1 and π. Now:

$$\begin{aligned} \int_{[\theta_2, \overline{\theta}]} \frac{\partial^2 S}{\partial h \partial \pi}(h, \pi)\, G(h)\, dh &= \int_{[\theta_2, \overline{\theta}]} \frac{\partial^2 S}{\partial \pi \partial h}(h, \pi)\, G(h)\, dh \\ &= \left[\frac{\partial S}{\partial \pi}(h, \pi)\, G(h) \right]_{\theta_2 -}^{\overline{\theta}} - \int_{[\theta_2, \overline{\theta}]} \frac{\partial S}{\partial \pi}(h, \pi)\, dG(h) \\ &= -\frac{\partial S}{\partial \pi}(\theta_2, \pi)\, G(\theta_2 -) - \int_{[\theta_2, \overline{\theta}]} \frac{\partial S}{\partial \pi}(h, \pi)\, dG(h) \\ &= -\frac{\partial S}{\partial \pi}(\theta_2, \pi)\, G(\theta_2) - \int_{(\theta_2, \overline{\theta}]} \frac{\partial S}{\partial \pi}(h, \pi)\, dG(h), \end{aligned}$$

where we have used the fact that $\frac{\partial S}{\partial \pi}(\,\overline{\theta}\,) = 0$;

$$\frac{\partial \phi}{\partial \pi} = \frac{G(\theta_2) - \phi}{\theta_2 - \theta_1}\, \frac{\partial \theta_2}{\partial \pi} = \frac{G(\theta_2) - \phi}{\pi};$$

and

$$\frac{\partial \theta_2}{\partial \pi} = \theta_1.$$

Substituting into (30), we therefore obtain

$$\begin{aligned}\frac{\partial \psi}{\partial \pi} &= -\frac{\partial S}{\partial \pi}(\theta_2, \pi)\, G(\theta_2) - \int_{(\theta_2, \overline{\theta}]} \frac{\partial S}{\partial \pi}(h, \pi)\, dG(h) - \frac{\partial S}{\partial h}(\theta_2, \pi)\, G(\theta_2)\, \theta_1 \\ &\quad + \left(\frac{\partial S}{\partial h}(\theta_2, \pi)\, \theta_1 + \frac{\partial S}{\partial \pi}(\theta_2, \pi) \right) \phi + S(\theta_2, \pi)\, \frac{G(\theta_2) - \phi}{\pi} \\ &= -\frac{\partial S}{\partial \pi}(\theta_2, \pi)\, (G(\theta_2) - \phi(\theta_1, \pi)) - \int_{(\theta_2, \overline{\theta}]} \frac{\partial S}{\partial \pi}(h, \pi)\, dG(h) \\ &\quad + \left(\frac{1}{\pi}\, S(\theta_2, \pi) - \theta_1\, \frac{\partial S}{\partial h}(\theta_2, \pi) \right) (G(\theta_2) - \phi(\theta_1, \pi))\,.\end{aligned}$$

The required formula now follows on rearranging. ■

In view of Proposition 32, it is clear that there are three main contributions to $\frac{\partial \psi}{\partial \pi}$, namely:

1. $\dfrac{1}{\pi}\, S(\theta_2, \pi) - \theta_1\, \dfrac{\partial S}{\partial h}(\theta_2, \pi) - \dfrac{\partial S}{\partial \pi}(\theta_2, \pi)$;
2. $G(\theta_2) - \phi(\theta_1, \pi)$;
3. $-\displaystyle\int_{(\theta_2, \overline{\theta}]} \frac{\partial S}{\partial \pi}(h, \pi)\, dG(h)$.

We discuss these contributions in turn.

The first contribution can be signed quite generally:

Proposition 33. *Suppose that $\theta_2 \leq \overline{\theta}$. Then*

$$\frac{1}{\pi}\, S(\theta_2, \pi) - \theta_1\, \frac{\partial S}{\partial h}(\theta_2, \pi) - \frac{\partial S}{\partial \pi}(\theta_2, \pi) > 0.$$

In other words, Contribution 1 is strictly positive.

Proof. Explicit calculation shows that

$$\frac{1}{\pi}\, S(\theta_2, \pi) - \theta_1\, \frac{\partial S}{\partial h}(\theta_2, \pi) - \frac{\partial S}{\partial \pi}(\theta_2, \pi) = \frac{N}{D},$$

where

$$N = 1 + (1+\pi)\left(\frac{\theta_2}{\beta(1+\pi)}\right)^{1/\rho} + \left(\frac{\overline{\theta}}{\beta(1+\pi)}\right)^{1/\rho} + (1+\pi)\left(\frac{\theta_2}{\beta(1+\pi)}\right)^{1/\rho}\left(\frac{\overline{\theta}}{\beta(1+\pi)}\right)^{1/\rho} + \rho\pi\left(\left(\frac{\overline{\theta}}{\beta(1+\pi)}\right)^{1/\rho} - \left(\frac{\theta_2}{\beta(1+\pi)}\right)^{1/\rho}\right).$$

and

$$D = \pi\left(1 + (1+\pi)\left(\frac{\theta_2}{\beta(1+\pi)}\right)^{1/\rho}\right)^{1-\rho}\left(1 + (1+\pi)\left(\frac{\overline{\theta}}{\beta(1+\pi)}\right)^{1/\rho}\right)^{1+\rho}.$$

Now the last term in the formula for N is non-negative, since $\theta_2 \leq \overline{\theta}$. Hence $N > 0$. Finally, it is clear that $D > 0$. ■

The second contribution can only be signed when $\theta_1 = \tau_1(\pi)$ (or, more generally, when $\theta_2 \in (\underline{\theta}, \overline{\theta})$ and $\psi(\theta_1) \leq F(\overline{\theta})$). This, however, is enough for the purpose of our comparative statics.

Proposition 34. *Suppose that:*

1. $\theta_2 \in (\underline{\theta}, \overline{\theta})$*;*
2. $\psi(\theta_1) \leq F(\overline{\theta})$*;*
3. *Assumption A4 holds.*

Then $G(\theta_2) - \phi(\theta_1, \pi) > 0$*. In other words, Contribution 2 is strictly positive.*

Proof. We break the proof down into the cases $\theta_2 \in (\underline{\theta}, \overline{\theta}_F)$ and $\theta_2 \in [\overline{\theta}_F, \overline{\theta})$. In the first case, the proof parallels that of Lemma 22. We have $[\theta_1, \theta_2] \subset \left(\frac{1}{1+\pi}\underline{\theta}, \overline{\theta}_F\right) \subset (0, \theta_M)$ and $\theta_2 > \underline{\theta}$. Assumption A4 therefore implies that $G(\theta_2) > \phi(\theta_1, \pi)$. In the second case, it follows from the proof of Lemma 23 that $G(\theta_2) - \phi(\theta_1, \pi) > 0$. ■

The third contribution cannot be signed under our primary assumptions. It is, however, worth drawing attention to three special cases in which it can be signed.

In all three cases, the comparative statics end up going the same way: $\frac{\partial \psi}{\partial \pi} > 0$, and therefore the allocation to the illiquid account will be increasing in π. We state these three cases as separate propositions, corresponding to the cases $\rho < 1$, $\rho = 1$ and $\rho > 1$.

Proposition 35. *Suppose that:*

1. $\rho < 1$;
2. $G' \leq 0$ *on* $(\underline{\theta}, \infty)$ *(i.e.* $\theta_M = \underline{\theta}$*);*
3. $\theta_2 \in \left(\underline{\theta}, \overline{\theta}\right)$.

Then $-\int_{\left(\theta_2, \overline{\theta}\right]} \frac{\partial S}{\partial \pi}(h, \pi)\, dG(h) \geq 0$. *In other words, Contribution 3 is non-negative.*

Proof. It is easy to show that we have

$$\frac{\partial S}{\partial \pi}(h, \pi) \left\{ \begin{array}{ll} > 0 & \text{if } \rho < 1 \\ = 0 & \text{if } \rho = 1 \\ < 0 & \text{if } \rho > 1 \end{array} \right\}$$

for all $h \in \left[\theta_2, \overline{\theta}\right)$. Furthermore, we have

$$\frac{\partial S}{\partial \pi}(\overline{\theta}, \pi) = 0 \quad \text{for all } \rho,$$

because $S(\overline{\theta}, \pi) = 1$. We can therefore proceed as follows.

First, we know that $\theta_2 \in \left(\underline{\theta}, \overline{\theta}\right)$. Hence $G' \leq 0$ on $\left(\theta_2, \overline{\theta}\right] \subset (\underline{\theta}, \infty)$. Second, $\rho < 1$. Hence $\frac{\partial S}{\partial \pi}(\cdot, \pi) \geq 0$ on $\left(\theta_2, \overline{\theta}\right] \subset \left[\theta_2, \overline{\theta}\right]$. Putting these two observations together gives us the required conclusion. ■

Remark 36. *If* $G' \leq 0$ *on* $(\underline{\theta}, \infty)$ *then necessarily* $\Delta G(\underline{\theta}) > 0$. *Hence it is essential for the proof of Proposition 35 that we restrict attention to* $\theta_2 > \underline{\theta}$.

Proposition 37. *Suppose that:*

1. $\rho = 1$;

2. $\theta_2 \in \left(0, \overline{\theta}\right)$.

Then $-\int_{\left(\theta_2,\overline{\theta}\right]} \dfrac{\partial S}{\partial \pi}(h,\pi)\, dG(h) = 0$. *In other words, Contribution 3 is zero.*

Proof. This follows at once from the fact that $\frac{\partial S}{\partial \pi}(\cdot,\pi) = 0$ on $\left[\theta_2, \overline{\theta}\right]$. ■

Proposition 38. *Suppose that:*

1. $\rho > 1$;
2. $G' \geq 0$ *on* $\left(0, \overline{\theta}\right)$ *(i.e.* $\theta_M = \overline{\theta}$*);*
3. $\theta_2 \in \left(0, \overline{\theta}\right)$.

Then $-\int_{\left(\theta_2,\overline{\theta}\right]} \dfrac{\partial S}{\partial \pi}(h,\pi)\, dG(h) \geq 0$. *In other words, Contribution 3 is non-negative.*

Proof. Note first that $G' \geq 0$ on $\left(\theta_2, \overline{\theta}\right) \subset \left(0, \overline{\theta}\right)$. Second, $\rho > 1$. Hence $\dfrac{\partial S}{\partial \pi}(\cdot,\pi) < 0$ on $\left(\theta_2, \overline{\theta}\right) \subset \left[\theta_2, \overline{\theta}\right]$. (Cf. the proof of Proposition .) Third, $\frac{\partial S}{\partial \pi}(\overline{\theta}, \pi) = 0$. Putting these three observations together, we obtain

$$\begin{aligned} \int_{\left(\theta_2,\overline{\theta}\right]} \frac{\partial S}{\partial \pi}(h,\pi)\, dG(h) &= \int_{\left(\theta_2,\overline{\theta}\right)} \frac{\partial S}{\partial \pi}(h,\pi)\, dG(h) + \int_{\left[\overline{\theta},\overline{\theta}\right]} \frac{\partial S}{\partial \pi}(h,\pi)\, dG(h) \\ &= \int_{\left(\theta_2,\overline{\theta}\right)} \frac{\partial S}{\partial \pi}(h,\pi)\, dG(h) \leq 0, \end{aligned}$$

as required. ■

Remark 39. *If* $G' \geq 0$ *on* $\left(0, \overline{\theta}\right)$ *then necessarily* $\Delta G(\overline{\theta}) < 0$. *The fact that* $\frac{\partial S}{\partial \pi}(\overline{\theta}, \pi) = 0$ *therefore plays an essential role in the proof of Proposition 38.*

L.2. The formula for $\frac{\partial \psi}{\partial \theta_1}$. Let $T(h;\pi)$ and $S(h;\pi) = \exp(-T(h;\pi))$ as in the preceding section. Then the formula for $\frac{\partial \psi}{\partial \theta_1}$ is given by the following proposition.

Proposition 40. *Suppose that $\theta_2 < \overline{\theta}$. Then*

$$\frac{\partial\psi}{\partial\theta_1}(\theta_1,\pi) = \left(\frac{\beta\, b\left(\frac{\theta_1}{\beta}\right)}{\theta_1\left(\theta_2 + \beta\, b\left(\frac{\theta_1}{\beta}\right)\right)}\left(G(\theta_2) - \phi(\theta_1,\pi)\right) + \frac{1}{\pi\,\theta_1}\left(G(\theta_2) - G(\theta_1)\right)\right) S(\theta_2,\overline{\theta}\,).$$

Proof. This is simply a restatement of Lemma 20. ■

In view of Proposition 40, there are two main contributions to $\frac{\partial\psi}{\partial\theta_1}$, namely

1. $G(\theta_2) - \phi(\theta_1,\pi)$;
2. $G(\theta_2) - G(\theta_1)$.

We have already given conditions under which the first is strictly positive (in Proposition 34). The second is strictly positive under the same conditions:

Proposition 41. *Suppose that:*

1. $\theta_2 \in \left(\underline{\theta}\,,\overline{\theta}\,\right)$;
2. $\psi(\theta_1) \le F(\overline{\theta}\,)$;
3. *Assumption A4 holds.*

Then $G(\theta_2) - G(\theta_1) > 0$.

Proof. We break the proof down into the cases $\theta_1 \in \left(\frac{1}{1+\pi}\,\underline{\theta}\,, \frac{1}{1+\pi}\,\overline{\theta}_F\right)$ and $\theta_1 \in \left[\frac{1}{1+\pi}\,\overline{\theta}_F, \frac{1}{1+\pi}\,\overline{\theta}\,\right)$. In the first case, the proof parallels that of Lemma 22. We have $[\theta_1,\theta_2] \subset \left(\frac{1}{1+\pi}\,\overline{\theta}_F,\overline{\theta}_F\right) \subset (0,\theta_M)$, and moreover $\theta_2 > \underline{\theta}$. Assumption A4 therefore implies that $G(\theta_2) > G(\theta_1)$. In the second case, Lemma 24 implies directly that that $G(\theta_2) > G(\theta_1)$. ■

M. The Strictly Slack Case

In this section we focus on the set W of (θ_1, π) such that

1. $\theta_1 \in \left(0, \overline{\theta}\,\right)$,
2. $\pi \in (0, \infty)$ and
3. $\theta_2 = (1 + \pi)\,\theta_1 > \overline{\theta}$.

In other words, we do not impose the requirement that $\theta_1 = \tau_1(\pi)$ (i.e. that θ_1 be optimal for the given π), but we do require that the maximum-penalty constraint is slack in the sense that no types are choosing to pay the penalty.

M.1. The formula for $\frac{\partial\psi}{\partial\pi}$. The formula for $\frac{\partial\psi}{\partial\pi}$ is given by the following proposition.

Proposition 42. *Suppose that $\theta_2 > \overline{\theta}$. Then*

$$\frac{\partial\psi}{\partial\pi}(\theta_1, \pi) = 0.$$

Proof. As in the proof of Lemma 21, we have $\psi(\theta_1, \pi) = \chi(\theta_1)$ for $\theta_1 \in \left(\frac{1}{1+\pi}\,\overline{\theta}\,, \overline{\theta}\,\right)$,where

$$\chi(\theta_1) = \frac{1}{\overline{\theta} - \theta_1} \int_{\theta_1}^{\overline{\theta}} G(\theta)\, d\theta.$$

Hence ψ is independent of π for such θ_1. ■

M.2. The formula for $\frac{\partial\psi}{\partial\theta_1}$. The formula for $\frac{\partial\psi}{\partial\theta_1}$ is given by the following proposition.

Proposition 43. *Suppose that $\theta_2 > \overline{\theta}$. Then*

$$\frac{\partial\psi}{\partial\theta_1}(\theta_1, \pi) = \frac{\chi(\theta_1) - G(\theta_1)}{\overline{\theta} - \theta_1}.$$

Proof. As already noted, we have $\psi(\theta_1, \pi) = \chi(\theta_1)$ for $\theta_1 \in \left(\frac{1}{1+\pi}\,\overline{\theta}\,, \overline{\theta}\,\right)$. Moreover

$$\chi'(\theta_1) = \frac{\chi(\theta_1) - G(\theta_1)}{\overline{\theta} - \theta_1},$$

as in the proof of Lemma 21. ■

In view of Proposition 43, there is really only one contribution to $\frac{\partial \psi}{\partial \theta_1}$, namely $\chi(\theta_1) - G(\theta_1)$. It is not possible to sign $\chi(\theta_1) - G(\theta_1)$ for all θ_1, but it is possible to sign it when $\theta_1 = \tau_1(\pi)$, and indeed much more generally when $\theta_1 \in \left(0, \overline{\theta}_F\right)$. As before, this is enough for the purpose of our comparative statics.

Proposition 44. *Suppose that* $\theta_1 \in \left(0, \overline{\theta}_F\right)$. *Then* $\chi(\theta_1) - G(\theta_1) > 0$.

Proof. This is simply a special case of Lemma 27. ■

N. The Intermediate Case

Up to now we have focussed on the comparative statics of τ_1. For example, we have shown that if A4 is satisfied and $\rho = 1$ then: (i) $\tau_1' < 0$ when $\tau_2(\pi) < \overline{\theta}$; and (ii) $\tau_1' = 0$ when $\tau_2(\pi) > \overline{\theta}$. However, this leaves open the question of what happens at the transition between the two cases. For example, does τ_1 jump up when $\tau_2(\pi) = \overline{\theta}\,$? Does it jump down? Or is there more than one value of π for which $\tau_2(\pi) = \overline{\theta}\,$?

In order to address these questions, we need to understand the comparative statics of τ_2. These comparative statics are quite complex in the binding case. However, they simplify as the borderline between the two cases is approached. Moreover they are simpler still in the slack case.

N.1. Comparative Statics of τ_2 in the Weakly Binding Case.

We begin this section by looking at the comparative statics of τ_2 when the maximum-penalty constraint is strictly binding (in the sense that $\tau_2(\pi) < \overline{\theta}$). More precisely, we show that $\tau_2'(\pi)$ satisfies a simple linear equation. We then go on to check whether this equation remains valid when the maximum-penalty constraint is only weakly binding (in the sense that $\tau_2(\pi) \uparrow \overline{\theta}\,$).

Proposition 45. *Suppose that* $\tau_2(\pi) < \overline{\theta}$. *Then*

$$D(\tau_1(\pi), \pi)\, \tau_2'(\pi) = N(\tau_1(\pi), \pi), \tag{31}$$

where

$$\begin{aligned} D(\theta_1,\pi) &= (1+\pi)\frac{\theta_1+\beta\, b\left(\frac{\theta_1}{\beta}\right)}{\theta_2+\beta\, b\left(\frac{\theta_1}{\beta}\right)}\left(G(\theta_2)-\phi(\theta_1,\pi)\right)+\left(\phi(\theta_1,\pi)-G(\theta_1)\right), \\ N(\theta_1,\pi) &= \left(\left(\phi(\theta_1,\pi)-G(\theta_1)\right)++\frac{\pi\,(1+\pi)}{S\,(\theta_2,\pi)}\int_{\left[\theta_2,\overline{\theta}\,\right]}\tfrac{\partial S}{\partial \pi}(h;\pi)\, d\widetilde{G}(h)\right)\theta_1 \end{aligned}$$

and $\widetilde{G}\in\mathcal{BV}(\left[\theta_2,\overline{\theta}\,\right],\mathbb{R})$ *is given by the formulae* $\widetilde{G}_L(\theta_2)=\phi(\theta_1,\pi)$ *and* $\widetilde{G}=G$ *on* $\left(\theta_2,\overline{\theta}\,\right]$.

Proof. We have

$$\tau_2(\pi)=(1+\pi)\,\tau_1(\pi)$$

and therefore

$$\tau_2'(\pi)=\tau_1(\pi)+(1+\pi)\,\tau_1'(\pi). \tag{32}$$

Now,

$$\frac{\partial\psi}{\partial\theta_1}(\tau_1(\pi),\pi)\,\tau_1'(\pi)+\frac{\partial\psi}{\partial\pi}(\tau_1(\pi),\pi)=0.$$

Hence, multiplying (32) through by $\frac{\partial\psi}{\partial\theta_1}(\tau_1(\pi),\pi)$, we obtain

$$\begin{aligned} \frac{\partial\psi}{\partial\theta_1}\,\tau_2' &= \frac{\partial\psi}{\partial\theta_1}\,\tau_1+(1+\pi)\,\frac{\partial\psi}{\partial\theta_1}\,\tau_1' \\ &= \frac{\partial\psi}{\partial\theta_1}\,\tau_1-(1+\pi)\,\frac{\partial\psi}{\partial\pi}, \end{aligned}$$

where we have suppressed the dependence of $\frac{\partial\psi}{\partial\theta_1}$ and $\frac{\partial\psi}{\partial\pi}$ on $\tau_1(\pi)$ and π, and the dependence of τ_1 and τ_2 on π. We may therefore put

$$D(\theta_1,\pi)=\frac{\pi\,\theta_1}{S(\theta_2,\overline{\theta}\,)}\,\frac{\partial\psi}{\partial\theta_1}(\theta_1,\pi)$$

and

$$N(\theta_1,\pi)=\frac{\pi\,\theta_1}{S(\theta_2,\overline{\theta}\,)}\left(\frac{\partial\psi}{\partial\theta_1}(\theta_1,\pi)\,\theta_1-(1+\pi)\,\frac{\partial\psi}{\partial\pi}(\theta_1,\pi)\right).$$

Equation (31) now follows on applying the formulae for $\frac{\partial\psi}{\partial\pi}(\theta_1,\pi)$ and $\frac{\partial\psi}{\partial\theta_1}(\theta_1,\pi)$ given in Propositions 32 and 40. ■

Equation (31) can be solved for $\tau_2'(\pi)$ under the conditions of Proposition 34, namely that: (i) $\tau_2(\pi)\in\left(\underline{\theta}\,,\overline{\theta}\,\right)$; (ii) $\psi(\tau_1(\pi))\le F(\,\overline{\theta}\,)$; and (iii) Assumption A4 holds. This is not, however, enough for our current purposes: we need to make sure that it can still be solved for $\tau_2'(\pi)$ in the limiting case $\tau_2(\pi)\uparrow\overline{\theta}$. To this end, recall that

$$V=\left\{(\theta_1,\pi)\mid\theta_1\in\left(0,\overline{\theta}\,\right),\ \pi\in(0,\infty)\,,\ \theta_2<\overline{\theta}\right\},$$

and put

$$\partial V=\left\{(\theta_1,\pi)\mid\theta_1\in\left(0,\overline{\theta}\,\right),\ \pi\in(0,\infty)\,,\ \theta_2=\overline{\theta}\right\}.$$

Furthermore, for all $(\theta_1,\pi)\in V\cup\partial V$, put

$$\begin{aligned}\underline{D}(\theta_1,\pi)\quad&=\quad(1+\pi)\,\frac{\theta_1+\beta\,b\left(\frac{\theta_1}{\beta}\right)}{\theta_2+\beta\,b\left(\frac{\theta_1}{\beta}\right)}\left(G_L(\theta_2)-\phi(\theta_1,\pi)\right)\\&\qquad+\left(\phi(\theta_1,\pi)-\max\left\{G(\theta_1),G_L(\theta_1)\right\}\right)\end{aligned}$$

and

$$\begin{aligned}\overline{D}(\theta_1,\pi)\quad&=\quad(1+\pi)\,\frac{\theta_1+\beta\,b\left(\frac{\theta_1}{\beta}\right)}{\theta_2+\beta\,b\left(\frac{\theta_1}{\beta}\right)}\left(G_L(\theta_2)-\phi(\theta_1,\pi)\right)\\&\qquad+\left(\phi(\theta_1,\pi)-\min\left\{G(\theta_1),G_L(\theta_1)\right\}\right).\end{aligned}$$

Then we have:

Lemma 46. *Suppose that $(\widetilde{\theta}_1,\widetilde{\pi})\in V\to(\theta_1,\pi)\in\partial V$. Then*

$$\underline{D}(\theta_1,\pi)\le\liminf D(\,\widetilde{\theta}_1,\widetilde{\pi}\,)\le\limsup D(\,\widetilde{\theta}_1,\widetilde{\pi}\,)\le\overline{D}(\theta_1,\pi).$$

Proof. Note first that b and ϕ are both continuous. Hence $b\left(\frac{\widetilde{\theta}_1}{\beta}\right)\to b\left(\frac{\theta_1}{\beta}\right)$ and $\phi(\,\widetilde{\theta}_1,\widetilde{\pi})\to\phi(\theta_1,\pi)$. Next, put $\widetilde{\theta}_2=(1+\pi)\,\widetilde{\theta}_1$ and $\theta_2=(1+\pi)\,\theta_1$. Then $\widetilde{\theta}_2\uparrow\theta_2$, and therefore $G(\widetilde{\theta}_2) \to G_L(\theta_2)$. Finally,

$$\begin{aligned} \min\{G(\theta_1), G_L(\theta_1)\} &\leq \liminf G(\widetilde{\theta}_1) \\ &\leq \limsup G(\widetilde{\theta}_1) \\ &\leq \max\{G(\theta_1), G_L(\theta_1)\}. \end{aligned}$$

The result follows. ■

The next step is to sign $\underline{D}$. This cannot be done everywhere on $V \cup \partial V$. But it can be done when $\theta_2 = \overline{\theta}$ and $\theta_1 = \tau_1(\pi)$. Indeed, it is enough to require that $\theta_2 \in (\overline{\theta}_F, \overline{\theta}]$ (i.e. we do not actually have to be on the boundary) and that $\psi(\theta_1, \pi) \leq F(\overline{\theta})$ (i.e. we do not actually have to be at an optimum). We begin with a lemma.

Lemma 47. *Suppose that:*

1. $\theta_2 \in (\overline{\theta}_F, \overline{\theta}]$;
2. $\psi(\theta_1) \leq F(\overline{\theta})$.

Then $G_L(\theta_2) > G(\theta_1) \geq G_L(\theta_1)$.

Proof. The proof is similar to that of Lemma 24. Note first that

$$G_L(\theta_2) \geq F(\overline{\theta})$$

(with strict inequality if $\theta_2 < \overline{\theta}$)

$$\geq \psi(\theta_1) = \Psi(\overline{\theta}; \theta_1) > \Psi(\theta_1; \theta_1) = G(\theta_1)$$

(by assumption, by definition of ψ, because $\theta_1 < \theta_2 \leq \overline{\theta}$ and by construction of Ψ respectively). Second, Lemma 10 tells us that $\psi > F(\overline{\theta})$ on $[\overline{\theta}_F, \overline{\theta})$. But we have $\psi(\theta_1) \leq F(\overline{\theta})$. Hence $\theta_1 < \overline{\theta}_F$ and therefore $G' \geq 0$ at $\theta_1 \in (0, \overline{\theta}_F) \subset (0, \theta_M)$. That is, $G(\theta_1) - G_L(\theta_1) = \Delta G(\theta_1) \geq 0$. ■

We can now sign $\underline{D}$.

Proposition 48. *Suppose that:*

1. $\theta_2 \in \left(\overline{\theta}_F, \overline{\theta}\,\right]$;
2. $\psi(\theta_1) \le F(\,\overline{\theta}\,)$.

Then $\underline{D}(\theta_1, \pi) > 0$.

Proof. Two things follow from Lemma 47. First, $G(\theta_1) \ge G_L(\theta_1)$. Hence the formula for $\underline{D}(\theta_1, \pi)$ simplifies to

$$\underline{D}(\theta_1, \pi) = (1+\pi)\,\frac{\theta_1 + \beta\, b\left(\frac{\theta_1}{\beta}\right)}{\overline{\theta} + \beta\, b\left(\frac{\theta_1}{\beta}\right)}\,(G_L(\theta_2) - \phi(\theta_1, \pi)) + (\phi(\theta_1, \pi) - G(\theta_1))\,.$$

In particular, $\underline{D}(\theta_1, \pi)$ is a strictly positive linear combination of the two terms $G_L(\theta_2) - \phi(\theta_1, \pi)$ and $\phi(\theta_1, \pi) - G(\theta_1)$. Second, $G_L(\theta_2) - G(\theta_1) > 0$. Hence the sum of the two terms $G_L(\theta_2) - \phi(\theta_1, \pi)$ and $\phi(\theta_1, \pi) - G(\theta_1)$ is strictly positive. It therefore suffices to show that each of these two terms is non-negative. We have

$$G_L(\theta_2) \ge F(\,\overline{\theta}\,) \ge \psi(\theta_1) = \Psi(\,\overline{\theta}\,; \theta_1)$$

(as in the proof of Lemma 47)

$$\ge \Psi(\theta_2; \theta_1) > \Psi(\theta_1; \theta_1)$$

(since $\Psi' \ge 0$ on $\left(\theta_1, \overline{\theta}_F\right)$ (by Proposition 18) and $\Psi' > 0$ on $\left[\overline{\theta}_F, \overline{\theta}\,\right)$ (by Proposition 19))

$$= G(\theta_1)$$

(again as in the proof of Lemma 47). In particular, since $\Psi(\theta_2; \theta_1) = \phi(\theta_1, \pi)$, we have $G_L(\theta_2) \ge \phi(\theta_1, \pi)$ and $\phi(\theta_1, \pi) > G(\theta_1)$. ■

Since $\underline{D} > 0$, finding the sign of N and finding the sign of $\tau_2'(\pi)$ amount to the same thing. Note first that

$$\tau_2(\pi) = (1+\pi)\,\tau_1(\pi)$$

and hence

$$\tau_2'(\pi) = (1+\pi)\,\tau_1'(\pi) + \tau_1(\pi).$$

We therefore face a tension. On the one hand, we are mainly interested in the case in which $\tau_1'(\pi) < 0$. For our purposes, then, the first contribution to $\tau_2'(\pi)$ (namely $(1+\pi)\,\tau_1'(\pi)$) is negative. However, the second contribution (namely $\tau_1(\pi)$) is necessarily positive. The net effect is therefore ambiguous. Worse still, what we really need to show for the purposes of comparative statics is that $\tau_2'(\pi) > 0$ (so that the curve $(\tau_1(\pi), \pi)$ crosses the boundary $\theta_2 = \overline{\theta}$ in a simple way). This is directly at odds with our interest in the case in which $\tau_1'(\pi) < 0$.

Fortunately, the problem of signing $\tau_2'(\pi)$ at the boundary is much simpler than the problem of signing $\tau_2'(\pi)$ in V. With this end in mind, for all $(\theta_1, \pi) \in \partial V$, put

$$\underline{N}(\theta_1, \pi) = (\phi(\theta_1, \pi) - \max\{G(\theta_1), G_L(\theta_1)\})\,\theta_1$$

and

$$\overline{N}(\theta_1, \pi) = (\phi(\theta_1, \pi) - \min\{G(\theta_1), G_L(\theta_1)\})\,\theta_1.$$

Then we have the following lemma.

Lemma 49. *Suppose that* $(\widetilde{\theta}_1, \widetilde{\pi}) \in V \to (\theta_1, \pi) \in \partial V$. *Then*

$$\underline{N}(\theta_1, \pi) \leq \liminf N(\widetilde{\theta}_1, \widetilde{\pi}) \leq \limsup N(\widetilde{\theta}_1, \widetilde{\pi}) \leq \overline{N}(\theta_1, \pi).$$

Proof. The proof is similar to that of Lemma 46. Put $\widetilde{\theta}_2 = (1+\pi)\,\widetilde{\theta}_1$ and $\theta_2 = (1+\pi)\,\theta_1$. Then $\widetilde{\theta}_2 \uparrow \theta_2 = \overline{\theta}$, and therefore $\frac{\partial S}{\partial \pi}(\widetilde{\theta}_2, \pi) \to 0$ and $\int_{(\theta_2, \overline{\theta}]} \frac{\partial S}{\partial \pi}(h; \pi)\, dG(h) \to 0$. Furthermore $\phi(\widetilde{\theta}_1, \widetilde{\pi}) \to \phi(\theta_1, \pi)$ and

$$\begin{aligned} \min\{G(\theta_1), G_L(\theta_1)\} &\leq \liminf G(\widetilde{\theta}_1) \\ &\leq \limsup G(\widetilde{\theta}_1) \\ &\leq \max\{G(\theta_1), G_L(\theta_1)\}. \end{aligned}$$

Passing to the limit in the formula given for N in the statement of Proposition 45, we therefore obtain the required result. ■

Combining Lemma 49 with the earlier Lemma 47, we obtain:

Proposition 50. *Suppose that:*

1. $\theta_2 = \overline{\theta}$*;*

2. $\psi(\theta_1) \leq F(\overline{\theta}\,)$*.*

Then $\underline{N}(\theta_1, \pi) > 0$.

Proof. The proof is similar to that of Proposition 48. First, because $\psi(\theta_1) \leq F(\overline{\theta}\,)$ and therefore $\theta_1 < \overline{\theta}_F$, the formula for $\underline{N}(\theta_1, \pi)$ simplifies to

$$\underline{N}(\theta_1, \pi) = (\phi(\theta_1, \pi) - G(\theta_1))\,\theta_1.$$

Second, we have

$$\Psi(\overline{\theta}\,; \theta_1) > \Psi(\theta_1; \theta_1) = G(\theta_1).$$

It remains only to note that, because $\theta_2 = \overline{\theta}$, we have $\phi(\theta_1, \pi) = \Psi(\overline{\theta}\,; \theta_1)$. ■

Combining Propositions 48 and 50, we see that $\tau_2'(\pi) > 0$ on ∂V. In other words, whatever the behaviour of the curve $(\tau_1(\pi), \pi)$ in V, it points out of V at ∂V. I.e. it can exit, but not enter, V at ∂V. In particular, there exists $\pi_1 \in (0, \infty)$ such that $\tau_2(\pi) < \overline{\theta}$ iff $\pi \in [0, \pi_1)$.

N.2. Comparative Statics of τ_2 in the Weakly Slack Case. We begin this section by looking at the comparative statics of τ_2 when the maximum-penalty constraint is strictly slack (in the sense that $\tau_2(\pi) > \overline{\theta}$). More precisely, we show that $\tau_2'(\pi)$ satisfies a simple linear equation. We then go on to check whether this equation remains valid when the maximum-penalty constraint is only weakly slack (in the sense that $\tau_2(\pi) \downarrow \overline{\theta}\,$). Our first proposition is analogous to Proposition 45.

Proposition 51. *Suppose that* $\tau_2(\pi) > \overline{\theta}$*. Then*

$$D(\tau_1(\pi), \pi)\,\tau_2'(\pi) = N(\tau_1(\pi), \pi), \tag{33}$$

where

$$D(\theta_1, \pi) = \frac{\chi(\theta_1) - G(\theta_1)}{\overline{\theta} - \theta_1}$$

and

$$N(\theta_1, \pi) = \frac{\chi(\theta_1) - G(\theta_1)}{\overline{\theta} - \theta_1}\,\theta_1.$$

Notice that, if $\chi(\theta_1) - G(\theta_1) > 0$, then we can divide through by $D(\tau_1(\pi), \pi)$ to conclude that $\tau_2'(\pi) = \theta_1$. Furthermore $\chi(\theta_1) - G(\theta_1) > 0$ if $\theta_1 = \tau_1(\pi)$, and indeed much more generally if $\theta_1 \in \left(0, \overline{\theta}_F\right)$. Cf. Proposition 44. But it does not hold for all $(\theta_1, \pi) \in W$.

Proof. As in the proof of Proposition 45, we have

$$\frac{\partial \psi}{\partial \theta_1} \tau_2' = \frac{\partial \psi}{\partial \theta_1} \tau_1 - (1 + \pi) \frac{\partial \psi}{\partial \pi}.$$

We may therefore put

$$D(\theta_1, \pi) = \frac{\partial \psi}{\partial \theta_1}(\theta_1, \pi)$$

and

$$N(\theta_1, \pi) = \frac{\partial \psi}{\partial \theta_1}(\theta_1, \pi)\, \theta_1 - (1 + \pi) \frac{\partial \psi}{\partial \pi}(\theta_1, \pi).$$

Equation (33) now follows on applying the formulae for $\frac{\partial \psi}{\partial \pi}(\theta_1, \pi)$ and $\frac{\partial \psi}{\partial \theta_1}(\theta_1, \pi)$ given in Propositions 42 and 43. ■

The next step is to ensure that equation (33) can still be solved for $\tau_2'(\pi)$ in the limiting case $\tau_2(\pi) \downarrow \overline{\theta}$. To this end, recall that

$$W = \left\{ (\theta_1, \pi) \mid \theta_1 \in \left(0, \overline{\theta}\right),\ \pi \in (0, \infty),\ \theta_2 > \overline{\theta} \right\},$$

and put

$$\partial W = \left\{ (\theta_1, \pi) \mid \theta_1 \in \left(0, \overline{\theta}\right),\ \pi \in (0, \infty),\ \theta_2 = \overline{\theta} \right\}.$$

Furthermore, for all $(\theta_1, \pi) \in W \cup \partial W$, put

$$\begin{aligned} \underline{D}(\theta_1, \pi) &= \frac{\chi(\theta_1) - \max\{G(\theta_1), G_L(\theta_1)\}}{\overline{\theta} - \theta_1}, \\ \overline{D}(\theta_1, \pi) &= \frac{\chi(\theta_1) - \min\{G(\theta_1), G_L(\theta_1)\}}{\overline{\theta} - \theta_1} \end{aligned}$$

and

$$\begin{aligned}\underline{N}(\theta_1,\pi) &= \frac{\chi(\theta_1)-\max\{G(\theta_1),G_L(\theta_1)\}}{\overline{\theta}-\theta_1}\theta_1,\\ \overline{N}(\theta_1,\pi) &= \frac{\chi(\theta_1)-\min\{G(\theta_1),G_L(\theta_1)\}}{\overline{\theta}-\theta_1}\theta_1.\end{aligned}$$

Then we have:

Lemma 52. *Suppose that* $(\widetilde{\theta}_1,\widetilde{\pi})\in W\to(\theta_1,\pi)\in\partial W$. *Then*

$$\underline{D}(\theta_1,\pi)\leq\liminf D(\widetilde{\theta}_1,\widetilde{\pi})\leq\limsup D(\widetilde{\theta}_1,\widetilde{\pi})\leq\overline{D}(\theta_1,\pi)$$

and

$$\underline{N}(\theta_1,\pi)\leq\liminf N(\widetilde{\theta}_1,\widetilde{\pi})\leq\limsup N(\widetilde{\theta}_1,\widetilde{\pi})\leq\overline{N}(\theta_1,\pi).$$

Proof. Note first that χ is continuous. Hence $\chi(\widetilde{\theta}_1)\to\chi(\theta_1)$. On the other hand, as in the proof of Lemma 46,

$$\begin{aligned}\min\{G(\theta_1),G_L(\theta_1)\} &\leq \liminf G(\widetilde{\theta}_1)\\ &\leq \limsup G(\widetilde{\theta}_1)\\ &\leq \max\{G(\theta_1),G_L(\theta_1)\}.\end{aligned}$$

The result follows. ■

The next step is to sign $\underline{D}$. This cannot be done everywhere on $W\cup\partial W$. But it can be done when $\theta_2=\overline{\theta}$ and $\theta_1=\tau_1(\pi)$. Indeed, it is enough to require that $\theta_2\in[\overline{\theta},\infty)$ (i.e. we do not actually have to be on the boundary) and that $\psi(\theta_1,\pi)\leq F(\overline{\theta})$ (i.e. we do not actually have to be at an optimum). We begin with a lemma.

Lemma 53. *Suppose that:*

1. $\theta_2\in[\overline{\theta},\infty)$;
2. $\psi(\theta_1)\leq F(\overline{\theta})$.

Then $G(\theta_1)\geq G_L(\theta_1)$.

Proof. The proof is identical to the relevant part of that of Lemma 47. Since $\psi(\theta_1) \leq F(\overline{\theta})$, we must have $\theta_1 < \overline{\theta}_F$. Hence $G' \geq 0$ at $\theta_1 \in (0, \overline{\theta}_F) \subset (0, \theta_M)$. ■

Proposition 54. *Suppose that:*

1. $\theta_2 \in [\overline{\theta}, \infty)$;
2. $\psi(\theta_1) \leq F(\overline{\theta})$.

   Then $\underline{D}(\theta_1, \pi), \underline{N}(\theta_1, \pi) > 0$.

Proof. Note first that, in view of Lemma 53, we have

$$\underline{D}(\theta_1, \pi) = \frac{\chi(\theta_1) - G(\theta_1)}{\overline{\theta} - \theta_1}$$

and

$$\underline{N}(\theta_1, \pi) = \frac{\chi(\theta_1) - G(\theta_1)}{\overline{\theta} - \theta_1}\theta_1.$$

Second, since $\psi(\theta_1) \leq F(\overline{\theta})$, we have $\theta_1 < \overline{\theta}_F$. Finally, Proposition 44 tells us that $\chi(\theta_1) - G(\theta_1) > 0$ for $\theta_1 \in (0, \overline{\theta}_F)$. ■

It follows from Proposition 54 that $\tau_2'(\pi) > 0$ on ∂W. In other words, whatever the behaviour of the curve $(\tau_1(\pi), \pi)$ in W, it points into W at ∂W. I.e. it can enter, but not exit, W at ∂W. In particular, there exists $\pi_2 \in (0, \infty)$ such that $\tau_2(\pi) > \overline{\theta}$ iff $\pi \in (\pi_2, \infty)$.

N.3. Comparative Statics of τ_2 in the Remaining Case. At this point we have established that there exist $0 < \pi_1 \leq \pi_2 < \infty$ such that $\tau_2(\pi) < \overline{\theta}$ iff $\pi \in [0, \pi_1)$ and $\tau_2(\pi) > \overline{\theta}$ iff $\pi \in (\pi_2, \infty)$.[14] The remaining question is therefore whether it is possible that $\pi_1 < \pi_2$, in other words that there is a non-trivial interval (π_1, π_2) over which $\tau_2(\pi) = \overline{\theta}$.

Suppose for a contradiction that there is such an interval. Then, over this interval, we must have both

$$\psi(\tau_1(\pi), \pi) = F(\overline{\theta}) \tag{34}$$

[14] That $\pi_1 \leq \pi_2$ follows at once from the fact that we cannot have $\tau_2(\pi) < \overline{\theta}$ and $\tau_2(\pi) > \overline{\theta}$ simultaneously.

(because $\tau_1(\pi)$ is the optimal θ_1) and

$$\tau_2(\pi) = \overline{\theta}. \tag{35}$$

Hence

$$F(\,\overline{\theta}\,) = \psi(\tau_1(\pi), \pi) = \Psi(\,\overline{\theta}\,; \tau_1(\pi), \pi) = \Psi(\tau_2(\pi); \tau_1(\pi), \pi)$$

(by equation (34), by definition of ψ and by equation (35))

$$= \phi(\tau_1(\pi); \pi) = \frac{1}{\tau_2(\pi) - \tau_1(\pi)} \int_{\tau_1(\pi)}^{\tau_2(\pi)} G(\theta)\, d\theta = \frac{1}{\overline{\theta} - \tau_1(\pi)} \int_{\tau_1(\pi)}^{\overline{\theta}} G(\theta)\, d\theta$$

(by construction of Ψ, by definition of ϕ, by equation (35) again). Multiplying through by $\overline{\theta} - \tau_1(\pi)$, we therefore obtain

$$\int_{\tau_1(\pi)}^{\overline{\theta}} G(\theta)\, d\theta = (\overline{\theta} - \tau_1(\pi))\, F(\overline{\theta}).$$

Differentiating with respect to π, we then obtain

$$-G(\tau_1(\pi))\, \tau_1'(\pi) = -\tau_1'(\pi)\, F(\overline{\theta})$$

or

$$(G(\tau_1(\pi)) - F(\overline{\theta}))\, \tau_1'(\pi) = 0.$$

But equation (35) implies that $(1 + \pi)\, \tau_1(\pi) = \overline{\theta}$ and therefore

$$\tau_1'(\pi) = -\frac{\theta_1}{1 + \pi} \neq 0.$$

We conclude that $G(\tau_1(\pi)) - F(\overline{\theta}) = 0$. This, however, is impossible. For we have

$$G(\tau_1(\pi)) = \Psi(\tau_1(\pi); \tau_1(\pi), \pi) \leq \Psi(\,\overline{\theta}_F\,; \tau_1(\pi), \pi) < \Psi(\,\overline{\theta}\,; \tau_1(\pi), \pi)$$

(by construction of Ψ, by Proposition 18 and by Proposition 19)

$$= \psi(\tau_1(\pi), \pi) = F(\overline{\theta})$$

(as above). The only possible conclusion is therefore that $\pi_1 = \pi_2$.

O. Comparative Statics without A4

We divide our discussion into the same three cases that we considered in Section L.1, namely:

1. $\rho < 1$ and $G' \leq 0$ on $(\underline{\theta}, \infty)$;
2. $\rho = 1$;
3. $\rho > 1$ and $G' \geq 0$ on $\left(0, \overline{\theta}\right)$.

Of these, the first is by far the simplest.

Proposition 55. *Suppose that $\rho < 1$ and $G' \leq 0$ on $(\underline{\theta}, \infty)$. Then $\underline{\tau} = \overline{\tau}$ for all $\pi \in (0, \infty)$. Furthermore there exists $\pi_1 \in (0, \infty)$ such that: the maximum-penalty constraint is strictly binding for all $\pi \in (0, \pi_1)$; and the maximum-penalty constraint is strictly slack for all $\pi \in (\pi_1, \infty)$. Finally:*

1. *$\underline{\tau} = \overline{\tau}$ is strictly decreasing on $(0, \pi_1)$; and*
2. *$\underline{\tau} = \overline{\tau}$ is constant on (π_1, ∞).*

In other words, for all values of the maximum penalty $\pi \in [0, \infty)$, there is a unique optimum within our one-parameter family. Furthermore there exists a critical level π_1 of π. Below π_1, the maximum-penalty constraint is strictly binding and the optimal savings target is strictly increasing in π. Above π_1, the maximum-penalty constraint is strictly slack and the optimal savings target is independent of π.

Proof. Since $G' \geq 0$ on $(0, \underline{\theta})$ and $G' \leq 0$ on $(\underline{\theta}, \infty)$, we can put $\theta_M = \underline{\theta}$. For this choice of θ_M, A4 holds. Indeed: the interval $[\underline{\theta}, \theta_M)$ is empty, and therefore G is certainly strictly increasing on $[\underline{\theta}, \theta_M)$! We may therefore apply the analysis of Sections K-N to conclude that there is a unique $\theta_1 = \tau_1(\pi)$ such that $\psi(\theta_1, \pi) = F(\overline{\theta})$, and that $\tau_1'(\pi) < 0$. ■

Remark 56. *There is also a direct proof of Theorem 55. A sketch of this proof runs as follows. Since* $G' \leq 0$ *on* $(\underline{\theta}, \infty)$, *we must have* $\overline{\theta}_F = \underline{\theta}$. *Furthermore we always have* $\theta_1 < \overline{\theta}_F$ *and* $\theta_2 > \underline{\theta}$; *and in the strictly binding case we also have* $\theta_2 < \overline{\theta}$. *Hence, in the strictly binding case, we have*

$$G(\theta_2) > F(\overline{\theta}) = \psi(\theta_1, \pi) = \Psi(\overline{\theta}; \theta_1, \pi) > \Psi(\theta_2; \theta_1, \pi) = \phi(\theta_1, \pi) > G(\theta_1).$$

In particular,

$$G(\theta_2) - \phi(\theta_1, \pi) > 0.$$

It then follows from the formulae for $\frac{\partial \psi}{\partial \pi}$ *and* $\frac{\partial \psi}{\partial \theta_1}$ *given in Propositions 32 and 40 – both of which feature the term* $G(\theta_2) - \phi(\theta_1, \pi)$ *– that*

$$\frac{\partial \psi}{\partial \pi}, \frac{\partial \psi}{\partial \theta_1} > 0.$$

That is, there is a unique $\theta_1 = \tau_1(\pi)$ *such that* $\psi(\theta_1, \pi) = F(\overline{\theta})$, *and* $\tau_1'(\pi) < 0$. *(The important point here is the fact that our assumption on* G *allows us to sign the core term* $G(\theta_2) - \phi(\theta_1, \pi)$, *and thereby the derivatives* $\frac{\partial \psi}{\partial \pi}$ *and* $\frac{\partial \psi}{\partial \theta_1}$, *directly.)*

Proposition 57. *Suppose that* $\rho = 1$. *Then there exists* $\pi_0 \in [0, \infty)$ *such that:* $\underline{\tau} < \overline{\tau}$ *for all* $\pi \in (0, \pi_0)$; *and* $\underline{\tau} = \overline{\tau}$ *for all* $\pi \in (\pi_0, \infty)$. *Furthermore there exists* $\pi_1 \in (\pi_0, \infty)$ *such that: the maximum-penalty constraint is strictly binding for all* $\pi \in (0, \pi_1)$; *and the maximum-penalty constraint is strictly slack for all* $\pi \in (\pi_1, \infty)$. *Finally:*

1. $\underline{\tau}$ *is constant, and* $\overline{\tau}$ *is strictly decreasing, on* $(0, \pi_0)$;
2. $\underline{\tau} = \overline{\tau}$ *is strictly decreasing on* (π_0, π_1); *and*
3. $\underline{\tau} = \overline{\tau}$ *is constant on* (π_1, ∞).

In other words, there are two critical levels of π, namely π_0 and π_1. Below π_0, there is a continuum of optima from within our one-parameter family; and, above π_0, there is a unique optimum from within our one-parameter family. Below π_1, the maximum-penalty constraint is strictly binding; and, above π_1, the maximum-penalty constraint is strictly slack. Furthermore, below π_0: the smallest of the possible optimal savings targets is strictly increasing in π; and the largest of the possible optimal savings targets is independent of π. Between π_0 and π_1: there is only one optimal savings target, and this is strictly increasing in π. And, above π_1: there is again only one optimal savings target, and this is independent of π.

In order to prove Proposition 57, we shall need three Lemmas.

Lemma 58. *Fix $\pi > 0$, and suppose that there exist $\theta_3, \theta_4 \in \left(\frac{1}{1+\pi}\,\underline{\theta}\,, \overline{\theta}_F\right)$ such that $\theta_3 < \theta_4$ and $\left\{\theta_1 \mid \psi(\theta_1, \pi) = F(\,\overline{\theta}\,)\right\} = [\theta_3, \theta_4]$. Then:*

1. *$G < G(\theta_3)$ on $(\,0, \theta_3)$;*
2. *$G = G(\theta_3)$ on $[\,\theta_3, (1+\pi)\,\theta_4)$;*
3. *$G > G(\theta_3)$ on $\left((1+\pi)\,\theta_4, \overline{\theta}\,\right)$.*

Furthermore $G(\theta_3) < F(\,\overline{\theta}\,)$.

In other words, if there is a multiplicity of optimal savings targets, then G must have a flat. Moreover the domain of this flat consists precisely of the half-open interval $[\,\theta_3, (1+\pi)\,\theta_4)$, where θ_3 is the smallest possible choice of θ_1 and θ_4 is the highest possible choice of θ_1.

{

, i.e. the set of θ for which $G(\theta) = G(\theta_3)$,

union of the intervals $[\theta_1, (1+\pi)\,\theta_1)$ associated with the optimal choices of θ_1.

}

Proof. Since $\psi(\theta_1, \pi) = F(\,\overline{\theta}\,)$ for all $\theta_1 \in [\theta_3, \theta_4]$, we can take derivatives on the right to conclude that

$$\frac{\partial \psi}{\partial \theta_1}(\theta_1, \pi) = 0$$

for all $\theta_1 \in [\theta_3, \theta_4)$. But

$$\begin{aligned} \frac{\partial \psi}{\partial \theta_1}(\theta_1, \pi) &= \left(\frac{1+\pi}{\pi\,\theta_1} \, \frac{\theta_1 + \beta\, b\!\left(\frac{\theta_1}{\beta}\right)}{\theta_2 + \beta\, b\!\left(\frac{\theta_1}{\beta}\right)} \left(G(\theta_2) - \phi(\theta_1, \pi)\right) \right. \\ &\qquad \left. + \frac{1}{\pi\,\theta_1} \left(\phi(\theta_1, \pi) - G(\theta_1)\right) \right) S(\theta_2, \overline{\theta}\,). \end{aligned}$$

Hence

$$G(\theta_2) - \phi(\theta_1, \pi) = \phi(\theta_1, \pi) - G(\theta_1) = 0 \tag{36}$$

for all $\theta_1 \in [\theta_3, \theta_4)$, where $\theta_2 = (1+\pi)\,\theta_1$. But $G' \geq 0$ on $\left(0, \overline{\theta}_F\right]$. Equation (36) therefore imples that $G = G(\theta_3)$ on $[\,\theta_3, (1+\pi)\,\theta_4)$, and that $G_L((1+\pi)\,\theta_4) = G(\theta_3)$. This establishes part 2 of the Lemma.

Now put $\widetilde{\theta}_3$ ■

Lemma 59. *Fix $\pi > 0$, suppose that there exist $\theta_3, \theta_4 \in \left(\frac{1}{1+\pi}\,\underline{\theta}\,, \overline{\theta}_F\right)$ such that $\theta_3 < \theta_4$ and $\left\{\theta_1 \mid \psi(\theta_1, \pi) = F(\,\overline{\theta}\,)\right\} = [\theta_3, \theta_4]$, and put $\theta_5 = (1+\pi)\,\theta_4$ and $\pi_0 = \frac{\theta_5 - \theta_3}{\theta_3}$. Then, for all $\widehat{\pi} \in (0, \pi_0)$,*

$$\left\{\theta_1 \mid \psi(\theta_1, \widehat{\pi}) = F(\,\overline{\theta}\,)\right\} = \left[\theta_3, \tfrac{1}{1+\widehat{\pi}}\,\theta_5\right].$$

In other words, if there is some $\pi > 0$ for which there is a multiplicity of optimal savings targets, then there is a whole range of π for which there is a multiplicity of optimal savings targets. Furthermore both of these multiplicities are associated with the same flat of G.

Proof. ... ■

Lemma 60. *Fix $\widehat{\pi}, \pi > 0$ and suppose that:*

1. *$\widehat{\pi} < \pi$;*

2. *there exist $\widehat{\theta}_3, \widehat{\theta}_4 \in \left(\frac{1}{1+\pi}\,\underline{\theta}\,, \overline{\theta}_F\right)$ such that $\widehat{\theta}_3 < \widehat{\theta}_4$ and $\left\{\theta_1 \mid \psi(\theta_1, \widehat{\pi}) = F(\,\overline{\theta}\,)\right\} = [\,\widehat{\theta}_3, \widehat{\theta}_4\,]$;*

3. *there exist* $\theta_3, \theta_4 \in \left(\frac{1}{1+\pi}\,\underline{\theta}\,, \overline{\theta}_F\right)$ *such that* $\theta_3 < \theta_4$ *and* $\left\{\theta_1 \mid \psi(\theta_1, \pi) = F(\,\overline{\theta}\,)\right\} = [\,\theta_3, \theta_4\,]$.

Then: $\widehat{\theta}_3 = \theta_3$; $\widehat{\theta}_4 = \frac{1+\pi}{1+\widehat{\pi}}\,\theta_4$; *and* $\pi < \widehat{\pi}_0 = \frac{(1+\widehat{\pi})\widehat{\theta}_4 - \widehat{\theta}_3}{\widehat{\theta}_3}$.

In other words, if there is a multiplicity of optimal savings targets associated with both $\widehat{\pi}$ and π, then both multiplicities derive from the same flat of G.

Proof. ... ■

We are now in a position to prove Proposition 57.

Proof. [Proof of Proposition 57] ... ■

Proposition 61. *Suppose that* $\rho > 1$ *and* $G' \geq 0$ *on* $\left(0, \overline{\theta}\right)$. *Then there there exists* $\pi_1 \in (0, \infty)$ *such that: the maximum-penalty constraint is strictly binding for all* $\pi \in (\,0, \pi_1)$; *and the maximum-penalty constraint is strictly slack for all* $\pi \in (\,\pi_1, \infty\,)$. *Furthermore:*

1. $\underline{\tau}$ *and* $\overline{\tau}$ *are both strictly decreasing on* $(0, \pi_1)$;
2. $\underline{\tau} = \overline{\tau}$ *is constant on* (π_1, ∞).

In other words, there exists a critical level π_1 of π. Below π_1: the maximum-penalty constraint is strictly binding; and the set of optimal savings targets is strictly increasing in π. Above π_1: the maximum-penalty constraint is strictly slack; and the optimal savings target is independent of π.

Proof. Let L be the locus of all those $(\theta_1, \pi) \in V$ such that $\psi(\theta_1, \pi) = F(\,\overline{\theta}\,)$, let L_θ be the projection of L onto the first coordinate, and let L_π be the projection of L onto the second coordinate. Then, in order to prove the result, it suffices to show that there is a non-increasing function $\varpi : L_\theta \to L_\pi$, the graph of which is L. For then the inverse $\tau : L_\pi \to L_\theta$ of ϖ is a strictly increasing correspondence.

Note first that, for all $(\theta_1, \pi) \in V$, we have

$$\frac{\partial \psi}{\partial \theta_1}(\theta_1, \pi) = \left(\frac{1+\pi}{\pi\,\theta_1} \frac{\theta_1 + \beta\, b\left(\frac{\theta_1}{\beta}\right)}{\theta_2 + \beta\, b\left(\frac{\theta_1}{\beta}\right)} \left(G(\theta_2) - \phi(\theta_1, \pi)\right) + \frac{1}{\pi\,\theta_1}\left(\phi(\theta_1, \pi) - G(\theta_1)\right) \right) S(\theta_2, \overline{\theta}\,)$$

and

$$\frac{\partial \psi}{\partial \pi}(\theta_1, \pi) = \left(\frac{1}{\pi}\, S(\theta_2, \pi) - \theta_1\, \frac{\partial S}{\partial h}(\theta_2, \pi)\right)(G(\theta_2) - \phi(\theta_1, \pi)) - \int_{\left[\theta_2, \overline{\theta}\right]} \frac{\partial S}{\partial \pi}(h, \pi)\, d\widetilde{G}(h),$$

where, as above, $\widetilde{G} \in \mathcal{BV}(\left[\theta_2, \overline{\theta}\,\right], \mathbb{R})$ is given by the formulae $\widetilde{G}_L(\theta_2) = \phi(\theta_1, \pi)$ and $\widetilde{G} = G$ on $\left(\theta_2, \overline{\theta}\,\right]$.[15]

Next, since $G' \geq 0$ on $\left(0, \overline{\theta}\right)$, we must have $\phi(\theta_1, \pi) - G(\theta_1) \geq 0$, $G(\theta_2) - \phi(\theta_1, \pi) \geq 0$ and $\widetilde{G}' \geq 0$ on $\left[\theta_2, \overline{\theta}\,\right)$. Hence $\frac{\partial \psi}{\partial \theta_1} \geq 0$.

Third, if in addition if $\psi(\theta_1, \pi) = F(\overline{\theta}\,)$, then we must have $\frac{\partial \psi}{\partial \pi} > 0$. Indeed, it is always the case that $\overline{\theta}_F < \overline{\theta}$ and $G > F(\overline{\theta}\,)$ on $\left(\overline{\theta}_F, \overline{\theta}\,\right)$. (See Lemma 9.) Moreover, if $\psi(\theta_1, \pi) = F(\overline{\theta}\,)$, then we also have $G(\theta_1) < F(\overline{\theta}\,)$. Overall, then, if $\psi(\theta_1, \pi) = F(\overline{\theta}\,)$ then G is non-trivial on $\left(\theta_1, \overline{\theta}\,\right)$. Now suppose for a contradiction that $\frac{\partial \psi}{\partial \pi} = 0$. Then we must have $G(\theta_2) - \phi(\theta_1, \pi) = 0$ (which is the same thing as saying that $\widetilde{G}' = 0$ on $\{\theta_2\}$) and $\widetilde{G}' = 0$ on $\left(\theta_2, \overline{\theta}\,\right)$. Moreover the former implies that $G(\theta_2) = G_L(\theta_2) = \phi(\theta_1, \pi) = G(\theta_1)$, and the latter implies that $G_L(\overline{\theta}\,) = G(\theta_2)$. So G is trivial on $\left(\theta_1, \overline{\theta}\,\right)$, which is the required contradiction.

Finally, since $\frac{\partial \psi}{\partial \pi} > 0$, there is a unique $\pi = \varpi(\theta_1)$ such that $\psi(\theta_1, \pi) = F(\overline{\theta}\,)$ and moreover

$$\varpi'(\theta_1) = -\frac{\frac{\partial \psi}{\partial \theta_1}}{\frac{\partial \psi}{\partial \pi}} \geq 0.$$

This completes the proof. ■

Remark 62. *It is interesting to compare the levels of uniqueness obtained in Propositions 55, 57 and 61. When $\rho < 1$, we have uniqueness for all $\pi \in (0, \infty)$. When $\rho = 1$, a limited form of non-uniqueness can develop: there exists $\pi_0 \in [0, \pi_1)$ such that there is non-uniqueness on $(0, \pi_0)$ and uniqueness on (π_0, ∞). And, when $\rho > 1$, non-uniqueness takes the form that one might expect in a convex optimization problem. However, we do at least get strict monotonicity on the whole of $(0, \pi_1)$.*

[15] For the definition of V, see the beginning of Section L.

P. Existence of a Full Optimum

Suppose that self 0 is required to pick a B satisfying Constraints 1 and 2. Then the utility curve (u, w) that results will satisfy the following three conditions:

I (u, w) is interior, in the sense that $u, w > U(0)$ on Θ.

M (u, w) is monotonic, in the sense that u is non-decreasing and w is non-increasing.

DE (u, w) satisfies the differential equation $\theta\, du + \beta\, dw = 0$.

If B is also convex, then (u, w) will also satisfy:

C (u, w) is continuous.

Now, the set Ω with which we have worked so far consists of utility curves (u, w) that satisfy I, BV, DE and C, where BV is the condition:

BV (u, w) is of bounded variation.

Since BV is weaker than M, this means that Ω contains all the utility curves that can result from convex B, and more besides. We have therefore solved a relaxed version of the convex-B problem. Since the solution of this relaxed problem is feasible in the convex-B problem, we have therefore also solved the convex-B problem. The purpose of the present section is to solve the general problem in which B is not required to be convex.

Suppose accordingly that Ω consists of all $(u, w) \in \mathcal{BV}(\Theta, \operatorname{ran}(U))^2$ such that $\theta\, du + \beta\, dw = 0$. In other words, let Ω consist of utility curves (u, w) that satisfy I, BV, DE but not C. Put $X = \mathcal{BV}(\Theta, \mathbb{R})^2$, $\Xi = \mathcal{BV}(\Theta, \operatorname{ran}(U))^2$ and $Z = \mathcal{BV}(\Theta, \mathbb{R})$. Then the objective function M and the constraint mappings G_1 and G_2 continue to make sense. The analysis of Luenberger (1969) therefore shows that $x_0 \in \Omega$ solves the problem

$$\begin{array}{ll} \text{maximize} & M(x) \\ \text{subject to} & \left\{ \begin{array}{c} x \in \Omega \\ G_1(x) \geq 0 \\ G_2(x) \geq 0 \end{array} \right\} \end{array}$$

iff there exist $\lambda_1, \lambda_2 \in Z^*$ such that:

1. $L(x_0, \lambda_1, \lambda_2) \geq L(x, \lambda_1, \lambda_2)$ for all $x \in \Omega$, where

$$L(x, \lambda_1, \lambda_2) = M(x) + \langle G_1(x), \lambda_1 \rangle + \langle G_2(x), \lambda_2 \rangle\,;$$

2. $G_1(x) \geq 0$, $\lambda_1 \geq 0$ and $\langle G_1(x), \lambda_1 \rangle = 0$;

3. $G_2(x) \geq 0$, $\lambda_2 \geq 0$ and $\langle G_2(x), \lambda_2 \rangle = 0$.

In other words, there exists multipliers λ_1 and λ_2 such that: (1) x_0 maximizes $L(\,\cdot\,, \lambda_1, \lambda_2)$ over Ω; (2) complementary slackness holds for the first constraint; and (3) complementary slackness holds for the second constraint.

At this point, however, we encounter an obstacle. While the dual space $\mathcal{C}(\Theta, \mathbb{R})^*$ of $\mathcal{C}(\Theta, \mathbb{R})$ has a convenient representation as the space $\mathcal{M}(\Theta, \mathbb{R})$, the dual space $\mathcal{BV}(\Theta, \mathbb{R})^*$ of $\mathcal{BV}(\Theta, \mathbb{R})$ does not have a similarly convenient representation. This makes it difficult to use the necessity part of the Lagrangean characterization of the optimum. We can, however, still hope to use the sufficiency part.

The idea here is to note that the elements of $\mathcal{M}(\Theta, \mathbb{R})$ can be used to induce continuous linear functionals on $\mathcal{BV}(\Theta, \mathbb{R})$. For example, $\mu \in \mathcal{M}(\Theta, \mathbb{R})$ induces $\mu_R \in \mathcal{BV}(\Theta, \mathbb{R})^*$ via the formula

$$\langle z, \mu_R \rangle = \int z_R \, d\mu,$$

where z_R denotes the right-continuous version of z. However, in pursuing this idea, it is important to note that μ also induces $\mu_L \in \mathcal{BV}(\Theta, \mathbb{R})^*$ via the formula

$$\langle z, \mu_L \rangle = \int z_L \, d\mu,$$

where z_L denotes the left-continuous version of z. In other words, there is no canonical association between elements of $\mathcal{M}(\Theta, \mathbb{R})$ and continuous linear functionals on $\mathcal{BV}(\Theta, \mathbb{R})$.

Our plan is therefore to start from a θ_1 such that $\Psi\big(\,\overline{\theta}\,; \theta_1\big) = F(\,\overline{\theta}\,)$, in the hope that $\Psi(\,\cdot\,; \theta_1)$ can be used to generate multipliers that can be used in the sufficiency part of the Lagrangean characterization of an optimum. Indeed, suppose that we are given such a θ_1. Then, bearing in mind that $\Delta\Psi(\theta_2; \theta_1) = 0$, we may put $d\widetilde{\Lambda}_1 = d\Psi(\,\cdot\,; \theta_1)$ on $[\underline{\theta}, \theta_2]$, $d\Lambda_2 = d\Psi(\,\cdot\,; \theta_1)$ on $\big[\theta_2, \overline{\theta}\,\big]$ and $d\Lambda_1 = \frac{1}{K'(w_0)}\, d\widetilde{\Lambda}_1$. Furthermore, if we let λ_1 and λ_2 be the continuous linear functionals induced on $\mathcal{BV}(\Theta, \mathbb{R})$ by $d\Lambda_1$ and $d\Lambda_2$ using integration with respect to the right-continuous versions of functions, then we have

$$\begin{aligned} L(x, \lambda_1, \lambda_2) \;&=\; \int \Big(\theta\, u(\theta) + w(\theta)\Big)\, dF(\theta) \\ &\quad + \int \Big(y - C(u(\theta)) - K(w(\theta))\Big)\, d\Lambda_1(\theta) \\ &\quad + \int \left(b\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big)\, u(\theta) - \frac{1}{\rho}\, a\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big) - w(\theta)\right) d\Lambda_2(\theta) \end{aligned}$$

for all $x \in X$. Our objective is then to show that the utility curve $x_0 = (u_0, w_0)$ associated with θ_1 maximizes $L(\,\cdot\,, \lambda_1, \lambda_2)$.

It suffices to show that, for all $x_1 \in \Omega$, the directional derivative $\nabla_x L(x_0, \lambda_1, \lambda_2)$ of L at x_0 in the direction $x = x_1 - x_0$ is non-positive. As in Section F, we have

$$\begin{aligned} \nabla_x L(x_0, \lambda_1, \lambda_2) &= \int \Big(\theta\, u + w\Big)\, dF - \int \Big(C'(u_0)\, u + K'(w_0)\, w\Big)\, d\Lambda_1 \\ &\qquad + \int \Big(b\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big)\, u - w\Big)\, d\Lambda_2 \\ &= \int \Big(\theta\, u + w\Big)\, dF - \int \Big(\tfrac{C'(u_0)}{K'(w_0)}\, u + w\Big)\, d\widetilde{\Lambda}_1 \\ &\qquad + \int \Big(b\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big)\, u - w\Big)\, d\Lambda_2. \end{aligned}$$

Furthermore, notwithstanding the fact that we are now working in a more general context, we can eliminate the terms $\int w\, dF$, $\int w\, d\widetilde{\Lambda}_1$ and $\int w\, d\Lambda_2$ using integration by parts.

Indeed, the general formula for integration by parts tells us that

$$\int_{[\underline{\theta}, \overline{\theta}]} w(\theta)\, dF(\theta) = [w\, F]_{\underline{\theta}-}^{\overline{\theta}} - \int_{[\underline{\theta}, \overline{\theta}]} F(\theta)\, dw(\theta) + \sum_{\theta \in [\underline{\theta}, \overline{\theta}]} \Delta w(\theta)\, \Delta F(\theta),$$

where

$$[w\, F]_{\underline{\theta}-}^{\overline{\theta}} = w(\,\overline{\theta}\,)\, F(\,\overline{\theta}\,) - w(\,\underline{\theta}-)\, F(\,\underline{\theta}-).$$

We therefore have

$$\int w\,dF = [w\,F]_{\underline{\theta}-}^{\overline{\theta}} - \int F\,dw + \sum \Delta w\,\Delta F$$

(where we have suppressed the dependence on θ and where the domains of all intergrals and sums are understood to be the whole of $[\,\underline{\theta}\,,\overline{\theta}\,]$)

$$= w(\,\overline{\theta}\,)\,F(\,\overline{\theta}\,) + \int F\,\frac{\theta}{\beta}\,du - \sum \frac{\theta}{\beta}\,\Delta u\,\Delta F$$

(because $F(\,\underline{\theta}-) = 0$ and $dw = -\frac{\theta}{\beta}\,du$)

$$= w(\,\overline{\theta}\,)\,F(\,\overline{\theta}\,) + \frac{1}{\beta}\int F\,\theta\,du - \frac{1}{\beta}\sum \theta\,\Delta u\,\Delta F.$$

Moreover

$$\int F\,\theta\,du = [(F\,\theta)\,u]_{\underline{\theta}-}^{\overline{\theta}} - \int u\,d(F\,\theta) + \sum \Delta(F\,\theta)\,\Delta u$$

(applying the general formula for integration by parts to $\int F\,\theta\,du$)

$$= \overline{\theta}\,u(\,\overline{\theta}\,)\,F(\,\overline{\theta}\,) - \int u\,(\theta\,dF + F\,d\theta) + \sum \theta\,\Delta F\,\Delta u$$

(since $F(\,\underline{\theta}-) = 0$, $d(F\,\theta) = \theta\,dF + F\,d\theta$ and $\Delta(F\,\theta) = \theta\,\Delta F$). Overall, then,

$$\int w\,dF = \left(\frac{\overline{\theta}}{\beta}u(\,\overline{\theta}\,) + w(\,\overline{\theta}\,)\right)F(\,\overline{\theta}\,) - \frac{1}{\beta}\int u\,(\theta\,dF + F\,d\theta).$$

By the same token, and bearing in mind that we did not use the fact that $\Delta F = 0$ in the derivation of the previous paragraph, we have

$$\int w\,d\widetilde{\Lambda}_1 = \left(\frac{\overline{\theta}}{\beta}u(\,\overline{\theta}\,) + w(\,\overline{\theta}\,)\right)\widetilde{\Lambda}_1(\,\overline{\theta}\,) - \frac{1}{\beta}\int u\,(\theta\,d\widetilde{\Lambda}_1 + \widetilde{\Lambda}_1\,d\theta)$$

and

$$\int w\,d\Lambda_2 = \left(\frac{\overline{\theta}}{\beta}u(\,\overline{\theta}\,) + w(\,\overline{\theta}\,)\right)\Lambda_2(\,\overline{\theta}\,) - \frac{1}{\beta}\int u\,(\theta\,d\Lambda_2 + \Lambda_2\,d\theta).$$

We therefore have

$$\nabla_x L(x_0, \lambda_1, \lambda_2) = \int u\, du^* + w(\,\overline{\theta}\,)\, r^*$$

where, as in section F above,

$$\begin{aligned} du^* &= -\frac{1}{\beta}\Big((1-\beta)\,\theta\, dF + F\, d\theta\Big) + \frac{1}{\beta}\Big(\Big(\theta - \beta\,\tfrac{C'(u_0)}{K'(w_0)}\Big)\, d\widetilde{\Lambda}_1 + \widetilde{\Lambda}_1\, d\theta\Big) \\ &\quad + \frac{1}{\beta}\Big(\Big(\theta + \beta\, b\Big(\tfrac{\theta}{(1+\pi)\,\beta}\Big)\Big)\, d\Lambda_2 + \Lambda_2\, d\theta\Big) + \frac{\overline{\theta}}{\beta}\Big(F(\,\overline{\theta}\,) - \widetilde{\Lambda}_1(\,\overline{\theta}\,) - \Lambda_2(\,\overline{\theta}\,)\Big)\, dI, \\ r^* &= F(\,\overline{\theta}\,) - \widetilde{\Lambda}_1(\,\overline{\theta}\,) - \Lambda_2(\,\overline{\theta}\,) \end{aligned}$$

and I is the distribution function of the unit mass at $\overline{\theta}$. Finally, by construction of $\widetilde{\Lambda}_1$ and Λ_2, we have $u^* = 0$ and $r^* = 0$. So in fact $\nabla_x L(x_0, \lambda_1, \lambda_2) = 0$. In particular, the utility curve $x_0 = (u_0, w_0)$ associated with θ_1 does indeed maximize $L(\,\cdot\,, \lambda_1, \lambda_2)$.

Remark 63. *Great care is needed in choosing the space Z. One possible choice is $\mathcal{C}(\Theta, \mathbb{R})$, the space of all continuous functions on Θ endowed with the* sup *norm. This choice has the advantage that there is a convenient representation for Z^*. However, it also requires that $\Omega \subset \mathcal{C}(\Theta, \mathbb{R})$, and this is not an economically reasonable restriction. Another possible choice is $\mathcal{B}(\Theta, \mathbb{R})$, the space of all bounded functions on Θ endowed with the* sup *norm. This choice has the advantage that it includes all economically relevant utility curves. Unfortunately, it leads to a different problem: the measures $d\Lambda_1$ and $d\Lambda_2$ associated with $\Psi(\,\cdot\,;\theta_1)$ do not induce continuous linear functionals on $\mathcal{B}(\Theta, \mathbb{R})$, since functions in $\mathcal{B}(\Theta, \mathbb{R})$ are not in general measurable. The results of Luenberger (1969) do not therefore apply. Our solution to this double problem is to use $\mathcal{BV}(\Theta, \mathbb{R})$. This space is big enough to include all economically relevant utility curves, but small enough that $d\Lambda_1$ and $d\Lambda_2$ can be used to induce continuous linear functionals on it (albeit not in a canonical way).*

Q. Distributions

Q.1. Beta Distribution. The density of the generalization of the Beta that we consider is proportional to

$$(x-a)^{\zeta-1}(b-x)^{\eta-1}$$

on the interval (a, b), where $0 < a < b$ and $\zeta, \eta > 0$. It is unbounded at a if $\zeta < 1$, in which case we require that $\underline{\theta} \in (a, b)$ in order to ensure that A1 is satisfied, and unbounded at b if $\eta < 1$, in which case we require that $\overline{\theta} \in (a, b)$ in order to ensure that A1 is satisfied.

There are then four main cases. Three of the cases are easy to describe:

Case 1 if $\zeta > 1$ and $\eta \geq 1$ then A3 is satisfied for all choices of $\underline{\theta}, \overline{\theta} \in (a, b)$;

Case 2 if $\zeta > 1$ and $\eta < 1$ then A3 is again satisfied for all choices of $\underline{\theta}, \overline{\theta} \in (a, b)$, albeit for somewhat different reasons;

Case 4 if $\zeta < 1$ and $\eta < 1$, then A3 is violated for some choices of $\underline{\theta}, \overline{\theta} \in (a, b)$.

Case 3 is more involved. If $\zeta < 1$ and $\eta \geq 1$, then A3 is satisfied for all choices of $\underline{\theta}, \overline{\theta} \in (a, b)$ iff

$$\frac{\left(\sqrt{1-\zeta} + \sqrt{\eta - 1}\sqrt{\frac{a}{b}}\right)^2}{1 - \frac{a}{b}} \geq 1 + \frac{1}{1-\beta}. \tag{37}$$

As this inequality makes clear, A3 is more likely to be satisfied if: either (i) ζ is close to 0 (i.e. the spike at a is very pronounced); or (ii) η is large (i.e. the density decays very quickly towards b); or (iii) $\frac{a}{b}$ is close to 1 (i.e. the density is concentrated in a narrow band).[16] It is also worth noting that, as $\frac{a}{b} \to 0$, the left-hand side of (37) converges to $1 - \zeta < 1$. Hence A3 is violated for some choices of $\underline{\theta}, \overline{\theta} \in (a, b)$ when $\zeta < 1$ and $\frac{a}{b}$ is small. This is in striking contrast with the standard case studied in both Rice and Hogg et al. In that case A3 is satisfied for all $\underline{\theta}, \overline{\theta} \in (a, b)$ when $\zeta < 1$ and $\frac{a}{b} = 0$.

Note finally that the right-hand side of (37) is strictly increasing in β. Hence, if we fix a distribution for which $\zeta < 1$ and $\eta \geq 1$, then the conclusion is that A3 will be satisfied provided that β is *far enough below* 1. I.e. A3 is more likely to be satisfied when the decision maker is *more* time-inconsistent.

Q.2. Cauchy Distribution. The density of the general form of the Cauchy distribution is proportional to

$$\left(1 + \left(\frac{x-\mu}{\sigma}\right)^2\right)^{-1}$$

[16] For the purposes of the present discussion, $\zeta \in (0, 1)$, $\eta \in [1, \infty)$ and $\frac{a}{b} \in (0, 1)$.

on $\mathbb{R}$, where $\mu \in \mathbb{R}$ is a location parameter and $\sigma > 0$ is a scale parameter. This distribution satisfies A3 for all $\underline{\theta}, \overline{\theta} \in (0, \infty)$ iff

$$\frac{\mu}{\sigma} \leq \sqrt{\frac{1-(1-\beta)^2}{(1-\beta)^2}}. \tag{38}$$

In other words, taking β as given, A3 is satisfied iff the distribution is not located too far to the right. If (38) does not hold then, for some choices of $\underline{\theta}, \overline{\theta} \in (0, \infty)$, G is first increasing (at $\underline{\theta}$), then decreasing, then increasing again, then finally decreasing again (at $\overline{\theta}$).

We can also make $1-\beta$ the subject of the inequality (38). Doing so, we find that A3 is satisfied for all $\underline{\theta}, \overline{\theta} \in (0, \infty)$ iff: either $\mu \leq 0$; or $\mu > 0$ and

$$1-\beta \leq \left(1+\frac{\mu^2}{\sigma^2}\right)^{-\frac{1}{2}}.$$

In other words, taking the parameters μ and σ of the Cauchy distribution as given, A3 is satisfied iff: either $\mu \leq 0$; or $\mu > 0$ and β is sufficiently close to 1. I.e. A3 is more likely to be satisfied when the decision maker is less time-inconsistent.

Q.3. Log-Gamma Distribution. The density of the Log-Gamma distribution is proportional to

$$x^{-\frac{\eta+1}{\eta}} (\log(x))^{\zeta-1}$$

on $(1, \infty)$, where $\zeta, \eta > 0$. It is unbounded at 1 if $\zeta < 1$, in which case we require that $\underline{\theta} > 1$ in order to ensure that A1 is satisfied. It violates A3 for some choices of $\underline{\theta}, \overline{\theta} \in (1, \infty)$ iff $\zeta < 1$ and $\eta > 1-\beta$. In other words, taking β as given, A3 is violated iff there is a singularity at 1 and the rate of decay at ∞ is sufficiently slow.

Note finally that, if we fix a distribution for which $\zeta < 1$, then the conclusion is that A3 will be satisfied provided that β is *far enough below* 1. I.e. A3 is more likely to be satisfied when the decision maker is *more* time-inconsistent.

Q.4. Pareto Distribution. The density of the Pareto type II distribution is proportional to

$$\left(1+\frac{x-\mu}{\sigma}\right)^{-\zeta-1}$$

on (μ, ∞), where $\mu \in \mathbb{R}$ is a location parameter, $\sigma > 0$ is a scale parameter and $\zeta > 0$ is a shape parameter. It violates A3 for some choices of $\underline{\theta}, \overline{\theta} \in (\mu, \infty)$ iff

$$\zeta < \frac{1}{1-\beta} \tag{39}$$

and

$$\frac{\mu}{\sigma} > \frac{1}{\zeta+1}\left(1+\frac{1}{1-\beta}\right). \tag{40}$$

In other words, it violates A3 iff its right-hand tail is sufficiently fat and, taking the fatness of the tail as given, it is located sufficiently far to the right. In particular, if $\frac{\mu}{\sigma} \leq 1$, then the Pareto type II distribution satisfies A3 for all $\underline{\theta}, \overline{\theta} \in (\mu, \infty)$. For in that case: either (i) $\zeta \geq \frac{1}{1-\beta}$ and therefore (39) is violated; or (ii) $\zeta < \frac{1}{1-\beta}$, in which case $\frac{1}{\zeta+1}\left(1+\frac{1}{1-\beta}\right) > \frac{1}{\zeta+1}(1+\zeta) = 1 \geq \frac{\mu}{\sigma}$ and therefore (40) is violated.

We can also make $1-\beta$ the subject of these inequalities. Doing so, we find that A3 is violated for some $\underline{\theta}, \overline{\theta} \in (\mu, \infty)$ iff

$$\left(\frac{\mu}{\sigma}(\zeta+1)-1\right)^{-1} < 1-\beta < \zeta^{-1}.$$

In particular, if $\frac{\mu}{\sigma} > 1$ and $\zeta > 1$ (so that $\left(\frac{\mu}{\sigma}(\zeta+1)-1\right)^{-1} < \zeta^{-1} < 1$), then A3 is satisfied iff β is either close enough to 1 or *far enough below* 1.