Optimal Default Policies in Defined Contribution Pension Plans when Employees are Biased

Asen Ivanov

Working Paper No. 858 April 2018 ISSN 1473-0278

School of Economics and Finance

Queen Mary
University of London

Optimal Default Policies in Defined Contribution Pension Plans when Employees are Biased

Asen Ivanov*
Queen Mary University of London

May 10, 2018

Abstract

This paper analyses a model in which employees are biased in their perception of their optimal contribution rates or asset allocations in defined contribution pension plans. The optimal default is characterised as a function of the parameters. It is shown that, for some values of the parameters, forcing employees to actively decide is the optimal default policy. The total loss in the population at the optimal default policy can be nonmonotone in the parameters in counterintuitive ways.

Keywords: optimal defaults, libertarian paternalism, nudging, pension plan design
JEL: D14, D91, J26, J32

*School of Economics and Finance, 327 Mile End Road, London, United Kingdom; tel. +44 (0)20 7882 5886; a.ivanov@qmul.ac.uk

1 Introduction

A large literature documents that the default contribution rate and the default asset allocation in defined contribution pension plans have a large affect on employees' saving for retirement.[1] This raises the question of what is the optimal default policy. This question can be broken down into two parts. First, what is the optimal default? Second, how does the optimal default compare to not setting a default and forcing employees to make a decision? The latter default policy is known as "active decision" (AD).

Although these questions have been studied previously, there is a gap in the existing literature. On the one hand, the use of defaults is often motivated by the need to "nudge" people towards better options, e.g., to better contribution rates or better asset allocations.[2] In particular, a common view is that, left to their own devices, individuals would make systematic mistakes in choosing options, e.g., by undersaving. On the other hand, the literature on optimal default policies assumes either that individuals know their optimal options or that they have unbiased beliefs about them.

The current paper aims to address this gap by presenting and analysing a model with the following key features. First, for each employee, there is an optimal option, x, which corresponds either to her optimal contribution rate or to the optimal fraction of her pension plan portfolio invested in stocks (with the rest of the portfolio being invested in bonds).

Second, each employee is biased in the sense that, if her optimal option is x, her preferred option (i.e., the one she would choose) absent a default is $x + b$. One can interpret the bias, b, as a mistake that is either due to some kind of irrationality or to the lack of information.[3] Consider the following examples: (i) if one considers $\beta = 1$ as the normative benchmark within the β-δ model (Phelps and Pollak (1968) and Laibson (1997)),[4] individuals with $\beta < 1$ would tend to save too little, especially if

[1] See Madrian and Shea (2001), Choi et al. (2002, 2003, 2004a, 2004b, 2006), Beshears et al. (2008), Carroll et al. (2009), and Chetty et al. (2014). Bronchetti et al. (2013) observe no default effect.

[2] For example, see Thaler and Sunstein (2003).

[3] This bias is similar in spirit to a bias in a model in Campbell (2016), which assumes that some consumers mistakenly believe that their utility from a given financial product is higher than it really is. Both in the current paper and in Campbell's model, the bias induces suboptimal choices.

[4] For example, this is the position taken in O'Donoghue and Rabin (1999) and Carroll et al. (2009). However, this position is controversial (see Bernheim (2009) and Bernheim and Rangel

they are naïve about their future savings behaviour;[5] (ii) if young people with exponential growth bias[6] underestimate the long-term difference in wealth that results from investing in stocks at a, say, seven percent annual return rather than in bonds at a, say, one percent annual return, they would underweight stocks in their retirement portfolios; (iii) if young people are not aware of the high historical equity premium[7] or of the fact that historically equity returns have dominated bond returns over long horizons, they are likely to underweight stocks in their retirement portfolios.[8,9] Alternatively, one can view the bias as reflecting an externality rather than a mistake. For example, if a perfectly rational employee saves little for retirement because she anticipates being bailed out by the government, she would not be undersaving from the point of view of her own interest, but would be undersaving from the point of view of the government.

Third, if the planner (e.g., the employer or government official designing the pension plan) uses a default, D, each employee updates her preferred option in the direction of the default. In particular, an employee's updated preferred option equals $\alpha D + (1 - \alpha)(x + b)$, where $0 < \alpha < 1$ captures the strength of the updating. This feature of the model is motivated by several pieces of evidence. Madrian and Shea (2001) find that employees hired before automatic enrolment for new hires allocated three times as much to new hires' default investment fund if they chose their asset allocation after the introduction of automatic enrolment for new hires. Madrian and Shea also report evidence that employees hired after automatic enrolment who had

(2009)).

[5]An empirical literature investigates whether people actually undersave (relative to the discounted utility benchmark). The findings in this literature are mixed. (For an overview of some of the issues as well as further references, see Poterba (2015).)

[6]For evidence for exponential growth bias, see Eisenstein and Hoch (2005), Stango and Zinman (2009), as well as Levy and Tasoff (2016).

[7]There is some suggestive evidence of such unawareness: Beshears et al. (2017) report that experimental subjects invest more heavily in stocks when they are shown data on historical returns.

[8]Bernartzi and Thaler (1999) report that experimental subjects invest more heavily in stocks when shown 30-year rather than 1-year return distributions. Beshears et al. (2017) find that this effect is not robust, at least when subjects are shown 5-year vs. 1-year return distributions.

[9]There is evidence that education (Campbell (2006) and Calvet, Cambell, and Sodini (2007)) and IQ (Grinblatt, Keloharju, and Linnainmaa (2011)) have a positive effect on stock market participation. This is consistent with the view that nonparticipation, or, more generally, low investment in equities, is often a mistake. Bach, Calvet, and Sodini (2016) show that wealthier households in Sweden hold riskier financial portfolios. The associated higher average returns substantially increase inequality in financial wealth. Thus, to the extent that limited exposure to stocks is a mistake, it is one that contributes to inequality.

opted out of the default contribution rate or default investment fund were still much more heavily invested in the default investment fund than employees hired before automatic enrolment. Using a clever econometric procedure to estimate employees' preferred contribution rates, Beshears et al. (2016) find that these are affected by the default contribution rate, especially in the case of young, low-income employees.[10] Finally, if employees update their preferred contribution rates based on the default, this would explain their reluctance to opt out of the default without any need to invoke unrealistically high opt-out costs (see section II D in Bernheim et al. (2015)).[11]

Fourth, if an employee has to decide actively, either because she opts out of the default or because the default policy is AD, she incurs a cost, c.

Fifth, employees are heterogeneous–their optimal options are assumed to be uniformly distributed on an interval of length 2ϵ, so that ϵ captures the degree of heterogeneity.

Finally, if an employee ends up with option x' when her optimal option is x, this entails a loss from deviations equal to $|x - x'|$. An optimal default policy is one that minimises the total loss in the population.

After presenting the details of this model, which we will refer to as the baseline model, in section 2, the paper proceeds as follows. Section 3 characterises the optimal default as a function of the parameters (α, b, c, ϵ). It is shown that, for each configuration of the parameters, there is a unique optimal default, which belongs to one of six qualitatively different types of defaults.

Section 4 characterises the optimal default policy. In particular, this section shows that, for some values of the parameters, AD is better than the optimal default. This occurs because a default is a blunt tool for influencing employees' preferred options and, as a result, the optimal default may end up making some employees more biased, either by exacerbating their existing bias or by inducing a larger bias in the opposite direction. Section 4 also reveals a complicated relationship between the parameters and the optimal default policy. Some patterns in this relationship are discussed.

[10] Somewhat at odds with this, Madrian and Shea (2001) find that employees hired before automatic enrolment for new hires had contribution rates that were unaffected by whether they chose their contribution rates before or after the introduction of automatic enrolment for new hires.

[11] The two popular explanations for why individuals update their preferred options in light of the default are that (i) they view the default as a recommendation by the employer or (ii) they use the default as a psychological anchor (Tversky and Kahneman (1974)).

Section 5 notes that, for each parameter, the total loss at the optimal default policy need not be monotone in it. Thus, somewhat surprisingly, if employees pay more attention to the default (a higher α), have a smaller bias (a lower $|b|$), or are less heterogeneous (a lower ϵ), this doesn't always lead to a lower total loss at the optimal default policy.

For purposes of tractability, the baseline model is designed to be as simple as possible. As a result, it lacks realism in many ways. Section 6 discusses (without analysing in depth) some variations of the baseline model that add realism in important ways.

One of the insights from this paper is just how complicated the analysis of the optimal default policy is. Section 7 discusses this complexity. Section 8 concludes.

The most closely related papers from the literture are Carrol et al. (2009) and Bernheim et al. (2015), which uses the behavioural welfare framework of Bernheim and Rangel (2009). Adopting $\beta = 1$ as the normative benchmark, Carrol et al. (2009) solve for the optimal default policy in a model in which employees with β-δ preferences procrastinate opting out of suboptimal defaults. That model is quite different from the baseline model. In Carroll et al. (2009), employees are not biased and their preferred options are not influenced by the default. On the other hand, the central issue in Carroll et al. (2009) is that employees procrastinate switching from the default. This issue is not present in the baseline model. However, there is a close correspondence between the parameter β which governs procrastination in Carroll et al. (2009) and a parameter in a variation on the baseline model discussed in section 6.2.

Bernheim and Rangel (2009) introduce a framework for conducting welfare analysis in situations in which individuals make different choices under different frames. Bernheim and Rangel's approach is to first prune frames in which individuals' choices are based on a lack of understanding of the environment. The remaining frames are treated as welfare relevant, i.e., contradictory choices under these frames are all treated as equally valid from a normative perspective.

Bernheim et al. (2015) apply this framework to the question of optimal default contribution rates. In particular, they consider three models of choice in which behaviour is frame-dependent: a model with partially naive β-δ employees (sophistication and complete naïveté being special cases), a model with inattentive employees, and an "anchoring" model in which employees update their preferred contribution rates based on the default. Given a model of choice, two other ingredients in the welfare analysis are the naturally occurring frame (i.e., the frame under which choices are actually made) and the set of welfare-relevant frames from which the outcome in the naturally occurring frame is evaluated. Because different defaults lead to different outcomes under the naturally occurring frame and employees may evaluate these outcomes differently under different welfare-relevant frames, it may not be possible to unambiguously pin down an optimal default. Surprisingly, for the pertinent parameter values, it is often possible to pin down the optimal default without much ambiguity both in the model with partially naive β-δ employees and the model with inattentive employees. However, in the anchoring model, one cannot make precise welfare statements without dramatically restricting the set of welfare-relevant frames. This normative ambiguity is especially relevant given that, according to Bernheim et al., the anchoring model provides the most plausible explanation for employees' reluctance to opt out of the default.[12]

Bernheim et al. (2015) and the current paper differ in a number of respects. First, the current paper assumes a single welfare-relevant perspective. On the one hand, this makes the current paper less general. On the other, it allows us to avoid dealing with normative ambiguity. Second, a central issue in Bernheim et al. (2015) is that inattentiveness or, just like in Carroll et al. (2009), $\beta < 1$ in the β-δ model with a no-commitment frame impede opting out. This issue is not present in the baseline model. However, it does play a central role in the variation on the baseline model discussed in section 6.2. Third, the approach in Bernheim et al. does not allow us to say much about the optimal default policy if employees' preferred contribution rates absent a default involve a bias.[13] Fourth, while the analysis in Bernheim et al. (2015) is largely geared towards default contribution rates, the current analysis is equally applicable to the default allocation to stocks.

Two other related studies are Carlin et al. (2013) and Wisson (2016). In these

[12]Bernheim et al. suggest that, in the anchoring model, one might be able to use AD as the only welfare-relevant frame because of its neutrality. This is clearly not appropriate in a context in which employees' preferred options absent a default involve a bias.

[13]In the model with partially naive β-δ employees and the model with inattentive employees, each employee's preferred contribution rate is the same in all frames. In the anchoring model, an employee's preferred contribution rate differs across frames, but the normative ambiguity does not allow us to say much about the optimal default policy.

papers, each individual doesn't know her optimal option, but has an unbiased belief about it and can learn about it if she incurs a cost. In Carlin et al. (2013), defaults are informative and, thus, discourage individuals from learning about their optimal options.[14] Because learning involves a positive externality, AD may be optimal. In Wisson (2016), individuals with β-δ preferences are excessively reluctant to incur the cost of learning or free-ride by letting others incur this cost. As a result, the optimal default policy may be an extreme default that induces individuals to incur the cost of learning.

2 The Baseline Model

The baseline model has the following components. First, for each employee, there is an optimal option, x, which corresponds either to her optimal contribution rate or to the optimal fraction of her pension plan portfolio invested in stocks. Employees' optimal options are assumed to be uniformly distributed on $[r, r + 2\epsilon] \subseteq [0, 1]$, where $\epsilon > 0$ captures the degree of heterogeneity. This assumption is clearly a simplification. Notably, it excludes the realistic possibility that there is a positive mass of employees with (i) optimal contribution rate equal to zero or to the maximum contribution rate that benefits from an employer match or (ii) optimal allocation to stocks that equals zero or one.

Second, each employee is biased in the sense that, if her optimal option is x, her preferred option absent a default is $x + b$. We assume $b \leq 0$, which is without loss of generality.[15] We opt for an additive bias because it greatly simplifies the analysis. However, there is no reason to believe that an additive bias is more realistic than, say, a multiplicative bias. In addition, with an additive bias, $x + b < 0$ for some employees if $|b| > r$. Allowing some employees to choose $x + b < 0$ absent a default is unrealistic given that employees typically cannot choose a negative contribution rate or to short-sell stocks in their retirement portfolios. To avoid this possibility, we

[14] A similar mechanism is at work in Caplin and Martin (2017).

[15] If the employee has a tendency to set the contribution rate or the allocation to stocks too high, we can simply think of x as the optimal fraction of her salary that she does not contribute towards her pension plan or as the optimal fraction of her portfolio invested in bonds, respectively. With this relabelling, the assumption $b \leq 0$ would be appropriate.

assume $|b| \leq r$.[16] This assumption may itself be unrealistic. E.g., it is plausible that, say, young, cash-strapped individuals have an optimal contribution rate of zero, so that $r = 0$ and $|b| > r$ for any $b < 0$. Note that, even if $|b| > r$, the consequences of incorrectly assuming $|b| \leq r$ would not be severe as long as $|b| - r$ is small relative to 2ϵ.[17]

Third, if the planner uses a default, D, each employee updates her preferred option to $\hat{x} = \alpha D + (1-\alpha)(x+b)$, where $0 < \alpha < 1$ captures the strength of the updating.[18] This assumptions may be unrealistic if D is very far from x–in that case it is plausible that the employee simply disregards x.

Fourth, if an employee has to decide actively, either because she opts out of the default or because the default policy is AD, she incurs a cost, $c \geq 0$. We will refer to c as the opt-out cost, even though there is no opting out under AD. We interpret c as reflecting only implementation costs, i.e., the time and effort required to contact the relevant people in the human resources department, to fill in the necessary paperwork, etc.[19]

Fifth, we assume a linear loss from deviations. If an employee ends up with option x' when her optimal option is x and (assuming there is a default) her updated preferred option is $\hat{x}$, this entails a loss from deviations equal to $|x - x'|$ and a perceived loss from deviations (from the point of view of the employee) equal to $|\hat{x} - x'|$.[20] In reality, factors such as employer-match and tax-bracket thresholds, compounding of returns, and nonlinear utility of consumption are all likely to render a nonlinear loss from deviations more appropriate.

[16]Assuming $|b| \leq r$ will also ensure that the constraints in problem (2) below are not inconsistent and that the optimal default is nonnegative, which will allow us to avoid some complications arising from a corner solution.

[17]Suppose that $|b| > r$. In this case, incorrectly assuming $|b| \leq r$ is like ignoring the employees with $x \in [r, |b|)$ and treating the employees with $x \in [|b|, r + 2\epsilon]$ as the whole population. Employees with $x \in [r, |b|)$ have mass $min\left(\frac{|b|-r}{2\epsilon}, 1\right)$, so that ignoring them would not affect the welfare analysis much as long as $|b| - r$ is small relative to 2ϵ.

[18]The appendix considers the cases $\alpha = 0$ and $\alpha = 1$.

[19]Sections 6.1 and 6.2 consider other possibilities.

[20]We could set these losses equal to $\lambda|x - x'|$ and $\lambda|\hat{x} - x'|$, respectively, where $\lambda > 0$. However, the optimal default policy would be the same if $\lambda = 1$ and the opt-out cost is c and if $\lambda = \lambda'$ and the opt-out cost is $\lambda' c$. Thus, $\lambda = 1$ is a normalisation. Note that this normalisation does not absolve us from the need to estimate λ if we wish to plug in realistic parameter values into the model. In particular, suppose that in reality the opt-out cost equals \$30 and the loss from a one-unit deviation equals $\lambda \neq 1$ in dollar terms. Then, we need to plug $c = \frac{30}{\lambda}$ into our model.

In this setup, how does an employee with optimal option x behave and what loss does she incur? Under AD, the employee incurs the cost c and chooses $x + b$. As a result, she incurs a loss equal to $c + |x + b - x| = c - b$.

How about if she faces a default, D? Letting $\Delta = x - D$, the loss associated with staying with the default is $|\Delta|$. However, the perceived loss is $|\hat{x} - D| = (1 - \alpha)|\Delta + b|$. The employee stays with the default if $c \geq (1 - \alpha)|\Delta + b|$, which can be written as $\Delta_L \leq \Delta \leq \Delta_R$, where $\Delta_L = -\frac{c}{1-\alpha} - b$ and $\Delta_R = \frac{c}{1-\alpha} - b$. If the employee opts out, she incurs the cost c and chooses $\hat{x}$, so that her loss is $c + |x - \hat{x}| = c + \alpha|\Delta - \Delta_1|$, where $\Delta_1 = \frac{1-\alpha}{\alpha} b$. Δ_1 is the value of Δ for which the employee's bias is completely eliminated after she observes the default, so that $\hat{x} = x$. The further away an employee's Δ is from Δ_1, the less suitable the default is for correcting her bias.

Given the remarks in the previous paragraph, the loss of an employee with optimal option x who faces a default D is captured by the following function:

$$L(\Delta, \alpha, b, c) = \begin{cases} |\Delta| & \text{if } \Delta_L \leq \Delta \leq \Delta_R \\ c + \alpha|\Delta - \Delta_1| & \text{otherwise} \end{cases}. \tag{1}$$

This function depends on x and D only through Δ, which greatly simplifies the analysis.

Figure 1 shows the graph of $L(\cdot, \alpha, b, c)$ for each of the cases $\Delta_L \leq \Delta_1$, $\Delta_1 < \Delta_L \leq 0$, and $\Delta_L > 0$. The following lemma establishes that key features of Figure 1 hold more generally.

Lemma 1

1) $L(\cdot, \alpha, b, c)$ *is continuous at* Δ_R*.*

2) If $\Delta_L \leq \Delta_1$*,* $L(\cdot, \alpha, b, c)$ *is continuous at* Δ_L*.*

3) If $\Delta_1 < \Delta_L \leq 0$*,* $L(\Delta_L, \alpha, b, c) < c < \lim_{\Delta \uparrow \Delta_L} L(\Delta, \alpha, b, c)$*.*

4) If $\Delta_L > 0$*,* $L(\Delta_L, \alpha, b, c) \leq \lim_{\Delta \uparrow \Delta_L} L(\Delta, \alpha, b, c)$*, the inequality being strict if and only if* $c > 0$*.*

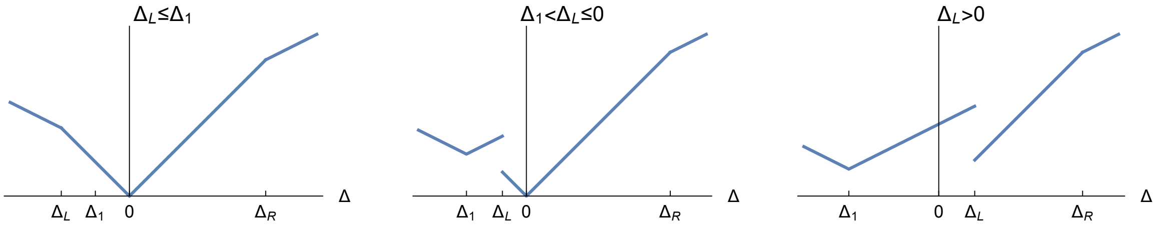

Figure 1: Graph of $L(\cdot, \alpha, b, c)$.

3 The Optimal Default

In this section, we set up and solve the planner's problem for finding the optimal default. The next section addresses the possibility that the optimal default policy is AD.

We assume that the planner faces two constraints. First, we assume $0 \le D \le 1$. This simply reflects the natural constraint that the planner cannot set a default contribution rate or a default allocation to stocks outside of $[0, 1]$. Second, we assume that D must be such that, for at least some value of $x \in [r, r + 2\epsilon]$, employees are willing to stay with the default. This assumption reflects the fact that, if the planner were to set a default from which everybody opts out, she is likely to face employees' resentment or even lawsuits.[21]

Given that the different possible optimal values of x lie in $[r, r + 2\epsilon]$, the corresponding Δ's lie in $[r - D, r + 2\epsilon - D]$, the latter interval being a translation of the former interval by D to the left. By choosing D, the planner shifts around the interval $[r - D, r + 2\epsilon - D]$ on the horizontal axis in each panel in Figure 1. Denoting the upper endpoint of this interval as $\bar{\Delta}$ (i.e., $\bar{\Delta} = r + 2\epsilon - D$), we can think of the planner as directly choosing $\bar{\Delta}$. For a given $\bar{\Delta}$, the position of the default, D, relative to $[r, r+2\epsilon]$ can be inferred from the position of $\Delta = 0$ relative to the interval $[\bar{\Delta} - 2\epsilon, \bar{\Delta}]$. E.g., if $\Delta = 0$ is below/in the lower end of/in the middle of/in the upper end of/above $[\bar{\Delta} - 2\epsilon, \bar{\Delta}]$, then D is below/in the lower end of/in the middle of/in the upper end of/above $[r, r + 2\epsilon]$.

The total loss in the population associated with a given value of $\bar{\Delta}$ equals $\frac{1}{2\epsilon}$ times the area under $L(\cdot, \alpha, b, c)$ over the interval $[\bar{\Delta} - 2\epsilon, \bar{\Delta}]$. The optimal $\bar{\Delta}$ solves:

[21] This constraint also significantly facilitates the analysis by making the planner's objective function in problem (2) below unimodal.

$$\min_{\substack{r+2\epsilon-1\leq\bar{\Delta}\leq r+2\epsilon,\\ \Delta_L\leq\bar{\Delta}\leq\Delta_R+2\epsilon}} \frac{1}{2\epsilon}\int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} L(\Delta,\alpha,b,c)d\Delta. \tag{2}$$

The first constraint is equivalent to $0 \leq D \leq 1$. The second constraint captures the requirement that at least some employees would be willing to stay with the default. The following proposition characterises the solution to problem (2).

Proposition 1 *Let $\Delta' = \frac{c}{1-\alpha} - \frac{2\alpha}{1-\alpha}\epsilon - b$ and $\Delta'' = \frac{c}{1+\alpha} + \frac{2\alpha}{1+\alpha}\epsilon + \frac{1-\alpha}{1+\alpha}b$. The unique solution to problem (2) is:*

$$\bar{\Delta}(\alpha,b,c,\epsilon) =$$

$$= \begin{cases} \epsilon & \text{if } \Delta_L \leq -\epsilon \\ \Delta_L + 2\epsilon & \text{if } L(\Delta_L,\alpha,b,c) < L(\Delta_L+2\epsilon,\alpha,b,c) \leq \lim_{\Delta\uparrow\Delta_L} L(\Delta,\alpha,b,c) \\ \Delta' & \text{if } \Delta' > \Delta_L, \Delta_1 \leq \Delta' - 2\epsilon < \Delta_L \\ \Delta'' & \text{if } \Delta_L < \Delta'' \leq \Delta_R, \Delta''-2\epsilon < \Delta_L, \Delta''-2\epsilon < \Delta_1 \\ \Delta_1+\epsilon & \text{if } \Delta_1+\epsilon > \Delta_R \\ \Delta_L & \text{if } L(\Delta_L-2\epsilon,\alpha,b,c) \leq L(\Delta_L) \end{cases} \tag{3}$$

$$= \begin{cases} \epsilon & \text{if } c \geq (1-\alpha)\epsilon - (1-\alpha)b \\ \Delta_L + 2\epsilon & \text{if } c < (1-\alpha)\epsilon - (1-\alpha)b, c \geq \epsilon \\ \Delta' & \text{if } c > \alpha\epsilon, c \geq 2\epsilon + \frac{1-\alpha}{\alpha}b, c < \epsilon \\ \Delta'' & \text{if } c > -\alpha(1-\alpha)\epsilon - (1-\alpha)b, c \geq (1-\alpha)\epsilon + \frac{1-\alpha}{\alpha}b, \\ & \quad c < (1-\alpha)\epsilon - (1-\alpha)b, c < 2\epsilon + \frac{1-\alpha}{\alpha}b \\ \Delta_1+\epsilon & \text{if } c < (1-\alpha)\epsilon + \frac{1-\alpha}{\alpha}b \\ \Delta_L & \text{if } c \leq \alpha\epsilon, c \leq -\alpha(1-\alpha)\epsilon - (1-\alpha)b \end{cases} \tag{4}$$

Expression (4) rewrites the conditions in each case in expression (3) in terms of the underlying parameters. By way of notation, let $\Psi(\alpha,b,c,\epsilon) = \int_{\bar{\Delta}(\alpha,b,c,\epsilon)-2\epsilon}^{\bar{\Delta}(\alpha,b,c,\epsilon)} L(\Delta,\alpha,b,c)d\Delta$, so that $\frac{1}{2\epsilon}\Psi(\alpha,b,c,\epsilon)$ is the total loss associated with $\bar{\Delta}(\alpha,b,c,\epsilon)$.

The first case in expressions (3) and (4) is illustrated in the two panels of Figure 2 (one panel for $\Delta_L \leq \Delta_1$ and one for $\Delta_1 < \Delta_L \leq 0$).[22] This case corresponds to a default placed in the middle of $[r, r+2\epsilon]$, with nobody opting out. We call it the all-stay centre (AS-C) default.

[22] $\Delta_L \leq -\epsilon$ rules out $\Delta_L > 0$.

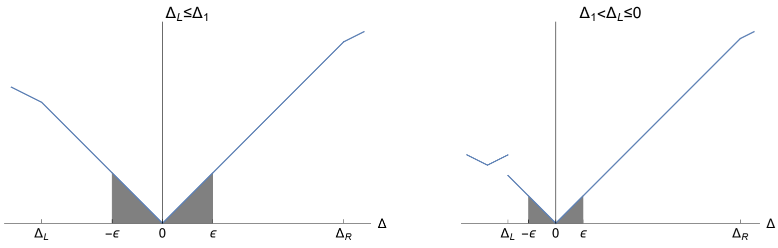

Figure 2: AS-C default ($\bar{\Delta}(\alpha, b, c, \epsilon) = \epsilon$). The shaded area equals $\Psi(\alpha, b, c, \epsilon)$.

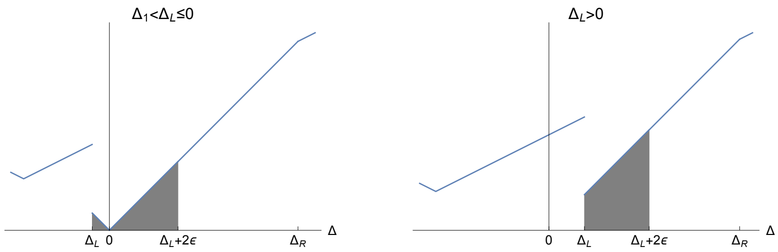

Figure 3: AS-L default ($\bar{\Delta}(\alpha, b, c, \epsilon) = \Delta_L + 2\epsilon$). The shaded area equals $\Psi(\alpha, b, c, \epsilon)$.

The second case in expressions (3) and (4) is illustrated in the two panels of Figure 3 (one panel for $\Delta_1 < \Delta_L \leq 0$ and one for $\Delta_L > 0$).[23] This case corresponds to a low default placed either inside $[r, r + 2\epsilon]$ near its lower endpoint (left panel) or below $[r, r + 2\epsilon]$ (right panel). Either way, nobody opts out. We call this the all-stay low (AS-L) default.

The third case in expressions (3) and (4) is illustrated in the two panels of Figure 4 (one panel for $\Delta_1 < \Delta_L \leq 0$ and one for $\Delta_L > 0$).[24] Note that Δ' is the value of Δ where the linear extension of the right branch of $c + \alpha|\Delta - \Delta_1|$ shifted to the right by 2ϵ and the linear extension of the right branch of $|\Delta|$ intersect,[25] which is why $L(\Delta') = L(\Delta' - 2\epsilon)$ in the figure. This case corresponds to a default placed either inside $[r, r + 2\epsilon]$ (left panel) or below $[r, r + 2\epsilon]$ (right panel).[26] Employees with high values of x stay while those with low values of x opt out. We call this the type-1 high-stay (HS1) default.

[23] $c < (1-\alpha)\epsilon - (1-\alpha)b$ and $c \geq \epsilon$ imply $c < (1-\alpha)c - (1-\alpha)b$. The latter inequality is equivalent to $\Delta_1 < \Delta_L$. Thus, the case $\Delta_L \leq \Delta_1$ is ruled out.

[24] Obviously, $\Delta_1 < \Delta_L$ rules out the case $\Delta_L \leq \Delta_1$.

[25] I.e., Δ' is the solution to $\Delta = c + \alpha(\Delta - 2\epsilon - \Delta_1)$.

[26] In the right panel, we could also have $\Delta' - 2\epsilon < 0$ (if ϵ were slightly larger).

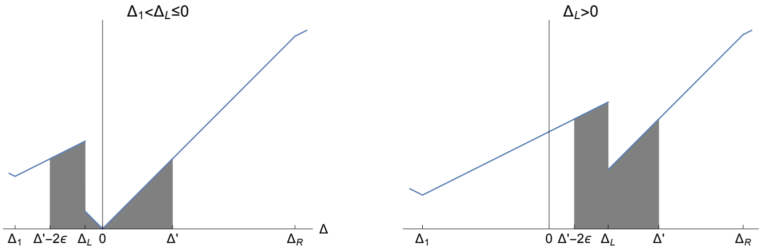

Figure 4: HS1 default ($\bar{\Delta}(\alpha, b, c, \epsilon) = \Delta'$). The shaded area equals $\Psi(\alpha, b, c, \epsilon)$.

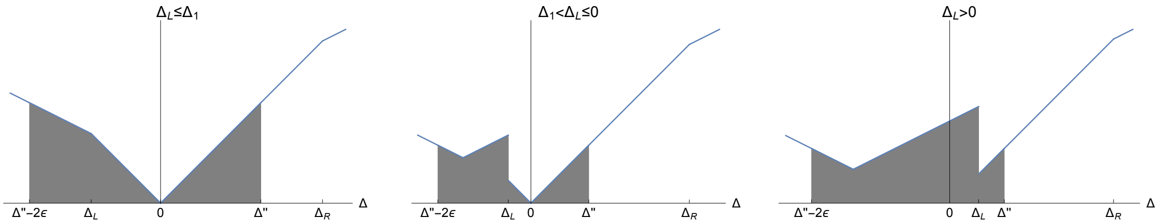

Figure 5: HS2 default ($\bar{\Delta}(\alpha, b, c, \epsilon) = \Delta''$). The shaded area equals $\Psi(\alpha, b, c, \epsilon)$.

The fourth case in expressions (3) and (4) is illustrated in the three panels of Figure 5 (one panel for $\Delta_L \leq \Delta_1$, one for $\Delta_1 < \Delta_L \leq 0$, and one for $\Delta_L > 0$). Note that Δ'' is the value of Δ where the linear extension of the left branch of $c+\alpha|\Delta-\Delta_1|$ shifted to the right by 2ϵ and the linear extension of the right branch of $|\Delta|$ intersect,[27] which is why $L(\Delta'') = L(\Delta'' - 2\epsilon)$ in the figure. This case corresponds to a default placed inside $[r, r+2\epsilon]$, nearer to the upper endpoint.[28] Employees with high values of x stay while those with low values of x opt out. We call this the type-2 high-stay (HS2) default.

The fifth case in expressions (3) and (4) is illustrated in the three panels of Figure 6 (one panel for $\Delta_L \leq \Delta_1$, one for $\Delta_1 < \Delta_L \leq 0$, and one for $\Delta_L > 0$). It corresponds to a default placed inside $[r, r+2\epsilon]$, nearer to the upper endpoint. Employees with intermediate values of x stay, while those with high or low values of x opt out. We call this the middle-stay high interior (MS-HI) default.

The sixth case in expressions (3) and (4) is illustrated in the two panels in Figure 7 (one panel for $\Delta_L > 0$ and $\Delta_L - 2\epsilon \geq \Delta_1$ as well as one panel for $\Delta_L > 0$ and

[27] I.e., Δ'' is the solution to $\Delta = c - \alpha(\Delta - 2\epsilon - \Delta_1)$.

[28] $\Delta'' - 2\epsilon < \Delta_L$ is equivalent to $\Delta'' < -(\Delta'' - 2\epsilon)$. Thus, $\Delta = 0$ is closer to Δ'' than to $\Delta'' - 2\epsilon$ in each panel of Figure 5.

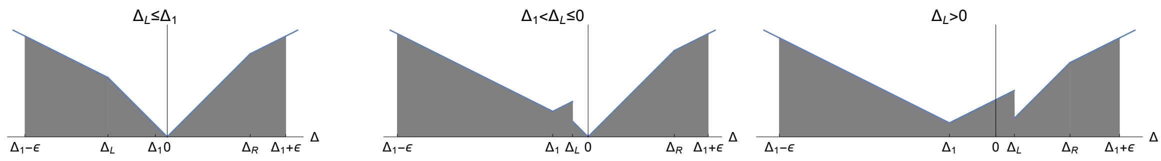

Figure 6: MS-HI default ($\bar{\Delta}(\alpha, b, c, \epsilon) = \Delta_1 + \epsilon$). The shaded area equals $\Psi(\alpha, b, c, \epsilon)$.

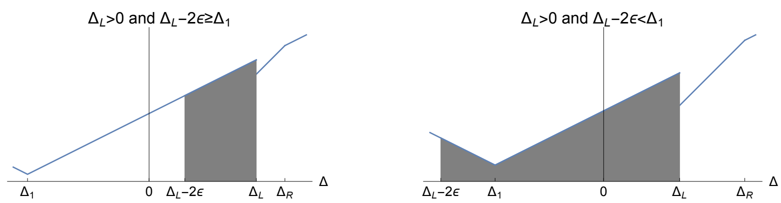

Figure 7: NS-corner default ($\bar{\Delta}(\alpha, b, c, \epsilon) = \Delta_L$). The shaded area equals $\Psi(\alpha, b, c, \epsilon)$.

$\Delta_L - 2\epsilon < \Delta_1$).[29,30] This corresponds to a default placed so that only an employee with $x = r$ (i.e., with the lowest possible x) is willing to stay. Because the set of employees with $x = r$ has mass zero, we call this the none-stay corner (NS-corner) default.

Note that three of the six types of defaults discussed above–namely the HS1, HS2, and NS-corner defaults–are at odds with the suggestion in Thaler and Sunstein (2003) that the optimal default minimise the number of people opting out.

The proof that the optimal default is given by expression (3) is based on the fact that, as $\bar{\Delta}$ moves away from $\bar{\Delta}(\alpha, b, c, \epsilon)$ on $[\Delta_L, \infty)$ in each panel of Figures 2-7, $\int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} L(\Delta, \alpha, b, c) d\Delta$ monotonically increases.

To conclude this section, let $R_{\text{AS-C}}$, $R_{\text{AS-L}}$, R_{HS1}, R_{HS2}, $R_{\text{MS-HI}}$, and $R_{\text{NS-corner}}$ denote the regions of the parameter space in which the conditions in case 1-6, respectively, in expressions (3) and (4) hold. For a region of the parameter space, R, let $R(\alpha, \epsilon)$ denote its section for fixed α and ϵ in $(|b|, c)$-space. The left panel in Figure 9 illustrates $R_{\text{AS-C}}(0.35, 0.1)$, $R_{\text{AS-L}}(0.35, 0.1)$, etc. The right panel is analogous, except that α is fixed at 0.65 instead of at 0.35. (For now ignore the shading near the origin in each panel in the figure.) The figure also shows in square brackets the value of

[29] $\Delta_L \le 0$ implies $L(\Delta_L - 2\epsilon, \alpha, b, c) > L(\Delta_L)$ (see Figure 1). Thus, $L(\Delta_L - 2\epsilon, \alpha, b, c) \le L(\Delta_L)$ implies $\Delta_L > 0$.

[30] In the left panel, we could also have $\Delta_L - 2\epsilon < 0$ (if ϵ were slightly larger).

$\bar{\Delta}(\alpha, b, c, \epsilon)$ in each region. This figure is discussed further in the next section.

4 Optimal Default Policy

The optimal default policy is either $\bar{\Delta}(\alpha, b, c, \epsilon)$ or AD depending on whether $\frac{1}{2\epsilon}\Psi(\alpha, b, c, \epsilon)$ or the total loss under AD, $c - b$, is smaller, respectively. Before characterising the optimal default policy, let us pause and consider the question: Why would AD ever be optimal? The reason is that a default is a blunt tool for influencing employees' preferred options. As a result, the optimal default may end up making some employees more biased, either by exacerbating their existing bias or by inducing a larger (in absolute value) bias in the opposite direction. In particular, for an employee with $\Delta > -b$, the default is below $x + b$ ($\Delta > -b$ is equivalent to $D < x + b$), so that it exacerbates her bias; for an employee with $\Delta < \frac{2-\alpha}{\alpha}b$, the default is so far above x that it not only undoes her bias, but also biases her by more than $|b|$ in the other direction ($\Delta < \frac{2-\alpha}{\alpha}b$ is equivalent to $\hat{x} > x - b$). The default attenuates the bias of an employee with $\frac{2-\alpha}{\alpha}b < \Delta < -b$ ($\frac{2-\alpha}{\alpha}b < \Delta < -b$ is equivalent to $|x - \hat{x}| < |b|$).

Given a default, we can classify each employee into one of the following four categories depending on how she fares under it relative to under AD.[31]

A) The employee's bias is made worse by the default, she opts out, and makes a worse choice than under AD.

B) The employee's bias is made worse by the default, she stays with the default, and she is worse off than under AD.

C) The employee's bias is made worse by the default, she stays with the default, and she is better off than under AD.

D) The employee's bias is attenuated by the default, which implies that she is better off than under AD.[32]

[31] We omit from this classification employees who fare equally well under the default and under AD. We can safely do this, because these employees are of measure zero.

[32] If the employee opts out, she is clearly better off than under AD. If she stays with the default and $\Delta \geq 0$, she is better off than under AD: $\Delta < -b$ (which holds if the employee's bias is attenuated) and $\Delta \geq 0$ imply $|\Delta| < c - b$. If she stays with the default and $\Delta < 0$, she is again better off than under AD: $(2-\alpha)b < \alpha\Delta$ (which holds if the employee's bias is attenuated), $-c-(1-\alpha)b \leq (1-\alpha)\Delta$

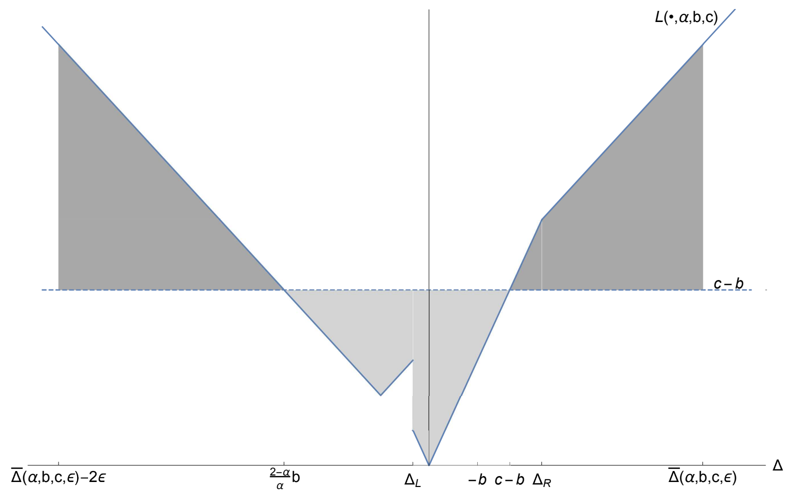

Figure 8: Example in which AD is better than the optimal default. Employees with $\Delta \in (\frac{2-\alpha}{\alpha}b, c-b)$ are better off under the optimal default. Employees with $\Delta \in [\bar{\Delta}(\alpha,b,c,\epsilon) - 2\epsilon, \bar{\Delta}(\alpha,b,c,\epsilon)] \setminus [\frac{2-\alpha}{\alpha}b, c-b]$ are better off under AD.

When the interests of employees in categories A) and B) outweigh the interests of those in categories C) and D), AD is the optimal default policy. To illustrate such a case, consider Figure 8. In the figure, employees with $\bar{\Delta}(\alpha,b,c,\epsilon) - 2\epsilon \leq \Delta < \frac{2-\alpha}{\alpha}b$ or $\Delta_R < \Delta \leq \bar{\Delta}(\alpha,b,c,\epsilon)$ fall in category A); employees with $c-b < \Delta \leq \Delta_R$ fall in category B); employees with $-b < \Delta < c-b$ fall in category C); employees with $\frac{2-\alpha}{\alpha}b < \Delta < -b$ fall in category D). The figure is drawn so that the interests of employees in categories A) and B) outweigh the interests of those in categories C) and D)–the two dark gray areas taken together are larger than the light gray area.

Let us now characterise the optimal default policy.

Proposition 2

1) The equation $\frac{1}{2\epsilon}\Psi(\alpha,b,c,\epsilon) = c-b$ *implicitly defines a unique function* c^{AD} : $\{(\alpha,b,\epsilon) \in (0,1) \times (-\infty,0] \times (0,\infty) : b \geq -\frac{\alpha\epsilon}{2}\} \to [0,\infty)$.

2) The set of optimal default policies is

(which holds given that the employee stays with the default), and $\Delta < 0$ imply $|\Delta| < c-b$ (to see this, add the first two inequalities).

$$\begin{cases} \{\bar{\Delta}(\alpha,b,c,\epsilon)\} & \text{if } c > c^{AD}(\alpha,b,\epsilon) \text{ or } b < -\frac{\alpha\epsilon}{2} \\ \{\bar{\Delta}(\alpha,b,c,\epsilon), AD\} & \text{if } c = c^{AD}(\alpha,b,\epsilon) \\ \{AD\} & \text{if } c < c^{AD}(\alpha,b,\epsilon) \end{cases}.$$

3) c^{AD} *is increasing in* b *and nondecreasing in* α. $c^{AD}(\alpha, -\frac{\alpha\epsilon}{2}, \epsilon) = 0$ *and* $c^{AD}(\alpha, 0, \epsilon) =$
$$\begin{cases} \epsilon\sqrt{\alpha(1-\alpha)} & \text{if } \alpha < 0.5 \\ 0.5\epsilon & \text{if } \alpha \geq 0.5 \end{cases}.$$

The left panel of Figure 9 shows the graph of $c^{\mathrm{AD}}(0.35, \cdot, 0.1)$ in $(|b|, c)$-space as the northern boundary of the shaded region around the origin. The right panel is analogous, except that α is fixed at 0.65. AD is optimal in the shaded region, which we will denote by $R_{\mathrm{AD}}(\alpha, \epsilon)$, and $\bar{\Delta}(\alpha, b, c, \epsilon)$ is optimal outside of it (or on its northern boundary).

Figure 9 holds ϵ constant at 0.1. The following proposition tells us how each panel in the figure changes when ϵ varies.

Proposition 3 *For* $k > 0$*, (i) the AS-C/AS-L/HS1/HS2/MS-HI/NS-corner default is the optimal default given* (α, b, c, ϵ) *if and only if it is the optimal default given* $(\alpha, kb, kc, k\epsilon)$ *and (ii) AD is an optimal default policy given* (α, b, c, ϵ) *if and only if it is an optimal default policy given* $(\alpha, kb, kc, k\epsilon)$*.*

The proposition implies that, for any arbitrary ϵ, we can obtain a version of each panel of Figure 9 by stretching each of the regions in the panel by a factor of $\frac{\epsilon}{0.1}$ away from the origin.

Note that Figure 9, coupled with the last observation, reveals a complicated relationship between, on the one hand, the parameters and, on the other hand, the optimal default and the optimal default policy. Below, some key patterns are identified.[33] Whenever possible, an explanation for the observed pattern is provided.[34]

1. Fixing α, b, and c, the MS-HI default is the optimal default (i.e., $\bar{\Delta}(\alpha, b, c, \epsilon) = \Delta_1 + \epsilon$) for ϵ large enough. This holds because, as $\epsilon \to \infty$, $\Delta_1 + \epsilon > \Delta_R$, so that the fifth case in expression (3) applies.

[33] Given that the inequalities in each case in expression (4) are linear in b and c, it is straightforward to verify that these patterns hold more generally, not just for the values of α and ϵ used in Figure 9.

[34] Regarding any statement below that holds if ϵ or $|b|$ is large enough, one should keep in mind that the constraints $|b| \leq r$ and $r + 2\epsilon \leq 1$ may render the values of ϵ or $|b|$, respectively, that are large enough infeasible.

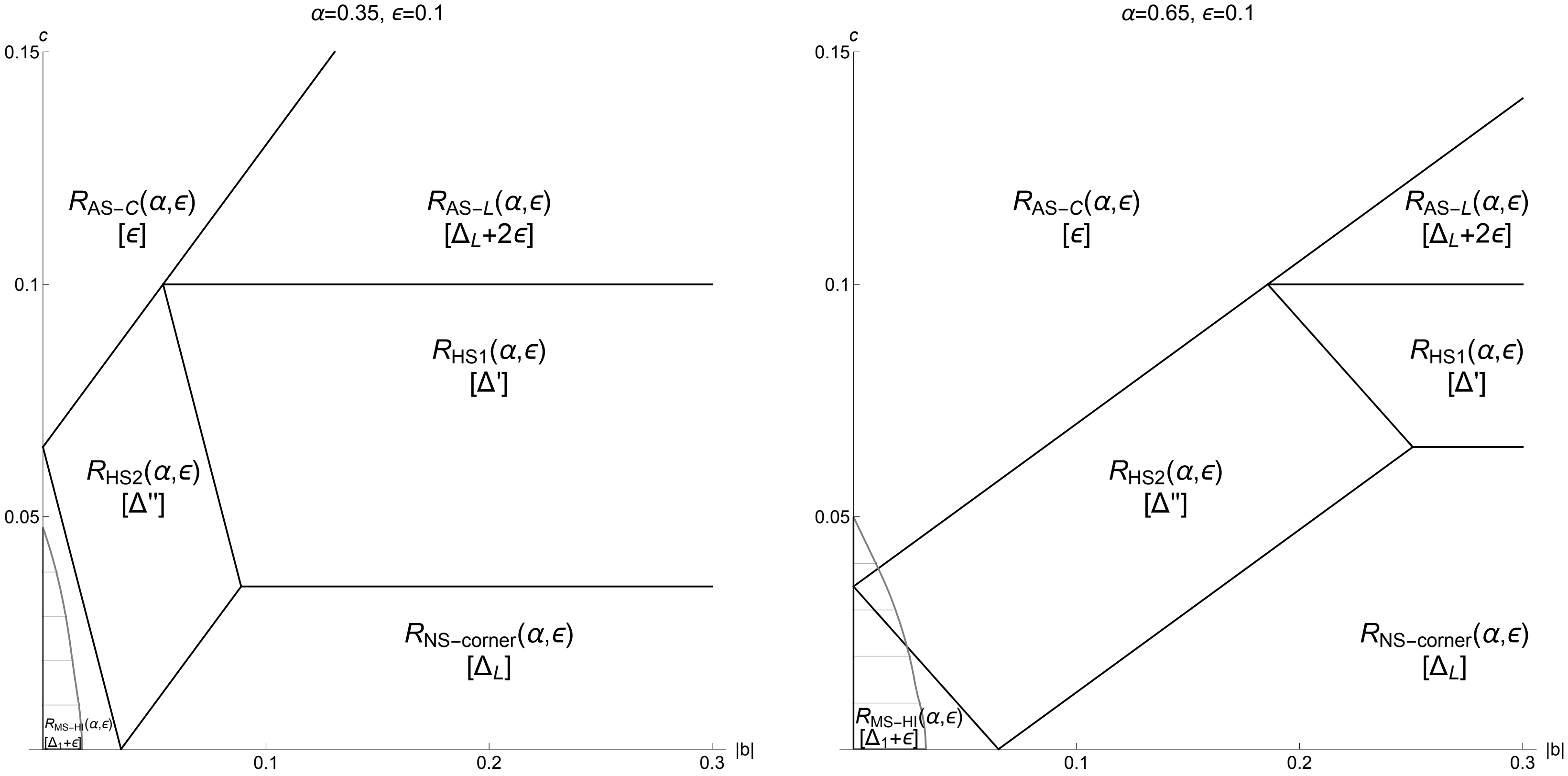

Figure 9: $R_{\text{AS-C}}(\alpha, \epsilon)$, $R_{\text{AS-L}}(\alpha, \epsilon)$, $R_{\text{HS1}}(\alpha, \epsilon)$, $R_{\text{HS2}}(\alpha, \epsilon)$, $R_{\text{MS-HI}}(\alpha, \epsilon)$, $R_{\text{NS-corner}}(\alpha, \epsilon)$, and $R_{\text{AD}}(\alpha, \epsilon)$. $R_{\text{AD}}(\alpha, \epsilon)$ is the shaded region. $\bar{\Delta}(\alpha, b, c, \epsilon)$ is given in square brackets.

2. Fixing α, b, and c, AD is the optimal default policy for ϵ large enough. This holds because, as $\epsilon \to \infty$, (i) the MS-HI default is the optimal default (by the previous point), (ii) the average value of $L(\cdot, \alpha, b, c)$ over $[\bar{\Delta}(\alpha, b, c, \epsilon) - 2\epsilon, \bar{\Delta}(\alpha, b, c, \epsilon)] = [\Delta_1 - \epsilon, \Delta_1 + \epsilon]$ becomes arbitrarily large, so that $\frac{1}{2\epsilon}\Psi(\alpha, b, c, \epsilon)$ becomes arbitrarily large, and (iii) the total loss associated with AD, $c - b$, is unaffected by ϵ. Intuitively, when ϵ is large, even with the optimal default most employees' Δ's lie both outside of $[\frac{2-\alpha}{\alpha}b, -b]$ and outside of $[\Delta_L, \Delta_R]$, so that they fall in category A) above.[35]

3. AD is optimal at $(\alpha, kb, kc, \epsilon)$ for small enough $k > 0$. By Proposition 3, this statement is equivalent to the statement "AD is optimal at $(\alpha, b, c, \frac{\epsilon}{k})$ for small enough $k > 0$", which holds by the previous point. Intuitively, when b and c are small, even with the optimal default most employees' Δ's lie both outside of $[\frac{2-\alpha}{\alpha}b, -b]$ and outside of $[\Delta_L, \Delta_R]$ (employees are willing to opt out given the low opt-out cost), so that they fall in category A) above.

4. Fixing α, ϵ, and b, the AS-C default is the optimal default policy for c high enough. This holds because (i) $\Delta_L \leq -\epsilon$ for c high enough, so that the AS-C default is the optimal default (see expression (3)), (ii) under the AS-C default, the total loss, $\frac{\epsilon}{2}$, does not depend on c, and (iii) the total loss under AD, $c - b$, grows without bound as c increases.

5. Fixing (α, ϵ) and fixing c at a small enough value, the NS-corner default is the optimal default policy for $|b|$ large enough. To see why this holds note the following. For $|b|$ large enough, $\Delta_L - 2\epsilon \geq 0$, so that $L(\Delta_L, \alpha, b, c) - L(\Delta_L - 2\epsilon, \alpha, b, c) = \Delta_L - c - \alpha(\Delta_L - 2\epsilon - \Delta_1) = 2(\alpha\epsilon - c)$. Thus, for $c \leq \alpha\epsilon$, $L(\Delta_L - 2\epsilon, \alpha, b, c) \leq L(\Delta_L, \alpha, b, c)$, so that the last case in expression (3) applies and the NS-corner default is the optimal default (i.e., $\bar{\Delta}(\alpha, b, c, \epsilon) = \Delta_L$). Moreover, for $|b|$ large enough, $[\Delta_L - 2\epsilon, \Delta_L) = [-\frac{c}{1-\alpha} - b - 2\epsilon, -\frac{c}{1-\alpha} - b) \subset (\frac{2-\alpha}{\alpha}b, -b)$. This means that, with $\bar{\Delta}(\alpha, b, c, \epsilon) = \Delta_L$, the bias of (almost) all employees is attenuated, so that they are better off than under AD.

[35] Moreover, some of these employees' Δ's lie far outside of $[\frac{2-\alpha}{\alpha}b, -b]$, so that the default severely exacerbates their bias and they opt out to highly unsuitable options.

6. Fixing ϵ and increasing α, $R_{\text{AD}}(\alpha, \epsilon)$ expands. This occurs because c^{AD} is non-decreasing in α.

7. Fixing ϵ and increasing α, the region in which the AS-C default is the optimal default policy (i.e., $R_{\text{AS-C}}(\alpha, \epsilon) \setminus R_{\text{AD}}(\alpha, \epsilon)$) expands. This occurs for the following reasons. As α increases, $\Delta_L = -\frac{c}{1-\alpha} - b \leq -\epsilon$ for more values of b and c (employees are more reluctant to opt out), so that the AS-C default is the optimal default for more values of b and c. This explains why $R_{\text{AS-C}}(\alpha, \epsilon)$ expands. However, as α increases, couldn't the expansion of $R_{\text{AD}}(\alpha, \epsilon)$ eat into $R_{\text{AS-C}}(\alpha, \epsilon)$, so that $R_{\text{AS-C}}(\alpha, \epsilon) \setminus R_{\text{AD}}(\alpha, \epsilon)$ fails to expand? The answer is "No" for the following reason. Suppose that the AS-C default is better than AD for some (α, b, c, ϵ). Increasing α cannot reverse this because the total loss under the AS-C default, $\frac{\epsilon}{2}$, and the total loss under AD, $c - b$, do not depend on α.

8. Fixing ϵ and increasing α, the region in which the AS-C or AS-L default is the optimal default policy (i.e., $\big(R_{\text{AS-C}}(\alpha, \epsilon) \cup R_{\text{AS-L}}(\alpha, \epsilon)\big) \setminus R_{\text{AD}}(\alpha, \epsilon)$) expands.

9. Fixing ϵ and increasing α, the region in which the AS-L default is the optimal default policy (i.e., $R_{\text{AS-L}}(\alpha, \epsilon) \setminus R_{\text{AD}}(\alpha, \epsilon)$) contracts.

10. Fixing α and increasing ϵ, $R_{\text{AD}}(\alpha, \epsilon)$ expands, while $R_{\text{AS-C}}(\alpha, \epsilon) \setminus R_{\text{AD}}(\alpha, \epsilon)$ and $\big(R_{\text{AS-C}}(\alpha, \epsilon) \cup R_{\text{AS-L}}(\alpha, \epsilon)\big) \setminus R_{\text{AD}}(\alpha, \epsilon)$ contract. This follows from the observation immediately following Proposition 3.

5 Nonmonotonicity of the Total Loss

For each parameter, the total loss at the optimal default, $\frac{1}{2\epsilon}\Psi(\alpha, b, c, \epsilon)$, need not be monotone in it. The same applies to the total loss at an optimal default policy, $\min\big(\frac{1}{2\epsilon}\Psi(\alpha, b, c, \epsilon), c - b\big)$. Thus, somewhat surprisingly, a higher α, a lower $|b|$, and a lower ϵ do not always lead to a lower total loss at the optimal default or at the optimal default policy.[36] These nonmonotonicities are illustrated via examples in the appendix.

[36] Increasing c shifts up the graph of $L(\cdot, \alpha, b, c)$, but also gives the planner more leverage because opting out of the default is more costly. Thus, it is not too surprising that the total loss at the optimal default and at the optimal default policy is nonmonotone in c.

6 Variations on the Baseline Model

6.1 Decision Costs

In the baseline model, we interpreted c as reflecting only implementation costs. Thus, we were implicitly assuming that employees face no decision costs, i.e., the time and cognitive costs of collecting information and thinking about what is the optimal option. Given the plausibility of nontrivial decision costs, it is tempting to reinterpret c as reflecting both implementation and decision costs. However, this is problematic because of the following features of the baseline model. First, all employees staying with a default avoid incurring c. However, this rules out the possibility that some employees incur the decision costs and conclude that staying with the default is a good idea. Second, employees' preferred options are not affected by whether they incurred the decision costs, which seems unrealistic.

To accommodate decision costs, one needs to modify the baseline model. A key obstacle lies in modelling an employee's decision whether to incur the decision costs.

6.2 "Irrational" Opt-Out Costs

Within a standard model, employees' strong reluctance to switch from pension plan default contribution rates implies opt-out costs in the thousands of dollars, which seem excessive.[37] The baseline model with a high α can produce a strong reluctance to switch without a need for excessive opt-out costs.[38] This holds because, for any fixed c, $\lim_{\alpha\to 1}\Delta_L = -\infty$ and $\lim_{\alpha\to 1}\Delta_R = \infty$.

Nevertheless, it is worth considering an alternative to the baseline model in which a part of an employee's opt-out cost is considered "irrational" and isn't counted as a loss in the planner's welfare function. In particular, suppose that the opt-out cost is given by $c = c_R + c_I$ (where $c_R, c_I \geq 0$ are the rational and irrational components, respectively) and let us modify $L(\Delta, \alpha, b, c)$ as follows:

$$L(\Delta, \alpha, b, c_R, c_I) = \begin{cases} |\Delta| & \text{if } \Delta_L \leq \Delta \leq \Delta_R \\ c_R + \alpha|\Delta - \Delta_1| & \text{otherwise} \end{cases}.$$

[37] See section 2.1.7 in DellaVigna (2009) and section IIB in Bernheim et al. (2015).

[38] Similalrly, the anchoring model in Bernheim et al. (2015) can produce a strong reluctance to switch without a need for excessive opt-out costs.

$L(\Delta, \alpha, b, c_R, c_I)$ takes into account that employees base their decision whether to opt out on $c = c_R + c_I$ (Δ_L and Δ_R are as defined before), but counts only c_R as a loss. In the same spirit, the planner considers $c_R - b$, rather than $c - b$, as the total loss from AD.

In this model, $c_I > 0$ plays a similar role to $\beta < 1$ in Carroll et al. (2009). In particular, both $c_I > 0$ in the current model and $\beta < 1$ in Carroll et al. (2009) capture a reluctance to switch from the default that the planner does not consider normatively relevant. Analogously, $c_I > 0$ plays a similar role to $\beta < 1$ in Bernheim et al. (2015) if only the full-commitment frame is considered welfare relevant.

Alternatively, c_I can be viewed as capturing a reluctance to switch due to inattentiveness as in Bernheim et al. (2015) if only the frame without inattentiveness is considered welfare relevant.

6.3 The Presence of Sophisticated Employees

In the baseline model, all employees share the same parameters α, b, and c. Such homogeneity is probably unrealistic. For example, it is natural to assume that some employees are more sophisticated and have a smaller bias (in absolute value). Also, Beshears et al. (2016) present evidence that low-income and young employees, who are plausibly less sophisticated, (i) are less likely to opt out of the default contribution rate, holding fixed the preferred contribution rates, and (ii) are probably more likely to update their preferred contribution rates in light of the default contribution rate. In the context of the baseline model, we can think of these employees as having a higher c and α.

A simple way to incorporate homogeneity along these dimensions is to assume that a fraction, ρ, of the population is sophisticated in the sense that they have zero bias and do not update their preferred options in light of the default. Sophisticated employees can also have their own opt-out cost parameter, c^S, which would be smaller than the cost parameter for unsophisticated employees, c^U. Given that sophisticated employees are less affected by the default in such a model, a natural guess is that the optimal default policy would be geared more towards the interests of unsophisticated employees, at least if ρ is not too high.

6.4 Endogenous Updating

In the baseline model, employees' updating rule is ad hoc. Instead, one could in principle consider a game-theoretic model in which employees use the default to update their beliefs about information that the planner might have and the planner sets the default with this in mind. In such an approach, employees update their beliefs in an endogenous fashion, which reduces the number of free parameters–there would no longer be a need for α.

It is worth noting that such a game-theoretic approach also has some disadvantages. First, one needs to make assumptions about the structure of information, i.e., about what the planner knows and what employees know. Second, in reality employees seem to update their preferred options based on the default even though planners' incentives are not aligned with employees' best interest.[39] This puts a strain on the assumption of equilibrium beliefs and suggests that a simple updating rule may be more realistic.[40] Third, it will likely be much harder to solve the model.

6.5 Delayed Opt-Outs

In the baseline model, employees make a once-and-for-all decision to opt out or not. In reality, many employees opt out eventually even if they don't do so immediately. One could incorporate the possibility of such delayed opt-outs into the baseline model without explicitly going to a dynamic model by modifying $L(\Delta, \alpha, b, c)$ as follows:

$$L(\Delta, \alpha, b, c, \gamma(\cdot)) = \gamma((1-\alpha)(\Delta+b))|\Delta| + \big(1-\gamma((1-\alpha)(\Delta+b))\big)\big(c+\alpha|\Delta-\Delta_1|\big), \quad (5)$$

where $(1-\alpha)(\Delta+b)$ is the difference between the employee's updated preferred option and the default, and $\gamma((1-\alpha)(\Delta+b))$ is the fraction of future work life with the current employer that the employee will spend at the default. This expression assumes that,

[39] For example, many employers benefit from high participation in their pension plans (e.g., because it is associated with reduced reporting requirements). This provides an incentive for setting a low default contribution rate, regardless of what contribution rate is optimal for employees. Also, the fear of lawsuits provides an incentive for employers to set the default asset allocation in a very conservative fashion, again regardless of employees' best interest.

[40] On the other hand, Altmann et al. (2013) provide evidence that, at least in a simple game in the lab, the extent to which people are willing to stay with the default is sensitive to the default-setter's incentives (as well as to the quality of their and the default-setter's information).

if the employee eventually opts out, she opts out to her updated preferred option, which entails a loss of $c + \alpha|\Delta - \Delta_1|$. The function $\gamma(\cdot)$ can be estimated from data on how the size of switches away from the default depends on the length of time since she joined the pension plan.

An alternative approach would be to have a dynamic model in the spirit of Carroll et al. (2009) in which the timing of the decision to opt-out is endogenous. Such a model would eliminate the need for an exogenous $\gamma(\cdot)$. Also, if the opt-out cost is randomly drawn each period, it is likely that employees who are more content with the default (i.e., employees with smaller values of $(1 - \alpha)|\Delta + b|$) would tend to wait for a lower value of the realised opt-out cost before opting out. A dynamic model can capture this while the model based on expression (5) does not. On the other hand, a dynamic model needs to make strong assumptions about how the employee behaves[41] and is likely to be harder to solve.

6.6 Jointly Setting the Default Contribution Rate and the Default Asset Allocation

Let the first and second component of (DP_1, DP_2) denote a default policy regarding the contribution rate and regarding the asset allocation, respectively. The baseline model can be applied separately for determining the optimal DP_1 and for determining the optimal DP_2. However, this neglects the fact that the cost of opting out of a default contribution rate is likely to be much smaller if the employee is simultaneously opting out of a default asset allocation (and vice versa). Thus, it might be necessary to think about how to optimally set DP_1 and DP_2 jointly.

6.7 A More Realistic Model

Any model that is simple enough to analyse is likely to be unrealistic in many ways. Therefore, it might be necessary to build a more realistic model combining many of the variations on the baseline model considered above as well as other features, such as a nonuniform distribution of optimal options, a more realistic updating rule,[42] or

[41] E.g., how does she discount the future? Does she correctly anticipate her own future behaviour?

[42] For example, we could assume that α, rather than being a constant, is a function of $|x + b - D|$. E.g., $\alpha(|x + b - D|) = \gamma_1 e^{-\gamma_2 |x+b-D|}$, where $\gamma_1 \in [0, 1]$ is a "level" parameter and $\gamma_2 \in [0, \infty)$ is a

(possibly) a multiplicative bias. Such a model would also have to incorporate a more realistic function capturing the loss from deviations, ideally one that is derived from a life-cycle model of consumption and takes into account features such as employer-match and tax-bracket thresholds. Although such a model will surely be too complex to analyse theoretically, it should not be too hard to solve numerically if one can take a stand on the relevant parameters, ideally based on empirical work as discussed in section 8.

7 Complexity

One of the insights from this paper is just how complicated the analysis of optimal default policies is. The baseline model was built to be as simple as possible. Yet, its analysis is still complicated in many ways. First, the optimal default is one of six qualitatively different kinds of defaults.

Second, I have been unable to find intuitions for the statements in Lemma 1, which is about the graph of $L(\cdot, \alpha, b, c)$. This is important because the proof of Proposition 1 relies heavily on the geometry of this graph. In a similar vein, I have been unable to find intuitions for the shapes and positions of many of the regions in Figure 9 or for how they shift as α or ϵ change. Although explanations were provided for some of the observed patterns, many of these explanations relied either on the geometry of the graph of $L(\cdot, \alpha, b, c)$, on Proposition 1, on Proposition 2 (whose proof is opaque, as explained below), or on Proposition 3 (whose proof relies in part on Proposition 1).

Third, the partial derivatives of Ψ with respect to each parameter are very complex (see Lemma 2 in the appendix). The complexity arises because, when we change a parameter, (i) $L(\cdot, \alpha, b, c)$ shifts in complicated ways and (ii) the limits of integration under $L(\cdot, \alpha, b, c)$, $\bar{\Delta}(\alpha, b, c, \epsilon)$ and $\bar{\Delta}(\alpha, b, c, \epsilon) - 2\epsilon$ also shift.[43]

Fourth, the proof of Proposition 2 is based largely on Lemma 3 in the appendix, which puts bounds on the partial derivatives of Ψ. Unfortunately, there is no apparent intuition for this lemma, and hence for Proposition 2, largely because of the complexity of the partial derivatives of Ψ.

parameter governing how quickly the employee's weight on the default declines.

[43]The shift in $\bar{\Delta}(\alpha, b, c, \epsilon)$ can often, but not always, be ignored based on envelope-theorem logic.

Fifth, we saw that the total loss at the optimal default or at an optimal default policy can be nonmonotone in each parameter in counterintuitive ways. This, too, is a consequence of the complex (often, opposing) forces governing the partial derivatives of Ψ.

Most of the variations of the baseline model considered in section 6 are likely to be even more complicated.

8 Concluding Remarks

The current paper attempts to fill a gap in the existing literature by analysing a model in which employees are biased in their perception of their optimal options. Within this model, we have gained several insights. First, we have shown that six qualitatively different kinds of defaults can be optimal. Second, we have observed an adverse effect of defaults on some employees' preferred options that can make AD the optimal policy. Third, we have discussed some patterns in the relationship between the parameters and the optimal default policy. Fourth, we have shown that the total loss at the optimal default and at the optimal default policy can be nonmonotone in the parameters in counterintuitive ways. Fifth, we have learned that the analysis of even this apparently simple model can be very complex.

Having said all this, the baseline model lacks realism in many ways. To be able to provide practical guidance to default-setters in the real world, two main challenges remain. First, we need to analyse variations on the baseline model along the lines discussed in section 6.

Second, much empirical work is needed to estimate parameters corresponding to the size of the bias, the strength of updating from the default,[44] implementation costs, decision costs, (possibly) irrational costs, the degree of heterogeneity in terms of optimal options, the fraction of sophisticated employees, the $\gamma(\cdot)$ function from section 6.5 on delayed opt-outs, and any parameters in a life cycle model from which a more realistic loss from deviations can be derived. Deciding what value to use for the bias in the models is likely to be especially hard and controversial. For example, there is no consensus in the existing literature on whether people undersave (see

[44]Beshears et al. (2016) estimate employees' preferred contribution rates for different defaults. This could allow us to pin down the strength of updating.

footnote 5). The question of whether people underweight stocks in their pension plan portfolios also hinges on complex issues, such as whether there is an equity premium puzzle, what is a reasonable expected return on the stock market going forward, and individuals' levels of risk aversion.

References

Altmann, Steffen, Armin Falk and Andreas Grunewald. 2013. "Incentives and Information as Driving Forces of Default Effects." IZA Discussion Paper No. 7610

Bach, Laurent, Laurent E. Calvet, and Paolo Sodini. 2016. "Rich pickings? Risk, return, and skill in the portfolios of the wealthy." working paper.

Benartzi, Shlomo, and Richard H. Thaler. 1999. "Risk aversion or myopia? Choices in repeated gambles and retirement investments." *Management Science* 45.3: 364-381.

Bernheim, B. Douglas. 2009. "Behavioral Welfare Economics." *Journal of the European Economic Association* 7 (2-3): 267319.

Bernheim, B. Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *Quarterly Journal of Economics* 124 (1): 51104.

Bernheim, B. Douglas, Andrey Fradkin, and Igor Popov. 2015. "The welfare economics of default options in 401 (k) plans." *American Economic Review* 105.9: 2798-2837.

Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian. 2008. "The Importance of Default Options for Retirement Saving Outcomes: Evidence from the United States." In: Kay SJ, Sinha T Lessons from Pension Reform in the Americas. Oxford University Press; p. 59-87.

Beshears John, James J. Choi, David Laibson, Brigitte C. Madrian, and Sean (Yixiang) Wang. 2016. "Who Is Easier to Nudge?" working paper.

Beshears J, J.Choi J, Laibson D, Madrian BC. 2017. "Does Aggregated Returns Disclosure Increase Portfolio Risk-Taking?" *Review of Financial Studies* 30(6):1971-2005.

Bronchetti, Erin Todd, Thomas S. Dee, David B. Huffman, and Ellen Magenheim. 2013. "When a Nudge Isnt Enough: Defaults and Saving among Low-Income Tax Filers." *National Tax Journal* 66 (3): 60934.

Calvet, Laurent E., John Y. Campbell, and Paolo Sodini. 2007. "Down or Out: Assessing the Welfare Costs of Household Investment Mistakes." *Journal of Political Economy* 115.5: 707-747.

Campbell John Y. 2006. "Household Finance, Presidential Address to the American Finance Association." *Journal of Finance* LXI (4):1553-1604.

Campbell John Y. 2016. "Restoring Rational Choice: The Challenge of Consumer Financial Regulation." *American Economic Review: Papers & Proceedings* [Internet] 106 (5): 1-30.

Caplin, Andrew, and Daniel Martin. 2017. "Defaults and Attention: The Drop Out Effect." *Revue conomique* 5: 45-54.

Carlin, Bruce Ian, Simon Gervais, and Gustavo Manso. 2013. "Libertarian paternalism, information production, and financial decision making." *The Review of Financial Studies* 26.9: 2204-2228.

Carroll, Gabriel D., James J. Choi, David Laibson, Brigitte C. Madrian, and Andrew Metrick. 2009. "Optimal Defaults and Active Decisions." *Quarterly Journal of Eco-*

nomics 124 (4): 163974.

Chetty, Raj, John N. Friedman, Sren Leth-Petersen, Torben Heien Nielsen, and Tore Olsen. 2014. "Active vs. Passive Decisions and Crowd-Out in Retirement Savings." *Quarterly Journal of Economics* 129.3: 1141-1219

Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick. 2002. "Defined Contribution Pensions: Plan Rules, Participant Choices, and the Path of Least Resistance." In *Tax Policy and the Economy*, edited by James M. Poterba, 67113. Cambridge, MA: MIT Press.

Choi, James, David Laibson, Brigitte Madrian, and Andrew Metrick. 2003. "Passive Decisions and Potent Defaults. National Bureau of Economic Research Working Paper 9917.

Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick. 2004a. "For Better or Worse: Default Effects and 401(k) Savings Behavior." *In Perspectives on the Economics of Aging*, edited by David A. Wise, 81126. Chicago: University of Chicago Press.

Choi, James, David Laibson, Brigitte C. Madrian, and Andrew Metrick. 2004b. "Employees Investment Decisions About Company Stock." In *Pension Design and Structure: New Lessons from Behavioral Finance*, edited by Olivia Mitchell and Stephen Utkus. Oxford: Oxford University Press.

Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick. 2006. "Saving for Retirement on the Path of Least Resistance." In *Behavioral Public Finance: Toward a New Agenda*, edited by Edward J. McCaffery and Joel Slemrod, 30451. New York: Russell Sage Foundation.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature* 47 (2): 31572.

Eisenstein, Eric M., and Stephen J. Hoch. 2005. "Intuitive compounding: framing, temporal perspective, and expertise." Manuscript, Temple University.

Grinblatt, Mark, Matti Keloharju, and Juhani Linnainmaa. 2011. "IQ and stock market participation." *The Journal of Finance* 66.6: 2121-2164.

Laibson, David. 1997. "Golden eggs and hyperbolic discounting." *The Quarterly Journal of Economics* 112.2 (1997): 443-478.

Levy, Matthew, and Joshua Tasoff. 2016. "Exponential-Growth Bias and Lifecycle Consumption." *Journal of the European Economic Association* 14.3

Madrian, Brigitte C., and Dennis F. Shea. 2001. "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior." *Quarterly Journal of Economics* 116 (4): 114987.

O'Donoghue, Ted, and Matthew Rabin. 1999. "Doing it now or later." *American Economic Review* 89.1: 103-124.

Phelps, Edmund S., and Robert A. Pollak. 1968. "On Second-Best National Saving and Game-Equilibrium Growth." *Review of Economic Studies* XXXV, 18599.

Poterba, James M. 2015. "Saver heterogeneity and the challenge of assessing retirement saving adequacy." *National Tax Journal* 68.2: 377.

Stango, Victor, and Jonathan Zinman. 2009. "Exponential Growth Bias and Household Finance." The Journal of Finance 64.6: 2807-2849.

Thaler, Richard H, and Cass R Sunstein. 2003. "Libertarian Paternalism." American Economic Review Papers and Proceedings 93 (2): 175-179.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185 (4157): 1124-1131.

Wisson, James. 2016. "Optimal Defaults and Uncertainty." Working paper.

9 Appendix: Examples Illustrating the Nonmonotonicity of the Total Loss

As noted in section 5, for each parameter α, b, c, or ϵ, the total loss at the optimal default, $\frac{1}{2\epsilon}\Psi(\alpha, b, c, \epsilon)$, need not be monotone in it. This is illustrated in Figure 10.

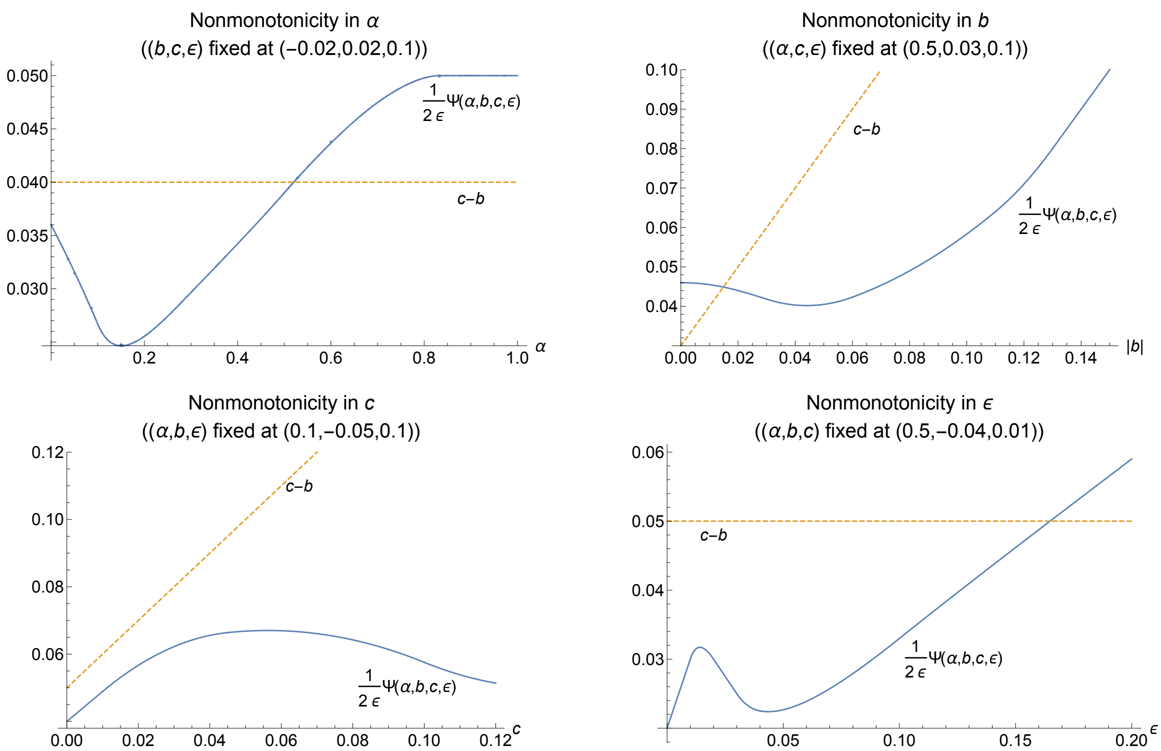

Figure 10: Nonmonotonicity of the total loss at the optimal default (solid line) and at the optimal default policy (lower envelope of solid and dashed lines).

For each parameter α, b, c, or ϵ, the total loss at the optimal default policy, $\min\left(\frac{1}{2\epsilon}\Psi(\alpha, b, c, \epsilon), c - b\right)$, need not be monotone in it either. In each panel of Figure

10, the total loss at an optimal default policy is given by the lower envelope of the solid and dashed lines. This lower envelope is clearly nonmonotone in each panel.

10 Appendix: Partial Derivatives of Ψ

Let Ψ_p denote the partial derivative of Ψ with respect to the parameter $p \in \{\alpha, b, c, \epsilon\}$. Let $-p$ denote the three parameters other than p and, with the usual abuse of notation, we will write $\bar{\Delta}(p, -p)$, $\Psi(p, -p)$, and $\Psi_p(p, -p)$.

In this section, we study $\Psi_\alpha, \Psi_b, \Psi_c$, and Ψ_ϵ. We do this for two reasons. First, Lemma 3 at the end of this section will play a crucial role in the proof of Proposition 2. Second, the complexity of these partial derivatives sheds some light on the nonmonotonicity of total losses discussed in section 5.

Let us start with the following result.

Lemma 2 *Letting $p \in \{\alpha, b, c\}$ and fixing $-p$, for almost all p, $\Psi_p(p, -p)$ exists and*[45]

$$\Psi_p(p,-p) =$$
$$\begin{cases} 0 & \text{if } (\alpha,b,c,\epsilon) \in R_{AS\text{-}C} \\ (\Delta_L + 2\epsilon - |\Delta_L|)\frac{\partial \Delta_L}{\partial p} & \text{if } (\alpha,b,c,\epsilon) \in R_{AS\text{-}L} \\ \big(c + \alpha|\Delta_L - \Delta_1| - |\Delta_L|\big)\frac{\partial \Delta_L}{\partial p} + & \\ \int_{(-\infty,\Delta_1]\cap[\Delta_L,\Delta_R]'\cap[\bar{\Delta}(\alpha,b,c,\epsilon)-2\epsilon,\bar{\Delta}(\alpha,b,c,\epsilon)]} \frac{\partial(c-\alpha(\Delta-\Delta_1))}{\partial p} d\Delta + & \\ \int_{[\Delta_1,\infty)\cap[\Delta_L,\Delta_R]'\cap[\bar{\Delta}(\alpha,b,c,\epsilon)-2\epsilon,\bar{\Delta}(\alpha,b,c,\epsilon)]} \frac{\partial(c+\alpha(\Delta-\Delta_1))}{\partial p} d\Delta & \text{if } (\alpha,b,c,\epsilon) \in \\ & R_{HS1} \cup R_{HS2} \cup R_{MS\text{-}HI} \\ \big(c + \alpha(\Delta_L - \Delta_1) - L(\Delta_L - 2\epsilon, \alpha, b, c)\big)\frac{\partial \Delta_L}{\partial p} + & \\ \int_{\min(\Delta_1,\Delta_L-2\epsilon)}^{\Delta_1} \frac{\partial(c-\alpha(\Delta-\Delta_1))}{\partial p} d\Delta + & \\ \int_{\max(\Delta_1,\Delta_L-2\epsilon)}^{\Delta_L} \frac{\partial(c+\alpha(\Delta-\Delta_1))}{\partial p} d\Delta & \text{if } (\alpha,b,c,\epsilon) \in R_{NS\text{-}corner} \end{cases} . \tag{6}$$

Fixing (α, b, c), for almost all ϵ, $\Psi_\epsilon(\alpha, b, c, \epsilon)$ exists and

[45] In the expression below, $[\Delta_L, \Delta_R]'$ denotes the complement of $[\Delta_L, \Delta_R]$.

$$\Psi_\epsilon(\alpha,b,c,\epsilon) = \begin{cases} 2L(\bar{\Delta}(\alpha,b,c,\epsilon),\alpha,b,c) & \textit{if } (\alpha,b,c,\epsilon) \notin R_{\textit{NS-corner}} \\ 2L(\Delta_L - 2\epsilon,\alpha,b,c) & \textit{if } (\alpha,b,c,\epsilon) \in R_{\textit{NS-corner}} \end{cases}. \tag{7}$$

For $p \in \{\alpha, b, c, \epsilon\}$, the formulas for Ψ_p in the lemma follow from (i) the Leibniz intergal rule, (ii) the fact that, for $\Delta \in (\Delta_L, \Delta_R)$, $\frac{\partial L(\Delta,\alpha,b,c)}{\partial p} = \frac{\partial|\Delta|}{\partial p} = 0$, and, in some cases, (iii) envelope-theorem-style considerations which allow us to ignore the dependence of $\bar{\Delta}(\alpha,b,c,\epsilon)$ on p when computing $\Psi_p(\alpha,b,c,\epsilon)$.

The lemma illustrates the complex mechanics governing Ψ_p ($p \in \{\alpha, b, c, \epsilon\}$), mechanics that differ in different regions of the parameter space. To appreciate this complexity, let us consider the geometric intuition behind the different cases in expressions (6) and (7).

Geometric intuition for the first case in expression (6):

This case corresponds to the first case in expression (3) and, hence, to $\bar{\Delta}(\alpha,b,c,\epsilon)$ illustrated in Figure 2. The area of the two shaded triangles under $L(\cdot,\alpha,b,c)$ in the figure does not depend on $p \in \{\alpha, b, c\}$, so that $\Psi_p = 0$.

Geometric intuition for the second case in expression (6):

This case corresponds to the second case in expression (3) and, hence, to $\bar{\Delta}(\alpha,b,c,\epsilon)$ illustrated in Figure 3. As p increases in this figure, both limits of integration in the figure, Δ_L and $\Delta_L + 2\epsilon$, increase at rate $\frac{\partial \Delta_L}{\partial p}$. As the right limit of integration increases, the shaded area under $L(\cdot,\alpha,b,c)$ increases at rate $\Delta_L + 2\epsilon$. As the left limit of integration increases, the shaded area under $L(\cdot,\alpha,b,c)$ decreases at rate $|\Delta_L|$. The upshot is the second piece in expression (6).

Geometric intuition for the third case in expression (6):

This case corresponds to Figures 4-6. Three things happen as p increases in these figures.[46]

First, Δ_L increases at rate $\frac{\partial \Delta_L}{\partial p}$. Given that $\Delta_L \in (\bar{\Delta}(\alpha,b,c,\epsilon) - 2\epsilon, \bar{\Delta}(\alpha,b,c,\epsilon))$,[47] an increase in Δ_L causes the shaded area under $L(\cdot,\alpha,b,c)$ in the figures to increase

[46] In this case, envelope-theorem-style considerations allow us to ignore the dependence of $\bar{\Delta}(\alpha,b,c,\epsilon)$ on p.

[47] This can be seen graphically in Figures 4-6. It is not hard to see this more formally by studying the inequalities in the third, fourth, and fifth cases in expression (3).

at rate $\lim_{\Delta\uparrow\Delta_L} L(\Delta,\alpha,b,c) - |\Delta_L| = c + \alpha|\Delta_L - \Delta_1| - |\Delta_L|$, which is the absolute size of the jump of $L(\cdot,\alpha,b,c)$ at Δ_L. This is reflected in the term before the two integrals.

Second, the leftmost segment of $L(\cdot,\alpha,b,c)$, which equals $c - \alpha(\Delta - \Delta_1)$, shifts up at rate $\frac{\partial(c-\alpha(\Delta-\Delta_1))}{\partial p}$. We are interested in the upward shift of $c - \alpha(\Delta - \Delta_1)$ only for values of Δ that are both less than Δ_1 and outside $[\Delta_L, \Delta_R]$ (which is where $L(\Delta,\alpha,b,c) = c - \alpha(\Delta - \Delta_1)$) and are, of course, also in $[\bar{\Delta}(\alpha,b,c,\epsilon) - 2\epsilon, \bar{\Delta}(\alpha,b,c,\epsilon)]$. This explains the domain of integration for the first integral.

Third, the part of $L(\cdot,\alpha,b,c)$, which equals $c + \alpha(\Delta - \Delta_1)$, shifts up at rate $\frac{\partial(c+\alpha(\Delta-\Delta_1))}{\partial p}$. We are interested in the upward shift of $c + \alpha(\Delta - \Delta_1)$ only for values of Δ that are both larger than Δ_1 and outside $[\Delta_L, \Delta_R]$ (which is where $L(\Delta,\alpha,b,c) = c + \alpha(\Delta - \Delta_1)$) and are, of course, also in $[\bar{\Delta}(\alpha,b,c,\epsilon) - 2\epsilon, \bar{\Delta}(\alpha,b,c,\epsilon)]$. This explains the domain of integration for the second integral.

Geometric intuition for the fourth case in expression (6):

This case corresponds to the last case in expression (3) and, hence, to $\bar{\Delta}(\alpha,b,c,\epsilon)$ illustrated in Figure 7. Three things happen as p increases in this figure.

First, both limits of integration in the figure, $\Delta_L - 2\epsilon$ and Δ_L, increase at rate $\frac{\partial\Delta_L}{\partial p}$. As the right limit of integration increases, the shaded area under $L(\cdot,\alpha,b,c)$ increases at rate $c + \alpha(\Delta_L - \Delta_1)$. As the left limit of integration increases, the shaded area under $L(\cdot,\alpha,b,c)$ decreases at rate $L(\Delta_L - 2\epsilon,\alpha,b,c)$. This is reflected in the term before the two integrals in the last piece in expression (6).

Second, the leftmost segment of $L(\cdot,\alpha,b,c)$ which equals $c - \alpha(\Delta - \Delta_1)$ shifts up at rate $\frac{\partial(c-\alpha(\Delta-\Delta_1))}{\partial p}$. We are interested in the upward shift of $c - \alpha(\Delta - \Delta_1)$ only if the lower limit of integration in Figure 7 (i.e., $\Delta_L - 2\epsilon$) is below Δ_1 (which is the case in the right panel) and, then, only for values between $\Delta_L - 2\epsilon$ and Δ_1. This explains the domain of integration for the first integral.

Third, the part of $L(\cdot,\alpha,b,c)$ which equals $c + \alpha(\Delta - \Delta_1)$ shifts up at rate $\frac{\partial(c+\alpha(\Delta-\Delta_1))}{\partial p}$. We are interested in the upward shift of $c + \alpha(\Delta - \Delta_1)$ only for values of Δ either between Δ_1 and Δ_L (in the case illustrated in the right panel of Figure 7) or between $\Delta_L - 2\epsilon$ and Δ_L (in the case illustrated in the left panel of Figure 7). This explains the domain of integration for the second integral.

Geometric intuition for the first case in expression (7):

This case corresponds to Figures 2-6. The mechanics governing Ψ_ϵ in this case depend on whether $(\alpha, b, c, \epsilon) \in R_{\text{AS-L}}$.

If $(\alpha, b, c, \epsilon) \in R_{\text{AS-L}}$, increasing ϵ shifts the right limit of integration in Figure 3 to the right at rate $\frac{d(2\epsilon)}{d\epsilon} = 2$. Thus, the shaded area under $L(\cdot, \alpha, b, c)$ in the figure increases at rate $2L(\bar{\Delta}(\alpha, b, c, \epsilon), \alpha, b, c)$.

If $(\alpha, b, c, \epsilon) \in R_{\text{AS-C}}$, R_{HS1}, R_{HS2}, or $R_{\text{MS-HI}}$, increasing ϵ shifts the left/right limit of integration in Figures 2, 4, 5, and 6 to the left/right at rate 1. As a result, the shaded area under $L(\cdot, \alpha, b, c)$ in these figures increases at rate $L(\bar{\Delta}(\alpha, b, c, \epsilon) - 2\epsilon, \alpha, b, c) + L(\bar{\Delta}(\alpha, b, c, \epsilon), \alpha, b, c) = 2L(\bar{\Delta}(\alpha, b, c, \epsilon), \alpha, b, c)$.

Geometric intuition for the second case in expression (7):

This case corresponds to the last case in expression (3) and, hence, to $\bar{\Delta}(\alpha, b, c, \epsilon)$ illustrated in Figure 7. As ϵ increases in this figure, the only thing that happens is that the left limit of intergration shifts left at rate $\frac{d(2\epsilon)}{\epsilon} = 2$. As a result the shaded area under $L(\cdot, \alpha, b, c)$ expands at rate $2L(\Delta_L - 2\epsilon, \alpha, b, c)$.

Despite the complex mechanics governing the partial derivatives of Ψ, we can nevertheless put bounds on Ψ_α, Ψ_b, and Ψ_c (in some regions of the parameter space).

Lemma 3 *Each of the following statements holds whenever the derivative of Ψ involved in the statement exists.*

1) $b > -\frac{\alpha\epsilon}{2}$ implies $\Psi_\alpha(\alpha, b, c, \epsilon) \geq 0$.

2) $b > -\frac{\alpha\epsilon}{2}$ implies $\Psi_b(\alpha, b, c, \epsilon) > -2\epsilon$.

3) $\Psi_c(\alpha, b, c, \epsilon) \leq 2\epsilon$, the inequality being strict whenever $c > 0$.

Unfortunately, there is no apparent intuition for the lemma. The proof of each part goes through different regions of the parameter space and, for each region, uses the relevant piece in expression (6). In some regions, there is a geometric intuition for why the relevant piece in expression (6) satisfies the given inequality. In other regions, the different terms in the relevant piece in expression (6) have different signs and only the algebra reveals that their relative magnitudes are such that the given inequality holds.

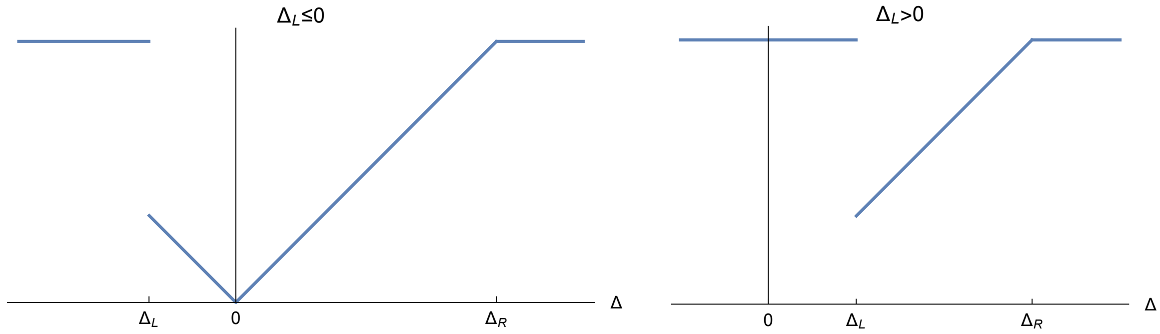

Figure 11: Graph of $L(\cdot, \alpha, b, c)$ when $\alpha = 0$.

11 Appendix: The Cases with $\alpha = 1$ and $\alpha = 0$

Let us start with the case $\alpha = 1$. In this case, $L(\Delta, \alpha, b, c) = |\Delta|$. The following proposition characterises the optimal default and the optimal default policy for this case.

Proposition 4 *Assume $\alpha = 1$. The unique solution to problem (2) is $\bar{\Delta} = \epsilon$. The set of optimal default policies is* $\begin{cases} \{\bar{\Delta} = \epsilon\} & \text{if } \frac{\epsilon}{2} < c - b \\ \{\bar{\Delta} = \epsilon, AD\} & \text{if } \frac{\epsilon}{2} = c - b \\ \{AD\} & \text{if } \frac{\epsilon}{2} > c - b \end{cases}$.

The validity of the proposition should be obvious given (i) the shape of the graph of $L(\cdot, \alpha, b, c)$ when $\alpha = 1$ and (ii) the fact that $\frac{\epsilon}{2}$ equals the total loss associated with the AS-C default.

Let us turn to the case when $\alpha = 0$. In this case,

$$L(\Delta, \alpha, b, c) = \begin{cases} |\Delta| & \text{if } \Delta_L \le \Delta \le \Delta_R \\ c - b & \text{otherwise} \end{cases}.$$

The left/right panel in Figure 11 depicts the graph of $L(\cdot, \alpha, b, c)$ for the case when $\Delta_L \le 0/\Delta_L > 0$.

We can state the following.

Proposition 5 *Assume $\alpha = 0$. The set of solutions to problem (2) is*

$$\begin{cases} \{\bar{\Delta} = \epsilon\} & \textit{if } \Delta_L \leq -\epsilon \\ \{\bar{\Delta} = \Delta_L + 2\epsilon\} & \textit{if } \Delta_L > -\epsilon, \Delta_R - \Delta_L \geq 2\epsilon \\ \{\bar{\Delta} \in [\Delta_R, \Delta_L + 2\epsilon]\} & \textit{if } 0 \leq \Delta_R - \Delta_L < 2\epsilon \end{cases} \tag{8}$$

$$= \begin{cases} \{\bar{\Delta} = \epsilon\} & \textit{if } c \geq \epsilon - b \\ \{\bar{\Delta} = \Delta_L + 2\epsilon\} & \textit{if } c < \epsilon - b, c \geq \epsilon \\ \{\bar{\Delta} \in [\Delta_R, \Delta_L + 2\epsilon]\} & \textit{if } 0 \leq c < \epsilon \end{cases} . \tag{9}$$

An optimal default is always an optimal default policy. AD is an optimal default policy if and only if $\Delta_R = \Delta_L$ (which can equivalently be written as $c = 0$).

That expression (8) gives the optimal defaults follows from considering, for each of the four cases in expression (8), how to place the interval $[\bar{\Delta} - 2\epsilon, \bar{\Delta}]$ in each of the panels of Figure 11 in order to minimise the area under the curve.[48] Note that any $\bar{\Delta}$ given by expression (8) satisfies the constraints in problem (2). In particular, $\bar{\Delta} \geq 0 \geq r + 2\epsilon - 1$; $|b| \leq r$ implies $\Delta_L \leq r$, which implies $\bar{\Delta} \leq r + 2\epsilon$; also, clearly, $\Delta_L \leq \bar{\Delta} \leq \Delta_R + 2\epsilon$.

Expression (9) merely rewrites the conditions in each case in expression (8) in terms of the underlying parameters. Given that $\Delta_R - \Delta_L < 2\epsilon$ implies $\Delta_L > -\epsilon$, it is obvious that the four cases in expression (8) cover the whole parameter space, so that each of expressions (8) and (9) completely characterises the optimal default as a function of the parameters.

Whenever $\Delta_R > \Delta_L$, any optimal default has the property that a positive measure of employees stay with the default and experience a loss strictly less than $c - b$ whilst any employees opting out experience a loss of $c - b$, so that any optimal default is clearly better than AD. When $\Delta_R = \Delta_L$, all employees, except those with $\Delta = -b$ (who are of measure zero), opt out of any default and experience a loss of $c - b$. Thus, both AD as well as any $\bar{\Delta}$ satisfying the constraints in problem (2) is an optimal default policy.[49]

[48] Only the left panel applies in the first case in expression (8).

[49] When $\Delta_L = \Delta_R$, $[\Delta_R, \Delta_L + 2\epsilon] = [-b, -b + 2\epsilon] \subseteq [r + 2\epsilon - 1, r + 2\epsilon]$. The set inclusion follows because (i) $-b \geq 0 \geq r + 2\epsilon - 1$ and (ii) $|b| \leq r$ implies $-b + 2\epsilon \leq r + 2\epsilon$. Thus, we can ignore the constraint $r + 2\epsilon - 1 \leq \bar{\Delta} \leq r + 2\epsilon$.

12 Appendix: Proofs and Further Results

Some of the proofs will make use of Figure 12. This figure is similar to Figure 9 except that (i) it does not depict $R_{AD}(\alpha, \epsilon)$, (ii) it applies to any values of $\alpha \leq 0.5$ and ϵ, and (iii) it shows the lines corresponding to the equations $\Delta_L = 0$ (lower dotted line) and $\Delta_L = \Delta_1$ (upper dotted line). In the region below the lower dotted line $\Delta_L > 0$; in the region above the upper dotted line $\Delta_L < \Delta_1$; in the region between the two dotted lines $\Delta_1 < \Delta_L < 0$. If $\alpha \geq 0.5$, the figure would be similar except that points with height $\alpha\epsilon$ would lie above points with height $(1-\alpha)\epsilon$.

12.1 Proof of Lemma 1

Proof of Statement 1):

$$\lim_{\Delta \downarrow \Delta_R} L(\Delta, \alpha, b, c) = c + \alpha(\Delta_R - \Delta_1) = \Delta_R = L(\Delta_R)$$

Q.E.D.

Proof of Statement 2):

When $\Delta_L \leq \Delta_1$,

$$\lim_{\Delta \uparrow \Delta_L} L(\Delta, \alpha, b, c) = c - \alpha(\Delta_L - \Delta_1) = -\Delta_L = L(\Delta_L)$$

Q.E.D.

Proof of Statement 3):

When $\Delta_1 < \Delta_L \leq 0$,

$$L(\Delta_L, \alpha, b, c) = -\Delta_L = c - \alpha(\Delta_L - \Delta_1) < c < c + \alpha(\Delta_L - \Delta_1) = \lim_{\Delta \uparrow \Delta_L} L(\Delta, \alpha, b, c)$$

Q.E.D.

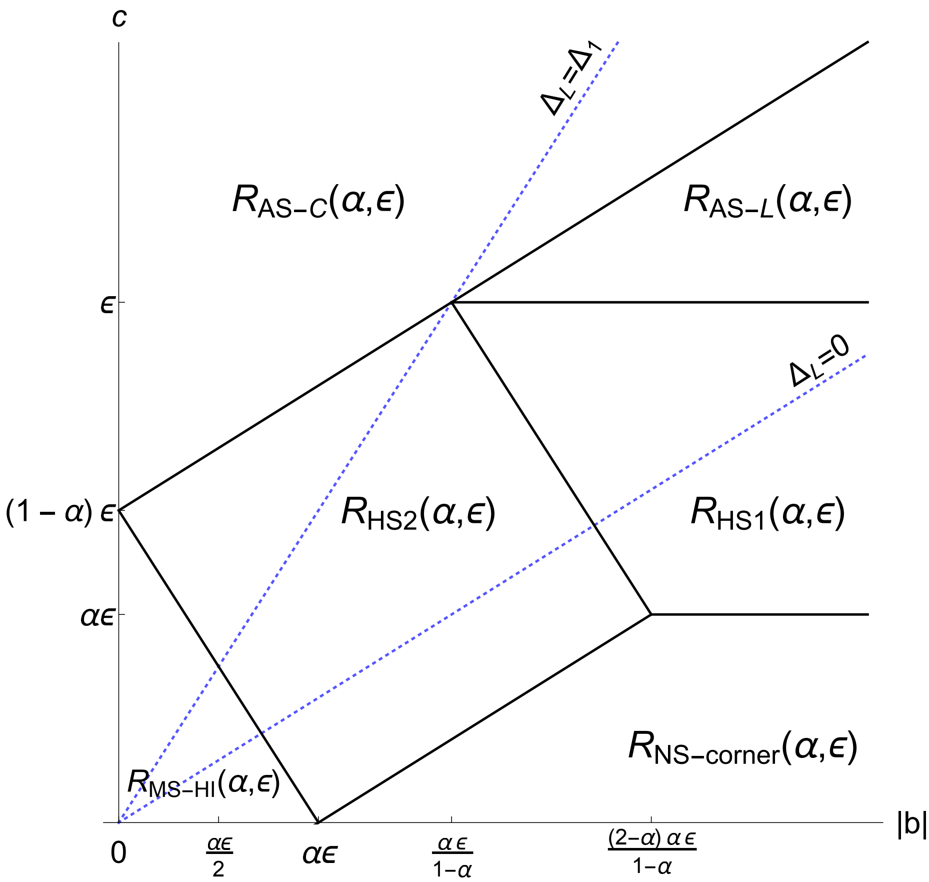

Figure 12: $R_{\text{AS-C}}(\alpha,\epsilon)$, $R_{\text{AS-L}}(\alpha,\epsilon)$, $R_{\text{HS1}}(\alpha,\epsilon)$, $R_{\text{HS2}}(\alpha,\epsilon)$, $R_{\text{MS-HI}}(\alpha,\epsilon)$, and $R_{\text{NS-corner}}(\alpha,\epsilon)$. These regions should be read off the figure ignoring the dotted lines. In the region below the lower dotted line $\Delta_L > 0$; in the region above the upper dotted line $\Delta_L < \Delta_1$; in the region between the two dotted lines $\Delta_1 < \Delta_L < 0$. The figure assumes $\alpha < 0.5$. If $\alpha \geq 0.5$, the figure would be similar except that points with height $\alpha\epsilon$ would lie above points with height $(1-\alpha)\epsilon$.

Proof of Statement 4):

When $\Delta_L > 0$,

$$\lim_{\Delta\uparrow\Delta_L} L(\Delta,\alpha,b,c) - L(\Delta_L,\alpha,b,c) = c + \alpha(\Delta_L - \Delta_1) - \Delta_L = 2c \geq 0$$

The last inequality is strict if and only if $c > 0$. Q.E.D.

12.2 Proof of Proposition 1

Let us start by showing that the inequalities in each case in expression (3) are equivalent to the inequalities in the corresponding case in expression (4). For cases in expression (3) in which no inequalities involve the function $L(\cdot, \alpha, b, c)$ (i.e., for the first, third, fourth, and fifth cases) this is trivial–one only needs to write out the variables in the inequalities in terms of the underlying parameters. We consider the remaining cases in the following claims.

Claim 1 $L(\Delta_L, \alpha, b, c) < L(\Delta_L + 2\epsilon, \alpha, b, c) \leq \lim_{\Delta \uparrow \Delta_L} L(\Delta, \alpha, b, c)$ *if and only if* $c < (1-\alpha)\epsilon - (1-\alpha)b, c \geq \epsilon$.

Proof:

$$\left| \begin{array}{l} \lim_{\Delta \uparrow \Delta_L} L(\Delta, \alpha, b, c) \geq L(\Delta_L + 2\epsilon, \alpha, b, c) \\ L(\Delta_L) < L(\Delta_L + 2\epsilon, \alpha, b, c) \end{array} \right. \iff$$

$$\left| \begin{array}{l} c + \alpha|\Delta_L - \Delta_1| \geq L(\Delta_L + 2\epsilon, \alpha, b, c) \\ L(\Delta_L) < L(\Delta_L + 2\epsilon, \alpha, b, c) \end{array} \right. \iff$$

$$\left| \begin{array}{l} c + \alpha|\Delta_L - \Delta_1| \geq \Delta_L + 2\epsilon \\ L(\Delta_L) < \Delta_L + 2\epsilon \end{array} \right. \iff$$

$$\left(\left| \begin{array}{l} \Delta_L < \Delta_1 \\ c - \alpha(\Delta_L - \Delta_1) \geq \Delta_L + 2\epsilon \\ -\Delta_L < \Delta_L + 2\epsilon \end{array} \right. \text{ or } \left| \begin{array}{l} \Delta_L \geq \Delta_1 \\ \Delta_L \leq 0 \\ c + \alpha(\Delta_L - \Delta_1) \geq \Delta_L + 2\epsilon \\ -\Delta_L < \Delta_L + 2\epsilon \end{array} \right. \text{ or } \right.$$

$$\left. \left| \begin{array}{l} \Delta_L \geq \Delta_1 \\ \Delta_L > 0 \\ c + \alpha(\Delta_L - \Delta_1) \geq \Delta_L + 2\epsilon \\ \Delta_L < \Delta_L + 2\epsilon \end{array} \right. \right) \iff$$

$$\left(\left| \begin{array}{l} \Delta_L < \Delta_1 \\ \Delta_L \leq -\epsilon \\ \Delta_L > -\epsilon \end{array} \right. \text{ or } \left| \begin{array}{l} \Delta_L \geq \Delta_1 \\ \Delta_L \leq 0 \\ c \geq \epsilon \\ \Delta_L > -\epsilon \end{array} \right. \text{ or } \left| \begin{array}{l} \Delta_L \geq \Delta_1 \\ \Delta_L > 0 \\ c \geq \epsilon \end{array} \right. \right) \iff$$

$$\left(\left|\begin{array}{l}\Delta_L \geq \Delta_1\\ \Delta_L \leq 0\\ c \geq \epsilon\\ \Delta_L > -\epsilon\end{array}\right. \text{ or } \left|\begin{array}{l}\Delta_L \geq \Delta_1\\ \Delta_L > 0\\ c \geq \epsilon\\ \Delta_L > -\epsilon\end{array}\right.\right) \iff$$

$$\left|\begin{array}{l}\Delta_L \geq \Delta_1\\ c \geq \epsilon\\ \Delta_L > -\epsilon\end{array}\right. \iff \left|\begin{array}{l}c \leq -\frac{1-\alpha}{\alpha}b\\ c \geq \epsilon\\ c < (1-\alpha)\epsilon - (1-\alpha)b\end{array}\right. \iff \left|\begin{array}{l}c \geq \epsilon\\ c < (1-\alpha)\epsilon - (1-\alpha)b\end{array}\right.$$

The second " $\iff$ " holds because $c + \alpha|\Delta_L - \Delta_1| \geq L(\Delta_L + 2\epsilon, \alpha, b, c)$ implies $\Delta_L + 2\epsilon \leq \Delta_R$ (see Figure 1), while $L(\Delta_L, \alpha, b, c) < L(\Delta_L + 2\epsilon, \alpha, b, c)$ implies $\Delta_L + 2\epsilon \geq 0$ (again, see Figure 1).

The last " $\iff$ " holds because, in the penultimate system, the last two inequalities imply the first.[50] Q.E.D.

Claim 2 $L(\Delta_L - 2\epsilon, \alpha, b, c) \leq L(\Delta_L)$ *if and only if* $c \leq \alpha\epsilon, c \leq -\alpha(1-\alpha)\epsilon - (1-\alpha)b$.

Proof:

$L(\Delta_L - 2\epsilon, \alpha, b, c) \leq L(\Delta_L)$ implies $\Delta_L > 0$ (see footnote 29) and, hence, $L(\Delta_L) = \Delta_L$.

Assume that $\Delta_L - 2\epsilon \geq \Delta_1$. Then:

$$\begin{aligned}&L(\Delta_L - 2\epsilon, \alpha, b, c) \leq \Delta_L \iff\\ &c + \alpha(\Delta_L - 2\epsilon - \Delta_1) \leq \Delta_L \iff\\ &c \leq \alpha\epsilon.\end{aligned}$$

Moreover, $\Delta_L - 2\epsilon \geq \Delta_1$ (which can be written as $c \leq -2(1-\alpha)\epsilon - \frac{1-\alpha}{\alpha}b$) and $c \leq \alpha\epsilon$ imply $c \leq -\alpha(1-\alpha)\epsilon - (1-\alpha)b$. Thus, $L(\Delta_L - 2\epsilon, \alpha, b, c) \leq \Delta_L$ holds if and only if $c \leq \alpha\epsilon$ and $c \leq -\alpha(1-\alpha)\epsilon - (1-\alpha)b$.

[50] This is easiest to see to plotting the regions in which $c \leq -\frac{1-\alpha}{\alpha}b$, $c \geq \epsilon$, and $c < (1-\alpha)\epsilon - (1-\alpha)b$ in (b, c)-space while holding α and ϵ constant at arbitrary values. A similar remark applies to other instances below when we say that some inequalities imply other inequalities.

Next, assume that $\Delta_L - 2\epsilon < \Delta_1$. Then:

$$\begin{aligned}
&L(\Delta_L - 2\epsilon, \alpha, b, c) \leq \Delta_L \iff \\
&c - \alpha(\Delta_L - 2\epsilon - \Delta_1) \leq \Delta_L \iff \\
&c \leq -\alpha(1-\alpha)\epsilon - (1-\alpha)b.
\end{aligned}$$

Moreover $\Delta_L - 2\epsilon < \Delta_1$ (which can be written as $c > -2(1-\alpha)\epsilon - \frac{1-\alpha}{\alpha}b$) and $c \leq -\alpha(1-\alpha)\epsilon - (1-\alpha)b$ imply $c \leq \alpha\epsilon$. Thus, $L(\Delta_L - 2\epsilon, \alpha, b, c) \leq \Delta_L$ holds if and only if $c \leq \alpha\epsilon$ and $c \leq -\alpha(1-\alpha)\epsilon - (1-\alpha)b$. Q.E.D.

We now show that $\bar{\Delta} = \bar{\Delta}(\alpha, b, c, \epsilon)$ minimises the objective function in problem (2) subject to the constraint $\bar{\Delta} \leq \Delta_L$. To see this, write:

$$\begin{aligned}
&\int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} L(\Delta, \alpha, b, c) d\Delta = \\
&\int_{-\infty}^{\bar{\Delta}} L(\Delta, \alpha, b, c) d\Delta - \int_{-\infty}^{\bar{\Delta}-2\epsilon} L(\Delta, \alpha, b, c) d\Delta = \\
&\int_{-\infty}^{\bar{\Delta}} L(\Delta, \alpha, b, c) d\Delta - \int_{-\infty}^{\bar{\Delta}} L(\Delta - 2\epsilon, \alpha, b, c) d\Delta = \\
&\int_{-\infty}^{\bar{\Delta}} \big(L(\Delta, \alpha, b, c) - L(\Delta - 2\epsilon, \alpha, b, c)\big) d\Delta.
\end{aligned}$$

It is apparent from Figures 2-6 that $L(\Delta, \alpha, b, c) < L(\Delta - 2\epsilon, \alpha, b, c)$ for all $\Delta_L \leq \Delta < \bar{\Delta}(\alpha, b, c, \epsilon)$. Thus, $\int_{-\infty}^{\bar{\Delta}} \big(L(\Delta, \alpha, b, c) - L(\Delta - 2\epsilon, \alpha, b, c)\big) d\Delta$ is decreasing in $\bar{\Delta}$ over $[\Delta_L, \bar{\Delta}(\alpha, b, c, \epsilon)]$.[51] Similarly, it is apparent from Figures 2-7 that $L(\Delta, \alpha, b, c) > L(\Delta - 2\epsilon, \alpha, b, c)$ for all $\Delta > \bar{\Delta}(\alpha, b, c, \epsilon)$. Thus, $\int_{-\infty}^{\bar{\Delta}} \big(L(\Delta, \alpha, b, c) - L(\Delta - 2\epsilon, \alpha, b, c)\big) d\Delta$ is increasing in $\bar{\Delta}$ over $[\bar{\Delta}(\alpha, b, c, \epsilon), \infty)$. Thus, $\int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} L(\Delta, \alpha, b, c) d\Delta$ has a unique minimum over $[\Delta_L, \infty)$ at $\bar{\Delta} = \bar{\Delta}(\alpha, b, c, \epsilon)$.

Next, we show that $\bar{\Delta} = \bar{\Delta}(\alpha, b, c, \epsilon)$ also satisfies the constraints $r + 2\epsilon - 1 \leq \bar{\Delta} \leq r + 2\epsilon$ and $\bar{\Delta} \leq \Delta_R + 2\epsilon$, so that these constraints are not binding. It is apparent from Figures 2-7 that (i) $\bar{\Delta}(\alpha, b, c, \epsilon) \geq 0$, so that $\bar{\Delta}(\alpha, b, c, \epsilon) \geq r + 2\epsilon - 1$ and (ii) $\bar{\Delta}(\alpha, b, c, \epsilon) \leq \Delta_R + 2\epsilon$. To see that $\bar{\Delta}(\alpha, b, c, \epsilon) \leq r + 2\epsilon$ as well, note that $|b| \leq r$

[51] In the sixth case in expressions (3) and (4), the latter statement is trivially true because the interval $[\Delta_L, \bar{\Delta}(\alpha, b, c, \epsilon)]$ consists of a single point.

implies $\Delta_L \leq r$[52] and consider each of the six cases in expressions (3) and (4). In the first, fourth, and fifth cases, the default, D, corresponding to $\bar{\Delta}(\alpha, b, c, \epsilon)$ lies in the interval $[r, r + 2\epsilon]$ (see Figures 2, 5, and 6), so that $D \geq 0$, which is equivalent to $\bar{\Delta}(\alpha, b, c, \epsilon) \leq r + 2\epsilon$. In the second case, $\bar{\Delta}(\alpha, b, c, \epsilon) = \Delta_L + 2\epsilon \leq r + 2\epsilon$. In the third case, $\bar{\Delta}(\alpha, b, c, \epsilon) = \Delta' = \frac{c}{1-\alpha} - \frac{2\alpha}{1-\alpha}\epsilon - b < \frac{2\epsilon}{1-\alpha} - \frac{2\alpha}{1-\alpha}\epsilon - b = 2\epsilon - b \leq 2\epsilon - b + r + b = r + 2\epsilon$.[53] In the sixth case, $\bar{\Delta}(\alpha, b, c, \epsilon) = \Delta_L \leq r < r + 2\epsilon$.

Finally, observe that $R_{\text{AS-C}}(\alpha, \epsilon)$, $R_{\text{AS-L}}(\alpha, \epsilon)$, etc., cover the whole $(|b|, c)$-space according to Figure 12, so that each of expressions (3) and (4) completely characterises the optimal default as a function of the parameters. Q.E.D.

12.3 Three Technical Lemmas

In this section we state and prove three technical lemmas. The first and third will be used in some of the remaining proofs of results from the main text. The second lemma is used in the proof of the third.

Before turning to these lemmas, note that, if we use the definition of $\Psi(\alpha, b, c, \epsilon)$ and divide cases 2-6 in expression (3) (or expression (4)) into subcases corresponding to the panels in Figures 3-7, we can write:

[52] $|b| \leq r \iff 0 \leq b + r \implies -\frac{c}{1-\alpha} \leq r + b \iff \Delta_L \leq r$.

[53] The first inequality follows because $c < \epsilon$ in the third case in expression (4)

$$\Psi(\alpha,b,c,\epsilon)=\begin{cases}\int_{-\epsilon}^{\epsilon}L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in R_{\text{AS-C}}\\ \int_{\Delta_L}^{\Delta_L+2\epsilon}L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in R_{\text{AS-L}}\cap R_{\Delta_1<\Delta_L\leq 0}\\ \int_{\Delta_L}^{\Delta_L+2\epsilon}L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in R_{\text{AS-L}}\cap R_{\Delta_L>0}\\ \int_{\Delta'-2\epsilon}^{\Delta'}L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in R_{\text{HS1}}\cap R_{\Delta_1<\Delta_L\leq 0}\\ \int_{\Delta'-2\epsilon}^{\Delta'}L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in R_{\text{HS1}}\cap R_{\Delta_L>0}\\ \int_{\Delta''-2\epsilon}^{\Delta''}L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in R_{\text{HS2}}\cap R_{\Delta_L\leq\Delta_1}\\ \int_{\Delta''-2\epsilon}^{\Delta''}L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in R_{\text{HS2}}\cap R_{\Delta_1<\Delta_L\leq 0}\\ \int_{\Delta''-2\epsilon}^{\Delta''}L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in R_{\text{HS2}}\cap R_{\Delta_L>0}\\ \int_{\Delta_1-\epsilon}^{\Delta_1+\epsilon}L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in R_{\text{MS-HI}}\cap R_{\Delta_L\leq\Delta_1}\\ \int_{\Delta_1-\epsilon}^{\Delta_1+\epsilon}L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in R_{\text{MS-HI}}\cap R_{\Delta_1<\Delta_L\leq 0}\\ \int_{\Delta_1-\epsilon}^{\Delta_1+\epsilon}L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in R_{\text{MS-HI}}\cap R_{\Delta_L>0}\\ \int_{\Delta_L-2\epsilon}^{\Delta_L}L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in R_{\text{NS-corner}}\cap R_{\Delta_L-2\epsilon\geq\Delta_1}\\ \int_{\Delta_L-2\epsilon}^{\Delta_L}L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in R_{\text{NS-corner}}\cap R_{\Delta_L-2\epsilon<\Delta_1}\end{cases} \tag{10}$$

In the expression above and the rest of this section, we let an "R" with an inequality as a subscript denote the region of the parameter space in which the inequality holds. Also, given $-p$ and $1 \leq i \leq 13$, let $p(-p, i)$ denote the set of admissible values of p that satisfy the inequalities in the i^{th} case in expression (10).[54]

Lemma 4 *Given $-p$ fixed at an arbitrary value, the set of admissible values of p can be partitioned into finitely many intervals, such that each interval is contained in $p(-p, i)$ for some i $(1 \leq i \leq 13)$.*

Proof:

For each i $(1 \leq i \leq 13)$, $p(-p, i)$ is defined by inequalities that are linear or quadratic in p.[55] Each linear inequality can be written in the form "p less than (or equal to) C" or the form "p greater than (or equal to) C" and each quadratic inequality can be written in the form "p greater than (or equal to) C_1 and p less

[54]The admissible values for $\alpha/b/c/\epsilon$ are $0 < \alpha < 1/-r \leq b \leq 0/c \geq 0/0 < \epsilon \leq \frac{1-r}{2}$.

[55]See expression (4) and write out $\Delta_L \leq \Delta_1$, $\Delta_1 < \Delta_L \leq 0$, $\Delta_L > 0$, $\Delta_L - 2\epsilon \geq \Delta_1$, and $\Delta_L - 2\epsilon < \Delta_1$ in terms of the underlying parameters. (To make each inequality linear or quadratic in p, we may need to multiply each side to get rid of any denominators.)

than (or equal to) C_2" or the form "p less than (or equal to) C_1 or p greater than (or equal to) C_2", where C, C_1, C_2 are constants given that $-p$ is fixed. Thus, $p(-p, i)$ equals the intersection of a finite number of sets (one for each inequality), each of which consists of one or two intervals. It follows that $p(-p, i)$ is the union of a finite number of intervals $I_{i,1}, \ldots, I_{i,J(i)}$. Furthermore, because the set of admissible values of p equals $\bigcup_{i=1}^{13} p(-p, i)$, it can equivalently be written as $\bigcup_{i=1}^{13} \left(\bigcup_{j=1}^{J(i)} I_{i,j} \right)$. Thus, the set of admissible values of p is the union of finitely many intervals each of which is contained in $p(-p, i)$ for some $1 \leq i \leq 13$. Although the intervals $\{I_{i,j}\}_{1 \leq i \leq 13, 1 \leq j \leq J(i)}$ may overlap, any overlaps between them can easily be eliminated by breaking them down into smaller intervals. Q.E.D.

Lemma 5 *Denoting by* $cl(\cdot)$ *the closure of a set, expression (10) can equivalently be written as:*

$$\Psi(\alpha,b,c,\epsilon) =$$

$$\begin{cases}
\epsilon^2 & \textit{if } (\alpha,b,c,\epsilon) \in cl(R_{\textit{AS-C}}) \\
\frac{\Delta_L^2}{2} + \frac{(\Delta_L+2\epsilon)^2}{2} & \textit{if } (\alpha,b,c,\epsilon) \in cl(R_{\textit{AS-L}}) \cap R_{\Delta_1 \le \Delta_L \le 0} \\
2\epsilon(\Delta_L+\epsilon) & \textit{if } (\alpha,b,c,\epsilon) \in cl(R_{\textit{AS-L}}) \cap R_{\Delta_L \ge 0} \\
(\Delta_L - \Delta' + 2\epsilon)\frac{\Delta'+c+\alpha\Delta_L-(1-\alpha)b}{2} + & \\
+\frac{\Delta_L^2}{2} + \frac{\Delta'^2}{2} & \textit{if } (\alpha,b,c,\epsilon) \in cl(R_{\textit{HS1}}) \cap R_{\Delta_1 \le \Delta_L \le 0} \\
(\Delta_L - \Delta' + 2\epsilon)\frac{\Delta'+c+\alpha\Delta_L-(1-\alpha)b}{2} + & \\
+\frac{(\Delta'-\Delta_L)(\Delta'+\Delta_L)}{2} & \textit{if } (\alpha,b,c,\epsilon) \in cl(R_{\textit{HS1}}) \cap R_{\Delta_L \ge 0} \\
(\Delta_L - \Delta'' + 2\epsilon)\frac{\Delta''-\Delta_L}{2} + \frac{\Delta_L^2}{2} + \frac{\Delta''^2}{2} & \textit{if } (\alpha,b,c,\epsilon) \in cl(R_{\textit{HS2}}) \cap R_{\Delta_L \le \Delta_1} \\
(\Delta_1 - \Delta'' + 2\epsilon)\frac{c+\Delta''}{2} + & \\
+(\Delta_L - \Delta_1)\frac{2c+\alpha\Delta_L-(1-\alpha)b}{2} + & \\
+\frac{\Delta_L^2}{2} + \frac{\Delta''^2}{2} & \textit{if } (\alpha,b,c,\epsilon) \in cl(R_{\textit{HS2}}) \cap R_{\Delta_1 \le \Delta_L \le 0} \\
(\Delta_1 - \Delta'' + 2\epsilon)\frac{c+\Delta''}{2} + & \\
+(\Delta_L - \Delta_1)\frac{2c+\alpha\Delta_L-(1-\alpha)b}{2} + & \\
+\frac{(\Delta''-\Delta_L)(\Delta''+\Delta_L)}{2} & \textit{if } (\alpha,b,c,\epsilon) \in cl(R_{\textit{HS2}}) \cap R_{\Delta_L \ge 0} \\
(\Delta_L - \Delta_1 + \epsilon)\frac{c-\alpha(\Delta_1-\epsilon)+(1-\alpha)b-\Delta_L}{2} + & \\
+\frac{\Delta_L^2}{2} + \frac{\Delta_R^2}{2} + & \\
+(\Delta_1 + \epsilon - \Delta_R)\frac{\Delta_R+c+\alpha(\Delta_1+\epsilon)-(1-\alpha)b}{2} & \textit{if } (\alpha,b,c,\epsilon) \in cl(R_{\textit{MS-HI}}) \cap R_{\Delta_L \le \Delta_1} \\
\epsilon\frac{2c-\alpha(\Delta_1-\epsilon)+(1-\alpha)b}{2} + & \\
+(\Delta_L - \Delta_1)\frac{2c+\alpha\Delta_L-(1-\alpha)b}{2}\frac{\Delta_L^2}{2} + \frac{\Delta_R^2}{2} + & \\
+(\Delta_1 + \epsilon - \Delta_R)\frac{\Delta_R+c+\alpha(\Delta_1+\epsilon)-(1-\alpha)b}{2} & \textit{if } (\alpha,b,c,\epsilon) \in cl(R_{\textit{MS-HI}}) \cap R_{\Delta_1 \le \Delta_L \le 0} \\
\epsilon\frac{2c-\alpha(\Delta_1-\epsilon)+(1-\alpha)b}{2} + & \\
+(\Delta_L - \Delta_1)\frac{2c+\alpha\Delta_L-(1-\alpha)b}{2} + & \\
+\frac{(\Delta_R-\Delta_L)(\Delta_R+\Delta_L)}{2} + & \\
+(\Delta_1 + \epsilon - \Delta_R)\frac{\Delta_R+c+\alpha(\Delta_1+\epsilon)-(1-\alpha)b}{2} & \textit{if } (\alpha,b,c,\epsilon) \in cl(R_{\textit{MS-HI}}) \cap R_{\Delta_L \ge 0} \\
2\epsilon(c+\alpha\Delta_L-(1-\alpha)b-\alpha\epsilon) & \textit{if } (\alpha,b,c,\epsilon) \in cl(R_{\textit{NS-corner}}) \cap R_{\Delta_L-2\epsilon \ge \Delta_1} \\
(\Delta_L - \Delta_1)\frac{2c+\alpha\Delta_L-(1-\alpha)b}{2} + & \\
+(\Delta_1 - \Delta_L + 2\epsilon)\frac{2c-\alpha(\Delta_L-2\epsilon)+(1-\alpha)b}{2} & \textit{if } (\alpha,b,c,\epsilon) \in cl(R_{\textit{NS-corner}}) \cap R_{\Delta_L-2\epsilon \le \Delta_1}
\end{cases} \tag{11}$$

Proof:

Consider expression (4). When $c = (1-\alpha)\epsilon - (1-\alpha)b$, $\Delta_L + 2\epsilon = \Delta'' = \epsilon$, so that the first inequality in the second case and the third inequality in the fourth case can be written as weak inequalities. When $c = \alpha\epsilon$, $\Delta' = \Delta_L$, so that the first inequality in the third case can be written as a weak one. When $c = \epsilon$, $\Delta' = \Delta_L + 2\epsilon$, so that the third inequality in the third case can be written as a weak one. When $c = -\alpha(1-\alpha)\epsilon - (1-\alpha)b$, $\Delta'' = \Delta_L$, so that the first inequality in the fourth case can be written as a weak one. When $c = 2\epsilon + \frac{1-\alpha}{\alpha}b$, $\Delta'' = \Delta'$, so that the last inequality in the fourth case can be written as a weak one. When $c = (1-\alpha)\epsilon + \frac{1-\alpha}{\alpha}b$, $\Delta'' = \Delta_1 + \epsilon$, so that the inequality in the fifth case can be written as a weak one. The upshot is that, for each case in (4), we can replace all strict inequalities with weak ones.

Consider R_{HS1}, the region of the parameter space in which the third case in expression (4) applies. We have $cl(R_{\text{HS1}}) = cl(R_{c>\alpha\epsilon} \cap R_{c\geq 2\epsilon + \frac{1-\alpha}{\alpha}b} \cap R_{c<\epsilon}) \subseteq \big(cl(R_{c>\alpha\epsilon}) \cap R_{c\geq 2\epsilon + \frac{1-\alpha}{\alpha}b} \cap cl(R_{c<\epsilon})\big) = \big(R_{c\geq\alpha\epsilon} \cap R_{c\geq 2\epsilon + \frac{1-\alpha}{\alpha}b} \cap R_{c\leq\epsilon}\big)$. The inclusion follows because the closure of the intersection is a subset of the intersection of the closures. Given that (i) we can write the conditions in the third case of (4) either as $(\alpha, b, c, \epsilon) \in R_{\text{HS1}}$ (this is just as in (4)) or as $(\alpha, b, c, \epsilon) \in \big(R_{c\geq\alpha\epsilon} \cap R_{c\geq 2\epsilon + \frac{1-\alpha}{\alpha}b} \cap R_{c\leq\epsilon}\big)$ (see the last paragraph) and (ii) $R_{\text{HS1}} \subseteq cl(R_{\text{HS1}}) \subseteq \big(R_{c\geq\alpha\epsilon} \cap R_{c\geq 2\epsilon + \frac{1-\alpha}{\alpha}b} \cap R_{c\leq\epsilon}\big)$, it follows that we can also write the conditions for this case as $(\alpha, b, c, \epsilon) \in cl(R_{\text{HS1}})$. Using an analogous logic for the other cases, we can write expression (4) as:

$$\bar{\Delta}(\alpha, b, c, \epsilon) = \begin{cases} \epsilon & \text{if } (\alpha, b, c, \epsilon) \in cl(R_{\text{AS-C}}) \\ \Delta_L + 2\epsilon & \text{if } (\alpha, b, c, \epsilon) \in cl(R_{\text{AS-L}}) \\ \Delta' & \text{if } (\alpha, b, c, \epsilon) \in cl(R_{\text{HS1}}) \\ \Delta'' & \text{if } (\alpha, b, c, \epsilon) \in cl(R_{\text{HS2}}) \\ \Delta_1 + \epsilon & \text{if } (\alpha, b, c, \epsilon) \in cl(R_{\text{MS-HI}}) \\ \Delta_L & \text{if } (\alpha, b, c, \epsilon) \in cl(R_{\text{NS-corner}}) \end{cases} \tag{12}$$

Using the definition of $\Psi(\alpha, b, c, \epsilon)$ and dividing cases 2-6 in expression (12) into exhaustive subcases, we get:

$$
\Psi(\alpha,b,c,\epsilon)=\begin{cases}
\int_{-\epsilon}^{\epsilon} L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in cl(R_{\text{AS-C}}) \\
\int_{\Delta_L}^{\Delta_L+2\epsilon} L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in cl(R_{\text{AS-L}})\cap R_{\Delta_1\le\Delta_L\le 0} \\
\int_{\Delta_L}^{\Delta_L+2\epsilon} L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in cl(R_{\text{AS-L}})\cap R_{\Delta_L\ge 0} \\
\int_{\Delta'-2\epsilon}^{\Delta'} L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in cl(R_{\text{HS1}})\cap R_{\Delta_1\le\Delta_L\le 0} \\
\int_{\Delta'-2\epsilon}^{\Delta'} L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in cl(R_{\text{HS1}})\cap R_{\Delta_L\ge 0} \\
\int_{\Delta''-2\epsilon}^{\Delta''} L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in cl(R_{\text{HS2}})\cap R_{\Delta_L\le\Delta_1} \\
\int_{\Delta''-2\epsilon}^{\Delta''} L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in cl(R_{\text{HS2}})\cap R_{\Delta_1\le\Delta_L\le 0} \\
\int_{\Delta''-2\epsilon}^{\Delta''} L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in cl(R_{\text{HS2}})\cap R_{\Delta_L\ge 0} \\
\int_{\Delta_1-\epsilon}^{\Delta_1+\epsilon} L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in cl(R_{\text{MS-HI}})\cap R_{\Delta_L\le\Delta_1} \\
\int_{\Delta_1-\epsilon}^{\Delta_1+\epsilon} L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in cl(R_{\text{MS-HI}})\cap R_{\Delta_1\le\Delta_L\le 0} \\
\int_{\Delta_1-\epsilon}^{\Delta_1+\epsilon} L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in cl(R_{\text{MS-HI}})\cap R_{\Delta_L\ge 0} \\
\int_{\Delta_L-2\epsilon}^{\Delta_L} L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in cl(R_{\text{NS-corner}})\cap R_{\Delta_L-2\epsilon\ge\Delta_1} \\
\int_{\Delta_L-2\epsilon}^{\Delta_L} L(\Delta,\alpha,b,c)d\Delta & \text{if } (\alpha,b,c,\epsilon)\in cl(R_{\text{NS-corner}})\cap R_{\Delta_L-2\epsilon\le\Delta_1}
\end{cases} \tag{13}
$$

Evaluating each of the thirteen pieces in expression (13) by computing the area under $L(\cdot,\alpha,b,c)$ over $[\bar{\Delta}(\alpha,b,c,\epsilon)-2\epsilon, \bar{\Delta}(\alpha,b,c,\epsilon)]$ as the sum of the areas of various triangles and/or trapezoids yields expression (11). Q.E.D.

Lemma 6 *For any two admissible values $p' < p''$ of $p \in \{\alpha,b,c,\epsilon\}$ and any fixed value of $-p$, $\Psi(\cdot,-p)$ is absolutely continuous on $[p',p'']$.*

Proof:

Given $-p$ and $1 \le i \le 13$, let $q(-p,i)$ denote the set of values of p that satisfy the inequalities in the i^{th} case in expression (11). Note the following.

First, given Lemma 4, the fact that $p(-p,i) \subseteq q(-p,i)$, and the fact that $q(-p,i)$ is closed, it follows that we can decompose $[p',p'']$ into finitely many intervals $[p_1,p_2],\ldots,[p_{n-1},p_n]$, such that $p' = p_1 < p_2 \ldots < p_{n-1} < p_n = p''$ and each interval is contained in $q(-p,i)$ for some i $(1 \le i \le 13)$.

Second, at any fixed value of $-p$, each piece in expression (11) is absolutely continuous in p on any closed and bounded interval because:

(i) the sum and product of two functions that are absolutely continuous on a closed and bounded interval is absolutely continuous on this interval;

(ii) if a function is absolutely continuous on a closed and bounded interval and is nowhere equal to zero on that interval, its inverse is also absolutely continuous on that interval;

(iii) a linear function is absolutely continuous;

(iv) when $p \in \{b, c, \epsilon\}$, each of Δ_L, Δ_R, Δ_1, Δ', and Δ'' is a linear function of p, and is hence absolutely continuous in p;

(v) when $p = \alpha$, each of Δ_L, Δ_R, Δ_1, Δ', and Δ'' can be obtained through addition, multiplication, and taking inverses of nowhere-zero linear functions of p, so that each of Δ_L, Δ_R, Δ_1, Δ', and Δ'' is absolutely continuous in p on any closed and bounded interval;

(vi) each of the thirteen pieces in expression (11) is obtained from Δ_L, Δ_R, Δ_1, Δ', Δ'', and the parameters through addition and multiplication, so that it is absolutely continuous in p on any closed and bounded interval.

Thus, $\Psi(\cdot, -p)$ is absolutely continuous on each interval $[p_k, p_{k+1}]$ $(1 \leq k \leq n-1)$.

It readily follows that $\Psi(\cdot, -p)$ is continuous on $[p', p'']$: if $\Psi(\cdot, -p)$ is continuous on each of two adjacent intervals $[p_{k-1}, p_k]$ and $[p_k, p_{k+1}]$, it must be continuous on $[p_{k-1}, p_{k+1}]$ and, by repeating this argument for $2 \leq k \leq n-1$, on $[p', p'']$.

It remains to show that $\Psi(\cdot, -p)$ is absolutely continuous on $[p', p'']$. Let $p \in (p_1, p_n]$ and let k' be such that $p \in (p_{k'}, p_{k'+1}]$. Then,

$$\Psi(p,-p) = \Psi(p',-p) + \textstyle\sum_{k=2}^{k=k'} \left(\Psi(p_k,-p) - \Psi(p_{k-1},-p)\right) + \Psi(p,-p) - \Psi(p_{k'},-p) =$$
$$\Psi(p',-p) + \textstyle\sum_{k=2}^{k=k'} \left(\int_{p_{k-1}}^{p_k} \Psi_p(\tilde{p},-p)d\tilde{p}\right) + \int_{p_{k'}}^{p} \Psi_p(\tilde{p},-p)d\tilde{p} =$$
$$\Psi(p',-p) + \textstyle\int_{p'}^{p} \Psi_p(\tilde{p},-p)d\tilde{p}$$

The penultimate equality follows from the absolute continuity of $\Psi(\cdot, -p)$ on each $[p_{k-1}, p_k]$. The last equality follows from the continuity of $\Psi(\cdot, -p)$ on $[p', p'']$. The fact that $\Psi(p, -p) = \Psi(p', -p) + \int_{p'}^{p} \Psi_p(\tilde{p}, -p)d\tilde{p}$ for any $p \in (p', p'']$ means that $\Psi(\cdot, -p)$ is absolutely continuous on $[p', p'']$. Q.E.D.

12.4 Proof of Lemma 2

Consider $\int_{l(p)}^{u(p)} f(\Delta, p) d\Delta$. Whenever (i) $f(\Delta, p)$ and $\frac{\partial f}{\partial p}(\Delta, p)$ are continuous and (ii) $l(p)$ and $u(p)$ are continuously differentiable, we can use the Leibniz integral rule:

$$\frac{\partial}{\partial p} \int_{l(p)}^{u(p)} f(\Delta, p) d\Delta = \int_{l(p)}^{u(p)} \frac{\partial f}{\partial p}(\Delta, p) d\Delta - f(l(p), p) \frac{dl}{dp}(p) + f(u(p), p) \frac{du}{dp}(p)$$

Whenever we need to apply the Leibniz integral rule, $f(\Delta, p)$ will be linear in Δ and p so that (i) will hold; $l(p)$ and $u(p)$ will be linear in p or will be of the form $\frac{1-\alpha}{\alpha} b$, $-\frac{c}{1-\alpha} - b$, or $\frac{c}{1-\alpha} - b$, so that (ii) will hold.

Given Lemma 4, the set of values of p ($p \in \{\alpha, b, c, \epsilon\}$) that are not in the interior of some case in expression (10) has measure 0. Thus, it suffices to prove the validity of expressions (6) and (7) for any value of p that lies in the interior of one of the cases in expression (10). For any such p, $\Psi_p(p, -p)$ can be computed by taking the derivative of the i^{th} piece of Ψ in expression (10). Thus, we can prove the lemma by showing, for each i ($1 \leq i \leq 13$), that the i^{th} piece in expression (10) has a derivative which equals the corresponding piece in expression (6) or (7).

For each i ($1 \leq i \leq 13$), we can proceed as follows. First, we decompose $[\bar{\Delta}(\alpha, b, c, \epsilon) - 2\epsilon, \bar{\Delta}(\alpha, b, c, \epsilon)]$ into (finitely many) nonoverlapping intervals, $I_1, I_2, \ldots$, over each of which $L(\cdot, \alpha, b, c)$ is linear. Then, we express $\int_{\bar{\Delta}(\alpha,b,c,\epsilon)-2\epsilon}^{\bar{\Delta}(\alpha,b,c,\epsilon)} L(\Delta, \alpha, b, c) d\Delta$ as $\int_{I_1} L(\Delta, \alpha, b, c) d\Delta + \int_{I_2} L(\Delta, \alpha, b, c) d\Delta + \ldots$ Finally, we apply the Leibniz integral rule to each of the latter integrals.

Going through this procedure for all thirteen cases in expression (10) is rather tedious and we omit this here. Instead, to show how the procedure works, we offer a proof for the tenth case in expression (10), which is a special case of the third case in expression (6) and the first case in expression (7).

The tenth piece in expression (10) can be decomposed as follows.

$$
\begin{aligned}
&\int_{\Delta_1-\epsilon}^{\Delta_1+\epsilon} L(\Delta,\alpha,b,c)d\Delta = \\
&= \int_{\Delta_1-\epsilon}^{\Delta_1} L(\Delta,\alpha,b,c)d\Delta + \int_{\Delta_1}^{\Delta_L} L(\Delta,\alpha,b,c)d\Delta + \\
&+ \int_{\Delta_L}^{0} L(\Delta,\alpha,b,c)d\Delta + \int_{0}^{\Delta_R} L(\Delta,\alpha,b,c)d\Delta + \int_{\Delta_R}^{\Delta_1+\epsilon} L(\Delta,\alpha,b,c)d\Delta = \\
&= \int_{\Delta_1-\epsilon}^{\Delta_1} \big(c-\alpha(\Delta-\Delta_1)\big)d\Delta + \int_{\Delta_1}^{\Delta_L} \big(c+\alpha(\Delta-\Delta_1)\big)d\Delta + \\
&+ \int_{\Delta_L}^{0} (-\Delta)d\Delta + \int_{0}^{\Delta_R} \Delta d\Delta + \int_{\Delta_R}^{\Delta_1+\epsilon} \big(c+\alpha(\Delta-\Delta_1)\big)d\Delta.
\end{aligned}
$$

Applying the Leibniz integral rule to each integral separately, we get:

$$
\begin{aligned}
&\Psi_p(\alpha,b,c,\epsilon) = \\
&= \int_{\Delta_1-\epsilon}^{\Delta_1} \frac{\partial\big(c-\alpha(\Delta-\Delta_1)\big)}{\partial p}d\Delta - \big(c+\alpha\epsilon\big)\frac{d(\Delta_1-\epsilon)}{dp} + c\frac{d\Delta_1}{dp} + \\
&+ \int_{\Delta_1}^{\Delta_L} \frac{\partial\big(c+\alpha(\Delta-\Delta_1)\big)}{\partial p}d\Delta - c\frac{d\Delta_1}{dp} + \big(c+\alpha(\Delta_L-\Delta_1)\big)\frac{d\Delta_L}{dp} + \Delta_L\frac{d\Delta_L}{dp} + \\
&+ \Delta_R\frac{d\Delta_R}{dp} + \int_{\Delta_R}^{\Delta_1+\epsilon} \frac{\partial\big(c+\alpha(\Delta-\Delta_1)\big)}{\partial p}d\Delta - \big(c+\alpha(\Delta_R-\Delta_1)\big)\frac{d\Delta_R}{dp} + \big(c+\alpha\epsilon\big)\frac{d(\Delta_1+\epsilon)}{dp} = \\
&= \big(c+\alpha(\Delta_L-\Delta_1)+\Delta_L\big)\frac{d\Delta_L}{dp} + \int_{\Delta_1-\epsilon}^{\Delta_1} \frac{\partial\big(c-\alpha(\Delta-\Delta_1)\big)}{\partial p}d\Delta + \\
&+ \int_{\Delta_1}^{\Delta_L} \frac{\partial\big(c+\alpha(\Delta-\Delta_1)\big)}{\partial p}d\Delta + \int_{\Delta_R}^{\Delta_1+\epsilon} \frac{\partial\big(c+\alpha(\Delta-\Delta_1)\big)}{\partial p}d\Delta + 2\big(c+\alpha\epsilon\big)\frac{d\epsilon}{dp}.
\end{aligned}
$$

The latter expression equals the third piece in expression (6) when $p \in \{\alpha, b, c\}$ and equals the first piece in expression (7) when $p = \epsilon$. Q.E.D.

12.5 Proof of Lemma 3

$b > -\frac{\alpha\epsilon}{2}$ implies $(\alpha, b, c, \epsilon) \in R_{\text{AS-C}} \cup (R_{\text{HS2}} \cap R_{\Delta_L \leq \Delta_1}) \cup R_{\text{MS-HI}}$ (see Figure 12). Thus, for the purposes of proving parts 1) and 2) of Lemma 3, we can restrict attention to these regions.

Proof of statement 1):

If $(\alpha, b, c, \epsilon) \in R_{\text{AS-C}}$, the first piece in expression (6) applies, so that $\Psi_\alpha(\alpha, b, c, \epsilon) = 0$.

Now, suppose $(\alpha, b, c, \epsilon) \in R_{\text{HS2}} \cap R_{\Delta_L \le \Delta_1}$. This corresponds to the left panel in Figure 5 and the third case in expression (6), so that:

$$\Psi_\alpha(\alpha, b, c, \epsilon) = -\int_{\Delta''-2\epsilon}^{\Delta_L} (\Delta + b) d\Delta > 0$$

The inequality follows because $\Delta'' - 2\epsilon < \Delta_L$ and $\Delta + b < 0$ for $\Delta \in [\Delta'' - 2\epsilon, \Delta_L]$.

Next, suppose $(\alpha, b, c, \epsilon) \in R_{\text{MS-HI}} \cap R_{\Delta_L \le \Delta_1}$. This corresponds to the left panel in Figure 6 and the third case in expression (6), so that:

$$\Psi_\alpha(\alpha, b, c, \epsilon) = -\int_{\Delta_1-\epsilon}^{\Delta_L} (\Delta + b) d\Delta + \int_{\Delta_R}^{\Delta_1+\epsilon} (\Delta + b) d\Delta > 0$$

The inequality follows because $\Delta_1 - \epsilon < \Delta_L$ (see the left panel in Figure 6), $\Delta + b < 0$ for $\Delta \in [\Delta_1 - \epsilon, \Delta_L]$, $\Delta_R < \Delta_1 + \epsilon$, and $\Delta + b > 0$ for $\Delta \in [\Delta_R, \Delta_1 + \epsilon]$.

Finally, suppose $(\alpha, b, c, \epsilon) \in (R_{\text{MS-HI}} \cap R_{\Delta_1 < \Delta_L \le 0}) \cup (R_{\text{MS-HI}} \cap R_{\Delta_L > 0})$. This corresponds to the centre and right panels in Figure 6 and the third case in expression (6), so that:

$$\begin{aligned}
&\Psi_\alpha(\alpha, b, c, \epsilon) = \\
&\big(c + \alpha|\Delta_L - \Delta_1| - |\Delta_L|\big)\Big(-\frac{c}{(1-\alpha)^2}\Big) - \int_{\Delta_1-\epsilon}^{\Delta_1} (\Delta + b) d\Delta + \int_{\Delta_1}^{\Delta_L} (\Delta + b) d\Delta + \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \int_{\Delta_R}^{\Delta_1+\epsilon} (\Delta + b) d\Delta > \\
&\big(c + \alpha|\Delta_L - \Delta_1| - |\Delta_L|\big)\Big(-\frac{c}{(1-\alpha)^2}\Big) - \int_{2\Delta_1-\Delta_R}^{\Delta_1} (\Delta + b) d\Delta + \int_{\Delta_1}^{\Delta_L} (\Delta + b) d\Delta > \\
&\big(c + \alpha|\Delta_L - \Delta_1| - |\Delta_L|\big)\Big(-\frac{c}{(1-\alpha)^2}\Big) - \int_{2\Delta_1-\Delta_R}^{\Delta_1} (\Delta_1 + b) d\Delta + \int_{\Delta_1}^{\Delta_L} (\Delta_1 + b) d\Delta = \\
&\big(c + \alpha|\Delta_L - \Delta_1| - |\Delta_L|\big)\Big(-\frac{c}{(1-\alpha)^2}\Big) - (\Delta_1 + b)(\Delta_R - \Delta_L)
\end{aligned}$$

The first inequality follows because $\Delta_1 - \epsilon < 2\Delta_1 - \Delta_R \le \Delta_1$[56], $\Delta + b < 0$ for

[56] To see this, note that $\Delta_1 - \epsilon < 2\Delta_1 - \Delta_R$ can be written as $\Delta_1 + \epsilon > \Delta_R$ and $2\Delta_1 - \Delta_R \le \Delta_1$ can be written as $\Delta_1 \le \Delta_R$.

$\Delta \in [\Delta_1 - \epsilon, 2\Delta_1 - \Delta_R]$, and $\Delta + b > 0$ for $\Delta \in [\Delta_R, \Delta_1 + \epsilon]$. The second inequality holds because $\Delta = \Delta_1$ maximises $\Delta + b$ over $[2\Delta_1 - \Delta_R, \Delta_1]$ and minimises $\Delta + b$ over $[\Delta_1, \Delta_L]$.

If $\Delta_1 < \Delta_L \leq 0$, the last line simplifies to $-\frac{2c}{(1-\alpha)^2}\Big(\alpha\Delta_L + \frac{(1-\alpha)^2}{\alpha}b\Big)$, which is greater than or equal to 0. If $\Delta_L > 0$, the last line simplifies to $\frac{2c}{1-\alpha}(\Delta_L - \Delta_1)$, which is also greater than or equal to 0. Q.E.D.

Proof of statement 2):

Suppose $(\alpha, b, c, \epsilon) \in R_{\text{AS-C}}$. This corresponds to the first case in expression (6), so that $\Psi_b(\alpha, b, c, \epsilon) = 0 > -2\epsilon$.

Next, suppose $(\alpha, b, c, \epsilon) \in R_{\text{HS2}} \cap R_{\Delta_L \leq \Delta_1}$. This which corresponds to the left panel in Figure 5 and the third case in expression (6), so that:

$$\Psi_b(\alpha, b, c, \epsilon) = \int_{\Delta'' - 2\epsilon}^{\Delta_L} (1-\alpha) d\Delta > 0 > -2\epsilon$$

Next, suppose $(\alpha, b, c, \epsilon) \in R_{\text{MS-HI}} \cap R_{\Delta_L \leq \Delta_1}$. This corresponds to the left panel in Figure 6 and the third case in expression (6), so that:

$$\Psi_b(\alpha, b, c, \epsilon) = \int_{\Delta_1 - \epsilon}^{\Delta_L} (1-\alpha) d\Delta - \int_{\Delta_R}^{\Delta_1 + \epsilon} (1-\alpha) d\Delta \geq 0 > -2\epsilon$$

The first inequality follows because, as can be seen in the left panel of Figure 6, the distance between Δ_L and $(\Delta_1 - \epsilon)$ is greater than or equal to the distance between Δ_R and $(\Delta_1 + \epsilon)$.

Finally, suppose $(\alpha, b, c, \epsilon) \in (R_{\text{MS-HI}} \cap R_{\Delta_1 < \Delta_L \leq 0}) \cup (R_{\text{MS-HI}} \cap R_{\Delta_L > 0})$. This corresponds to the centre and right panels in Figure 6 and the third case in expression (6), so that:

$$\begin{aligned}&\Psi_b(\alpha, b, c, \epsilon) = \\ &-\big(c + \alpha|\Delta_L - \Delta_1| - |\Delta_L|\big) + \int_{\Delta_1 - \epsilon}^{\Delta_1} (1-\alpha) d\Delta - \int_{\Delta_1}^{\Delta_L} (1-\alpha) d\Delta - \int_{\Delta_R}^{\Delta_1 + \epsilon} (1-\alpha) d\Delta\end{aligned}$$

If $\Delta_1 < \Delta_L \leq 0$, the last line simplifies to $-2\Delta_L$, which is greater than or equal to 0 and, hence, greater than -2ϵ. If $\Delta_L > 0$, the last line equals 0, which is greater than -2ϵ. Q.E.D.

Proof of statement 3):

Suppose $(\alpha, b, c, \epsilon) \in R_{\text{AS-C}}$. This corresponds to the first case in expression (6), so that $\Psi_c(\alpha, b, c, \epsilon) = 0 < 2\epsilon$.

Next, suppose $(\alpha, b, c, \epsilon) \in R_{\text{AS-L}}$, which corresponds to Figure 3 and the second case in expression (6). We have:

$$\Psi_c = (\Delta_L + 2\epsilon - |\Delta_L|)(-\frac{1}{1-\alpha}) < 0 < 2\epsilon$$

The first inequality holds because $\Delta_L + 2\epsilon > |\Delta_L|$, which can be seen in Figure 3.

Next, suppose $(\alpha, b, c, \epsilon) \in R_{\text{HS1}} \cup R_{\text{HS2}} \cup R_{\text{MS-HI}}$. This corresponds to Figures 4-6 and the third case in expression (6), so that:

$$\begin{aligned}
&\Psi_c(\alpha, b, c, \epsilon) = \\
&\Big(c + \alpha|\Delta_L - \Delta_1| - |\Delta_L|\Big)\Big(-\frac{1}{1-\alpha}\Big) + \\
&\int_{(-\infty,\Delta_1]\cap[\Delta_L,\Delta_R]'\cap[\bar{\Delta}(\alpha,b,c,\epsilon)-2\epsilon,\bar{\Delta}(\alpha,b,c,\epsilon)]} 1 d\Delta + \int_{[\Delta_1,\infty)\cap[\Delta_L,\Delta_R]'\cap[\bar{\Delta}(\alpha,b,c,\epsilon)-2\epsilon,\bar{\Delta}(\alpha,b,c,\epsilon)]} 1 d\Delta \leq \\
&\int_{(-\infty,\Delta_1]\cap[\Delta_L,\Delta_R]'\cap[\bar{\Delta}(\alpha,b,c,\epsilon)-2\epsilon,\bar{\Delta}(\alpha,b,c,\epsilon)]} 1 d\Delta + \int_{[\Delta_1,\infty)\cap[\Delta_L,\Delta_R]'\cap[\bar{\Delta}(\alpha,b,c,\epsilon)-2\epsilon,\bar{\Delta}(\alpha,b,c,\epsilon)]} 1 d\Delta = \\
&\int_{[\Delta_L,\Delta_R]'\cap[\bar{\Delta}(\alpha,b,c,\epsilon)-2\epsilon,\bar{\Delta}(\alpha,b,c,\epsilon)]} 1 d\Delta \leq 2\epsilon
\end{aligned}$$

When $c > 0$, the last inequality is strict because (i) $[\bar{\Delta}(\alpha, b, c, \epsilon) - 2\epsilon, \bar{\Delta}(\alpha, b, c, \epsilon)]$ has length 2ϵ and (ii) $[\bar{\Delta}(\alpha, b, c, \epsilon) - 2\epsilon, \bar{\Delta}(\alpha, b, c, \epsilon)] \cap [\Delta_L, \Delta_R]$ has positive measure.

Finally, suppose $(\alpha, b, c, \epsilon) \in R_{\text{NS-corner}}$. This corresponds to Figure 7 and the fourth case in expression (6), so that:

$$\begin{aligned}
&\Psi_c(\alpha, b, c, \epsilon) = \\
&\Big(c + \alpha(\Delta_L - \Delta_1) - L(\Delta_L - 2\epsilon, \alpha, b, c)\Big)\Big(-\frac{1}{1-\alpha}\Big) + \int_{\min(\Delta_1,\Delta_L-2\epsilon)}^{\Delta_1} 1 d\Delta + \int_{\max(\Delta_1,\Delta_L-2\epsilon)}^{\Delta_L} 1 d\Delta = \\
&\Big(c + \alpha(\Delta_L - \Delta_1) - L(\Delta_L - 2\epsilon, \alpha, b, c)\Big)\Big(-\frac{1}{1-\alpha}\Big) + 2\epsilon \leq \\
&\Big(c + \alpha(\Delta_L - \Delta_1) - L(\Delta_L, \alpha, b, c)\Big)\Big(-\frac{1}{1-\alpha}\Big) + 2\epsilon \leq 2\epsilon
\end{aligned}$$

The first inequality follows because $L(\Delta_L - 2\epsilon, \alpha, b, c) \leq L(\Delta_L, \alpha, b, c)$. The second inequality holds because (i) $L(\Delta_L - 2\epsilon, \alpha, b, c) \leq L(\Delta_L, \alpha, b, c)$ implies $\Delta_L > 0$ (see footnote 29) and (ii) the size of the jump in $L(\cdot, \alpha, b, c)$ at Δ_L is nonnegative when $\Delta_L > 0$ (by Lemma 1). In fact, the second inequality is strict whenever $c > 0$ because, in that case, the size of the jump in $L(\cdot, \alpha, b, c)$ at Δ_L is positive when $\Delta_L > 0$ (by Lemma 1). Q.E.D.

12.6 Proof of Proposition 2

To prove the Proposition, we make use of a sequence of claims.

Claim 1 $\frac{1}{2\epsilon}\Psi(\alpha, b, 0, \epsilon) \geq -b$ *if and only if* $b \geq -\frac{\alpha\epsilon}{2}$.

Proof:

When $c = 0$, expression (4) simplifies to:

$$\bar{\Delta}(\alpha, b, 0, \epsilon) = \begin{cases} \Delta_1 + \epsilon & \text{if } b > -\alpha\epsilon \\ -b & \text{if } b \leq -\alpha\epsilon \end{cases},$$

and $L(\Delta, \alpha, b, 0) = \alpha|\Delta - \Delta_1|$. Thus,

$$\begin{aligned} \frac{1}{2\epsilon}\Psi(\alpha, b, 0, \epsilon) &= \begin{cases} \frac{1}{2\epsilon}\int_{\Delta_1-\epsilon}^{\Delta_1+\epsilon} \alpha|\Delta - \Delta_1| d\Delta & \text{if } b > -\alpha\epsilon \\ \frac{1}{2\epsilon}\int_{-b-2\epsilon}^{-b} \alpha|\Delta - \Delta_1| d\Delta & \text{if } b \leq -\alpha\epsilon \end{cases} \\ &= \begin{cases} \frac{\alpha\epsilon}{2} & \text{if } b > -\alpha\epsilon \\ \frac{b^2}{2\alpha\epsilon} + b + \alpha\epsilon & \text{if } -2\alpha\epsilon < b \leq -\alpha\epsilon \\ -b - \alpha\epsilon & \text{if } b \geq -2\alpha\epsilon \end{cases}. \end{aligned}$$

Note that $\frac{1}{2\epsilon}\Psi(\alpha, \cdot, 0, \epsilon)$ is continuous, constant with value $\frac{\alpha\epsilon}{2}$ on $(-\alpha\epsilon, 0]$, decreasing with absolute slope less than 1 on $(-2\alpha\epsilon, -\alpha\epsilon]$, and decreasing with absolute slope 1 on $(-\infty, -2\alpha\epsilon]$. Thus, the graph of $-b$ crosses the graph of $\frac{1}{2\epsilon}\Psi(\alpha, \cdot, 0, \epsilon)$ only once from below at $b = -\frac{\alpha\epsilon}{2}$. Q.E.D.

Claim 2 $\lim_{c\to\infty} \frac{1}{2\epsilon}\Psi(\alpha, b, c, \epsilon) < \lim_{c\to\infty}(c - b)$.

Proof:

$$\lim_{c\to\infty}\frac{1}{2\epsilon}\Psi(\alpha,b,c,\epsilon)=\lim_{c\to\infty}\frac{1}{2\epsilon}\int_{-\epsilon}^{\epsilon}|\Delta|d\Delta=\lim_{c\to\infty}\frac{1}{2\epsilon}\epsilon^2=\frac{\epsilon}{2}<\infty=\lim_{c\to\infty}(c-b)$$

The first equality follows because, as c approaches ∞, (α,b,c,ϵ) must eventually enter into and remain in $R_{\text{AS-C}}$ (see Figure 12). Thus, as c goes to ∞, the losses under the optimal default are limited, while the losses under AD grow without bound. Q.E.D.

Claim 3 *The function $\frac{1}{2\epsilon}\Psi(\alpha,b,c,\epsilon)-(c-b)$ is decreasing in c.*

This lemma basically follows from statement 3) in Lemma 3. Given that there is no simple intuition for statement 3) in Lemma 3, there is no simple intuition for Claim 3.

Proof:

Fix c_1,c_2, where $0\leq c_1<c_2$. By the absolute continuity of $\Psi(\alpha,b,c,\epsilon)$ (see Lemma 6), and hence of $\frac{1}{2\epsilon}\Psi(\alpha,b,c,\epsilon)-(c-b)$, in c on $[c_1,c_2]$:

$$\frac{1}{2\epsilon}\Psi(\alpha,b,c_2,\epsilon)-(c_2-b)-\big(\frac{1}{2\epsilon}\Psi(\alpha,b,c_1,\epsilon)-(c_1-b)\big)=$$
$$\int_{c_1}^{c_2}\Big(\frac{1}{2\epsilon}\Psi_c(\alpha,b,c,\epsilon)-1\Big)dc.$$

By statement 3) in Lemma 3, the integrand in the last integral is negative for almost all c, so that the integral is negative as well. Q.E.D.

Now, we are ready to prove statements 1) and 2) in Proposition 2.

Proof of statements 1) and 2) in Proposition 2:

Suppose $b\geq-\frac{\alpha\epsilon}{2}$. It follows directly from Claims 1-3 and the continuity of $\frac{1}{2\epsilon}\Psi(\alpha,b,c,\epsilon)-(c-b)$ in c[57] that there exists a unique value of c, denoted $c^{\text{AD}}(\alpha,b,\epsilon)$, such that $\frac{1}{2\epsilon}\Psi(\alpha,b,c^{\text{AD}}(\alpha,b,\epsilon),\epsilon)=c^{\text{AD}}(\alpha,b,\epsilon)-b$. Furthermore, it follows directly from Claim 3 that (i) AD is optimal if and only if $c\leq c^{\text{AD}}(\alpha,b,\epsilon)$ and (ii) $\bar{\Delta}(\alpha,b,c,\epsilon)$ is optimal if and only if $c\geq c^{\text{AD}}(\alpha,b,\epsilon)$.

[57]Absolute continuity implies continuity.

Now, suppose $b < -\frac{\alpha\epsilon}{2}$. It follows directly from Claims 1 and 3 that $\bar{\Delta}(\alpha, b, c, \epsilon)$ is the unique optimal default policy. Q.E.D.

It remains to prove statement 3) in Proposition 2. Let $w(\alpha, b, c, \epsilon) = \frac{1}{2\epsilon}\Psi(\alpha, b, c, \epsilon) - (c - b)$. We break down the proof of statement 3) into a sequence of claims.

Claim 4 *$c^{AD}(\alpha, b, \epsilon)$ is increasing in b.*

This claim follows mainly from Claim 3 and statement 2) in Lemma 3. Given that there is no simple intuition for Claim 3 and statement 2) in Lemma 3, there is no simple intuition for Claim 4.

Proof:

Fix $\alpha, b', b'', \epsilon$, such that $-\frac{\alpha\epsilon}{2} \leq b' < b'' \leq 0$. Assume that $c^{\mathrm{AD}}(\alpha, b', \epsilon) \geq c^{\mathrm{AD}}(\alpha, b'', \epsilon)$. We will arrive at a contradiction.

By the definition of c^{AD}, we have:

$$w(\alpha, b', c^{\mathrm{AD}}(\alpha, b', \epsilon), \epsilon) = w(\alpha, b'', c^{\mathrm{AD}}(\alpha, b'', \epsilon), \epsilon).$$

Subtracting $w(\alpha, b', c^{\mathrm{AD}}(\alpha, b'', \epsilon), \epsilon)$ from both sides:

$$\begin{aligned} w(\alpha, b', c^{\mathrm{AD}}(\alpha, b', \epsilon), \epsilon) - w(\alpha, b', c^{\mathrm{AD}}(\alpha, b'', \epsilon), \epsilon) = \\ w(\alpha, b'', c^{\mathrm{AD}}(\alpha, b'', \epsilon), \epsilon) - w(\alpha, b', c^{\mathrm{AD}}(\alpha, b'', \epsilon), \epsilon). \end{aligned}$$

By the absolute continuity of $\Psi(\alpha, \cdot, c^{\mathrm{AD}}(\alpha, b'', \epsilon), \epsilon)$, and hence of $w(\alpha, \cdot, c^{\mathrm{AD}}(\alpha, b'', \epsilon), \epsilon)$, on $[b', b'']$, we can write the latter equality as:

$$w(\alpha, b', c^{\mathrm{AD}}(\alpha, b', \epsilon), \epsilon) - w(\alpha, b', c^{\mathrm{AD}}(\alpha, b'', \epsilon), \epsilon) = \int_{b'}^{b''} \frac{\partial w}{\partial b}(\alpha, b, c^{\mathrm{AD}}(\alpha, b'', \epsilon), \epsilon) db$$

or, equivalently as:

$$w(\alpha, b', c^{\mathrm{AD}}(\alpha, b', \epsilon), \epsilon) - w(\alpha, b', c^{\mathrm{AD}}(\alpha, b'', \epsilon), \epsilon) = \int_{b'}^{b''} \big(\frac{1}{2\epsilon}\Psi_b(\alpha, b, c^{\mathrm{AD}}(\alpha, b'', \epsilon), \epsilon) + 1\big) db$$

By Claim 3, $w(\alpha, b, \cdot, \epsilon)$ is decreasing, so that the left-hand side is nonpositive. By statement 2) in Lemma 3, the integrand on the right-hand side is positive for all $b \in (b', b'')$, so that the right-hand-side is positive. We have arrived at a contradiction. Q.E.D.

Claim 5 *$c^{AD}(\alpha, b, \epsilon)$ is nondecreasing in α.*

This claim follows mainly from Claim 3 and statement 1) in Lemma 3. Given that there is no simple intuition for Claim 3 and statement 1) in Lemma 3, there is no simple intuition for Claim 5.

Proof:

Fix $\alpha', \alpha'', b, \epsilon$, such that $-\frac{2b}{\epsilon} \leq \alpha' < \alpha'' < 1$. Assume that $c^{\mathrm{AD}}(\alpha', b, \epsilon) > c^{\mathrm{AD}}(\alpha'', b, \epsilon)$.[58] We will arrive at a contradiction.

By the definition of c^{AD}, we have:

$$w(\alpha', b, c^{\mathrm{AD}}(\alpha', b, \epsilon), \epsilon) = w(\alpha'', b, c^{\mathrm{AD}}(\alpha'', b, \epsilon), \epsilon)$$

Subtracting $w(\alpha', b, c^{\mathrm{AD}}(\alpha'', b, \epsilon), \epsilon)$ from both sides:

$$\begin{aligned} w(\alpha', b, c^{\mathrm{AD}}(\alpha', b, \epsilon), \epsilon) - w(\alpha', b, c^{\mathrm{AD}}(\alpha'', b, \epsilon), \epsilon) = \\ w(\alpha'', b, c^{\mathrm{AD}}(\alpha'', b, \epsilon), \epsilon) - w(\alpha', b, c^{\mathrm{AD}}(\alpha'', b, \epsilon), \epsilon) \end{aligned}$$

By the absolute continuity of $\Psi(\cdot, b, c^{\mathrm{AD}}(\alpha'', b, \epsilon), \epsilon)$, and hence of $w(\cdot, b, c^{\mathrm{AD}}(\alpha'', b, \epsilon), \epsilon)$, on $[\alpha', \alpha'']$, we can write the latter equality as:

$$w(\alpha', b, c^{\mathrm{AD}}(\alpha', b, \epsilon), \epsilon) - w(\alpha', b, c^{\mathrm{AD}}(\alpha'', b, \epsilon), \epsilon) = \int_{\alpha'}^{\alpha''} \frac{\partial w}{\partial \alpha}(\alpha, b, c^{\mathrm{AD}}(\alpha'', b, \epsilon), \epsilon) d\alpha$$

or, equivalently as:

$$w(\alpha', b, c^{\mathrm{AD}}(\alpha', b, \epsilon), \epsilon) - w(\alpha', b, c^{\mathrm{AD}}(\alpha'', b, \epsilon), \epsilon) = \int_{\alpha'}^{\alpha''} \frac{1}{2\epsilon} \Psi_\alpha(\alpha, b, c^{\mathrm{AD}}(\alpha'', b, \epsilon), \epsilon) d\alpha$$

[58] $-\frac{2b}{\epsilon} < \alpha'$ and $-\frac{2b}{\epsilon} < \alpha''$ guarantee that c^{AD} is defined at (α', b, ϵ) and (α'', b, ϵ), respectively.

By Claim 3, $w(\alpha, b, \cdot, \epsilon)$ is decreasing, so that the left-hand side is negative. By statement 1) in Lemma 3, the integrand on the right-hand side is nonnegative for all $\alpha \in (\alpha', \alpha'')$, so that the right-hand-side is nonnegative. We have arrived at a contradiction. Q.E.D.

Claim 6 $c^{AD}(\alpha, -\frac{\alpha\epsilon}{2}, \epsilon) = 0$

Proof:

Consider the equation $\frac{1}{2\epsilon}\Psi(\alpha, -\frac{\alpha\epsilon}{2}, c, \epsilon) = c - \frac{\alpha\epsilon}{2}$. By Claim 1, $c = 0$ solves it. By statement 1) in Proposition 2, $c^{\mathrm{AD}}(\alpha, -\frac{\alpha\epsilon}{2}, \epsilon)$ is the unique value of c that solves it. Thus, it must be that $c^{\mathrm{AD}}(\alpha, -\frac{\alpha\epsilon}{2}, \epsilon) = 0$. Q.E.D.

Claim 7 $c^{AD}(\alpha, 0, \epsilon) = \begin{cases} \sqrt{\alpha(1-\alpha)}\epsilon & \textit{if } \alpha < 0.5 \\ 0.5\epsilon & \textit{if } \alpha \geq 0.5 \end{cases}$

Proof:

If $b = 0$, (α, b, c, ϵ) is either in $R_{\text{AS-C}}$ or in $R_{\text{MS-HI}} \cap R_{\Delta_L \leq \Delta_1}$ (see Figure 12).[59] Thus, $(\alpha, 0, c^{\mathrm{AD}}(\alpha, 0, \epsilon), \epsilon)$ is either in $R_{\text{AS-C}}$ or in $R_{\text{MS-HI}} \cap R_{\Delta_L \leq \Delta_1}$.

First, assume $(\alpha, 0, c^{\mathrm{AD}}(\alpha, 0, \epsilon), \epsilon) \in R_{\text{AS-C}}$, i.e., $-\frac{c^{\mathrm{AD}}(\alpha,0,\epsilon)}{1-\alpha} \leq -\epsilon$ or, equivalently $c^{\mathrm{AD}}(\alpha, 0, \epsilon) \geq (1-\alpha)\epsilon$. In this case, the first piece in expression (11) for Ψ applies, so that $c^{\mathrm{AD}}(\alpha, 0, \epsilon)$ is defined by $c^{\mathrm{AD}}(\alpha, 0, \epsilon) = \frac{\epsilon}{2}$. If $\alpha \geq 0.5$, $c^{\mathrm{AD}}(\alpha, 0, \epsilon) = \frac{\epsilon}{2}$ is consistent with $c^{\mathrm{AD}}(\alpha, 0, \epsilon) \geq (1-\alpha)\epsilon$. Thus, $c^{\mathrm{AD}}(\alpha, 0, \epsilon) = \frac{\epsilon}{2}$ if $\alpha \geq 0.5$.

Next, assume $(\alpha, 0, c^{\mathrm{AD}}(\alpha, 0, \epsilon), \epsilon) \in R_{\text{MS-HI}} \cap R_{\Delta_L \leq \Delta_1}$, i.e., $\epsilon > \frac{c^{\mathrm{AD}}(\alpha,0,\epsilon)}{1-\alpha}$ and $-\frac{c^{\mathrm{AD}}(\alpha,0,\epsilon)}{1-\alpha} \leq 0$. The latter two inequalities can be stated equivalently as $0 \leq c^{\mathrm{AD}}(\alpha, 0, \epsilon) < (1-\alpha)\epsilon$. In this case, the ninth piece in expression (11) for Ψ applies. This piece simplifies to $2c\epsilon - \frac{c^2}{1-\alpha} + \alpha\epsilon^2$ given that $b = 0$. Thus, $c^{\mathrm{AD}}(\alpha, 0, \epsilon)$ solves $\frac{1}{2\epsilon}\left(2c\epsilon - \frac{c^2}{1-\alpha} + \alpha\epsilon^2\right) = c$. Thus, $c^{\mathrm{AD}}(\alpha, 0, \epsilon) = \epsilon\sqrt{\alpha(1-\alpha)}$. If $\alpha < 0.5$, $c^{\mathrm{AD}}(\alpha, 0, \epsilon) = \epsilon\sqrt{\alpha(1-\alpha)}$ is consistent with $0 \leq c^{\mathrm{AD}}(\alpha, 0, \epsilon) < (1-\alpha)\epsilon$. Thus, $c^{\mathrm{AD}}(\alpha, 0, \epsilon) = \epsilon\sqrt{\alpha(1-\alpha)}$ if $\alpha < 0.5$. Q.E.D.

[59] Although it is not clear from the figure, the point $(\alpha, b, c, \epsilon) = (\alpha, 0, (1-\alpha)\epsilon, \epsilon)$ satisfies the condition in the first case in expression (4), so that this point is in $R_{\text{AS-C}}$.

12.7 Proof of Proposition 3

Part (i) follows from replacing (α, b, c, ϵ) with $(\alpha, kb, kc, k\epsilon)$ in expression (4)[60] and observing that the conditions in the i^{th} ($1 \leq i \leq 6$) case in expression (4) hold for $(\alpha, kb, kc, k\epsilon)$ if and only if they hold for (α, b, c, ϵ).

Turning to part (ii), note that it follows from expression (1) that $L(\Delta, \alpha, kb, kc) = kL(\frac{\Delta}{k}, \alpha, b, c)$. Utilising this, we have:

$$
\begin{aligned}
&\Psi(\alpha, kb, kc, k\epsilon) = \\
&\int_{\bar{\Delta}(\alpha,kb,kc,k\epsilon)-2k\epsilon}^{\bar{\Delta}(\alpha,kb,kc,k\epsilon)} L(\Delta, \alpha, kb, kc)d\Delta = \\
&\int_{k\bar{\Delta}(\alpha,b,c,\epsilon)-2k\epsilon}^{k\bar{\Delta}(\alpha,b,c,\epsilon)} kL(\frac{\Delta}{k}, \alpha, b, c)d\Delta = \\
&k^2 \int_{k\bar{\Delta}(\alpha,b,c,\epsilon)-2k\epsilon}^{k\bar{\Delta}(\alpha,b,c,\epsilon)} L(\frac{\Delta}{k}, \alpha, b, c)d\frac{\Delta}{k} = \\
&k^2 \int_{\bar{\Delta}(\alpha,b,c,\epsilon)-2\epsilon}^{\bar{\Delta}(\alpha,b,c,\epsilon)} L(\tilde{\Delta}, \alpha, b, c)d(\tilde{\Delta}) = \\
&k^2\Psi(\alpha, b, c, \epsilon),
\end{aligned}
$$

where the last equality utilises a change of the variable of integration.

Thus, $c - b \leq \frac{1}{2\epsilon}\Psi(\alpha, b, c, \epsilon)$ (i.e., given (α, b, c, ϵ), AD is an optimal default policy) if and only if $kc - kb \leq \frac{1}{2k\epsilon}\Psi(\alpha, kb, kc, k\epsilon)$ (i.e., given $(\alpha, kb, kc, k\epsilon)$, AD is an optimal default policy). Q.E.D.

[60]Note that, although expression (4) doesn't make this explicit, Δ_L, Δ_1,Δ', and Δ'' change when we change the parameters.

School of Economics and Finance

Queen Mary
University of London

School of Economics and Finance
Queen Mary University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 7356
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/research/workingpapers/