Optimal Pension Plan Default Policies when Employees are Biased*

Asen Ivanov†

Queen Mary University of London

November 8, 2020

Abstract

What is the optimal default contribution rate or default asset allocation in pension plans? Could active decision (i.e., not setting a default and forcing employees to make a decision) be optimal? These questions are studied in a model in which each employee is biased regarding her optimal contribution rate or asset allocation. In this model, active decision is never optimal and the optimal default is, depending on parameter values, one of three defaults. The paper explores how the parameters affect the average loss in the population at the optimal default. Policy implications are discussed.

Keywords: optimal defaults, libertarian paternalism, nudging, pension plan design

JEL: D14, D91, J26, J32

*I would like to thank Sarolta Laczó and Sujoy Mukerji for helpful comments.

†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 affect employees' contribution rates and asset allocations in these plans.[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 "active decision" (AD), i.e., to not setting a default and forcing employees to make a decision?

Although these questions have been studied previously, most of the existing literature on optimal default policies assumes either that individuals know their optimal options or that they have unbiased beliefs about them. This assumption runs counter to the common view that, left to their own devices, individuals would make systematic mistakes (e.g., by undersaving) and defaults can be used to "nudge" them towards better options.[2]

The current paper analyses a model in which employees are biased. In particular, for each employee, there is an optimal option, x, which corresponds 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 risk-free bonds). However, each employee is biased in the sense that, if her optimal option is x, her preferred option (i.e., the one she would choose) is $x - b$ (where $b > 0$ without loss of generality). One can interpret the bias, b, as a systematic mistake due to a reasoning mistake or a lack of information. Alternatively, one can view the bias as reflecting an externality. For example, if a perfectly rational, well-informed employee saves little for retirement because she anticipates being bailed out by the government, she would not be under-

[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. Laibson (2020) argued that the effect of the default contribution rate on savings is smaller than previously thought, especially in the long-run.

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

saving from the point of view of her own interest, but would be undersaving from the point of view of the government.[3]

Other key features of the model are that (i) employees' optimal options are uniformly distributed on an interval $[\underline{x}, \overline{x}]$ of width 2ϵ, (ii) if an employee decides actively, either because she opts out of the default or because the default policy is AD, she incurs a cost c, (iii) if an employee ends up with option x' when her optimal option is x, this entails a loss from deviations equal to $l(x - x')$, and (iv) each employee's loss equals her loss from deviations plus, possibly, the cost c.

Let us now sketch the main findings in the model. First, for a given default, suppose there is (i) an employee who thinks that the default is too high for her but is only just willing to stay with it and (ii) another employee whose optimal option is just slightly lower than that of the former employee and is hence only just willing to opt out. It turns out that the latter employee's loss is discontinuously higher than that of the former employee. This feature of the model leads to the optimal default policy being, for some parameter values, the default such that employees with $x = \underline{x}$ think it's too high, but are only just willing to stay with it. We'll call this the $\underline{x}$-PEL default, where "PEL" stands for "perceived equal loss (from staying or opting out)". Because even employees with the lowest possible optimal option stay with it, the $\underline{x}$-PEL default avoids the aforementioned discontinuous jump in losses.

Second, given their downward bias, employees are willing to stay with defaults far below their optimal options. Moreover, it turns out that, if $l(\cdot)$ is strictly convex, some employees may stay with a given default at a loss exceeding the loss from opting out. This feature of the model leads to the optimal default policy being, for some parameter values, the default such that it is below $\overline{x}$ and employees with $x = \overline{x}$ stay at a loss equal to the loss from opting out. We'll call this the $\overline{x}$-EL default, where "EL" stands for "equal loss (from staying or opting out)". The $\overline{x}$-EL default ensures

[3] For concreteness, some of the language below is based on the former interpretation.

that no employee stays with the default at a loss greater than the loss from opting out.

Third, given the parameters b, c, and ϵ, the optimal default policy is one of the following three defaults: the $\underline{x}$-PEL default, the $\overline{x}$-EL default, and the default placed in the middle of $[\underline{x}, \overline{x}]$. We'll call the latter the centre (C) default.

Finally, the average loss in the population at the optimal default is weakly increasing in b, either weakly decreasing or nonmonotone in c, and strictly increasing in ϵ.[4]

There is a burgeoning literature on optimal default policies. Most of the papers focus on factors omitted from the model in the current paper, such as employee heterogeneity in terms of the bias or the cost of deciding actively (Goldin and Reck (2019)), updating from the default (Goldin and Reck (2019) and Ivanov (2019)), procrastination in opting-out (Carrol et al. (2009)), decision costs (Carlin et al. (2013), Wisson (2016), Caplin and Martin (2017), Blumenstock et al. (2018)), normative ambiguity (Bernheim et al. (2015)), and normatively irrelevant cost of deciding actively (Goldin and Reck (2019)).[5] Most of these papers do not allow agents to be biased vis-à-vis their optimal options. The two exceptions are Goldin and Reck (2019) and Ivanov (2019).

Goldin and Reck (2019) allow for a bias vis-à-vis optimal options in an extension of their baseline model. This extension is more general than the model in the current paper. Notably, it allows (i) the planner to count a part of each employee's cost of deciding actively as normatively irrelevant and (ii) agents to be heterogeneous in terms of the cost of deciding actively and the magnitude of the bias. However, this generality allows Goldin and Reck only to partially analyse their model.[6] In contrast,

[4]The appendix explores how the parameters affect the optimal default.

[5]The appendix discusses these factors further.

[6]In particular, they derive an approximation of the first-order condition that an optimal default must satisfy, discuss whether marginally raising or lowering a default that is optimal in the absence

the current paper performs a more complete analysis by explicitly solving for the optimal default policy.

Ivanov (2019) modifies the model in the current paper by allowing employees, upon seeing a default, to update their preferred options in the direction of the default. One finding in that paper is that, in addition to the C and $\underline{x}$-PEL defaults here, three other defaults emerge as candidates for being the optimal default.[7]

One advantage of the current paper relative to Ivanov (2019) is that the analysis here is relatively straightforward, whereas the analysis in Ivanov (2019) is very complex and few intuitions for the main results emerge. Another advantage is that the strictly convex loss from deviations assumed here is probably more realistic than the linear loss from deviations assumed in Ivanov (2019). Of course, these advantages come at the cost of assuming away updating from the default.[8]

The paper proceeds as follows. Section 2 provides some examples of reasoning mistakes and lack of information in the context of contribution rates and in the context of asset allocation. Section 3 presents the model. Sections 4 and 5 cover the main findings. Section 6 takes a critical look at the model's assumptions. Section 7 concludes with some policy implications.

2 Reasoning Mistakes and Lack of Information

As mentioned, one interpretation of the bias is that it is a systematic mistake due to a reasoning mistake or a lack of information. The current section aims to elucidate this interpretation via some illustrative examples.

of any bias improves social welfare when a bias is present, and discuss how the presence of a bias might undermine the case for AD.

[7]One of those three other defaults in Ivanov (2019) is similar to the $\overline{x}$-EL default here.

[8]The appendix discusses evidence on updating from the default. In a nutshell, the evidence supports such updating in the context of asset allocation and is mixed in the context of contribution rates.

2.1 Mistakes and Undersaving

Let us consider two examples. The first one is about naiveté. There is evidence that employees think they will save more for retirement in the future, but then fail to follow through.[9] It is plausible that such naiveté about one's own future behaviour is due to a reasoning mistake (perhaps employees neglect to ask themselves why they expect their future "selves" to be more self-disciplined than their current "selves") or to a lack of information (perhaps any naiveté would evaporate if individuals were aware of the empirical evidence on naiveté). Observe that current saving is probably reduced if employees believe their future selves will make up for low current saving. Thus, naiveté probably leads to undersaving for retirement.

The second example is about myopia. Loosely speaking, myopia refers to a failure to think about the future (a reasoning mistake).[10] Because one would save more if one took the future into consideration, myopia leads to undersaving.

Note that these examples are merely illustrative. I do not wish to argue that individuals actually undersave for retirement. In fact, there is currently a heated theoretical and empirical debate on the issue.[11] I steer clear of this debate, merely noting that the current paper might be of interest to those who believe that individuals do indeed undersave.

2.2 Mistakes and Underinvestment in Stocks

Let us consider three examples. First, if young employees suffer from exponential growth bias,[12] they would underestimate, for a given expected annual equity premium,

[9]Choi et al. (2002). For evidence on naiveté more generally, see section 3.5 in Ericson and Laibson (2018).

[10]For an attempt to formally model myopia, see Gabaix and Laibson (2017).

[11]For an overview of the issues as well as further references, see Poterba (2015) as well as section 4.6 in Ericson and Laibson (2018).

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

the expected long-term difference in wealth that results from investing in stocks rather than in bonds. Second, perhaps employees are not aware of the high historical equity premium.[13] Third, perhaps young employees are not aware of the fact that historically equity returns have dominated bond returns over long horizons.[14] In each of these examples, employees would invest too small a fraction of their pension plan portfolios in stocks as a result of either a reasoning mistake (in the first example) or a lack of information (in the second and third examples).[15]

3 The Model

3.1 The Setup

The model setup is as follows. First, for each employee, there is an exogenously given 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 risk-free bonds). Employees' optimal options are uniformly distributed on $[\underline{x}, \overline{x}] \subseteq [0, 1]$, where $\underline{x} < \overline{x}$. Let 2ϵ denote the width of this interval, so that ϵ captures the degree of heterogeneity in the population.

Second, each employee is biased in the sense that, if her optimal option is x, her preferred option is $x - b$. We assume $b > 0$, which is without loss of generality.[16] Note that, if $b > \underline{x}$, then $x - b < 0$ for some employees. Allowing some employees to

[13] 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.

[14] Benartzi 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.

[15] Again, these examples are meant to be merely illustrative. I am not aware of any direct evidence on whether employees actually invest too small a fraction of their pension plan portfolios in stocks.

[16] If each employee sets the contribution rate or the allocation to stocks too high, we can 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 $b > 0$ would be appropriate.

choose $x-b<0$ 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 assume $b \leq \underline{x}$.

Third, 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 $c>0$. We interpret c as reflecting implementation costs, i.e., the time and effort required to contact the relevant people in human resources, to fill in the necessary paperwork, etc.

Fourth, if an employee ends up with option x' when her optimal option is x, we assume this entails a loss from deviations equal to $l(x-x')$ and a perceived loss from deviations (from the point of view of the employee) equal to $l(x-b-x')$. We maintain throughout the following assumptions on $l(\cdot)$: (i) $l(\cdot)$ is continuous, (ii) $l(0)=0$, (iii) $l(\cdot)$ is strictly decreasing on $(-\infty, 0]$ and strictly increasing on $[0, \infty)$, and (iv) $\lim_{z \to -\infty} l(z) = \lim_{z \to \infty} l(z) = \infty$. These assumptions should be uncontroversial. Some of the results below also assume that $l(\cdot)$ is differentiable on $(0, \infty)$, that $l(\cdot)$ is symmetric (i.e., $l(z) = l(-z)$ for all z), or that $l(\cdot)$ is strictly convex.[17]

3.2 Employee Behaviour

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 $c + l(x-(x-b)) = c + l(b)$.

How about if she faces a default, D? Letting $\Delta = x - D$, the loss associated with staying with the default is $l(\Delta)$. However, the perceived loss is $l(x-b-D) = l(\Delta - b)$. The employee stays with the default if and only if $c \geq l(\Delta - b)$. The latter inequality can be written as $\Delta_L \leq \Delta \leq \Delta_R$, where Δ_L and Δ_R are the two values of Δ that

[17]Figures below are drawn based on a symmetric and strictly convex $l(\cdot)$. However, explanations in the text referring to these figures do not rely on these properties of $l(\cdot)$ unless these properties are explicitly assumed.

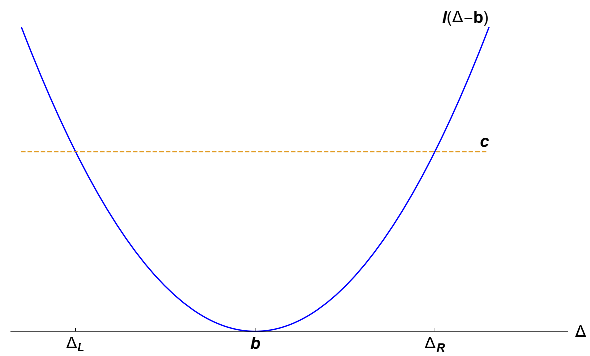

Figure 1: Determination of Δ_L and Δ_R.

solve the equation $c = l(\Delta - b)$ (see Figure 1). Note that, given that an employee with $\Delta = b$ thinks the default is ideal, an employee with $\Delta = \Delta_L/\Delta = \Delta_R$ thinks the default is too high/low but is only just willing to stay with it. If an employee opts out, she incurs the cost c and chooses $x - b$, so that her loss is the same as under AD, namely $c + l(b)$.

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:

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

Figure 2 shows the graph of $L(\cdot)$ for each of the cases $\Delta_L \leq 0$ and $\Delta_L > 0$. The following lemma establishes that key features of Figure 2 hold more generally.[18]

Lemma 1

1) If $\Delta_L \leq 0$, $L(\Delta_L) < c$.

2) If $\Delta_L > 0$, $L(\Delta_L) < l(b)$.

3) If $l(\cdot)$ is strictly convex, $L(\Delta_R) > c + l(b)$.

[18] All proofs are in the appendix.

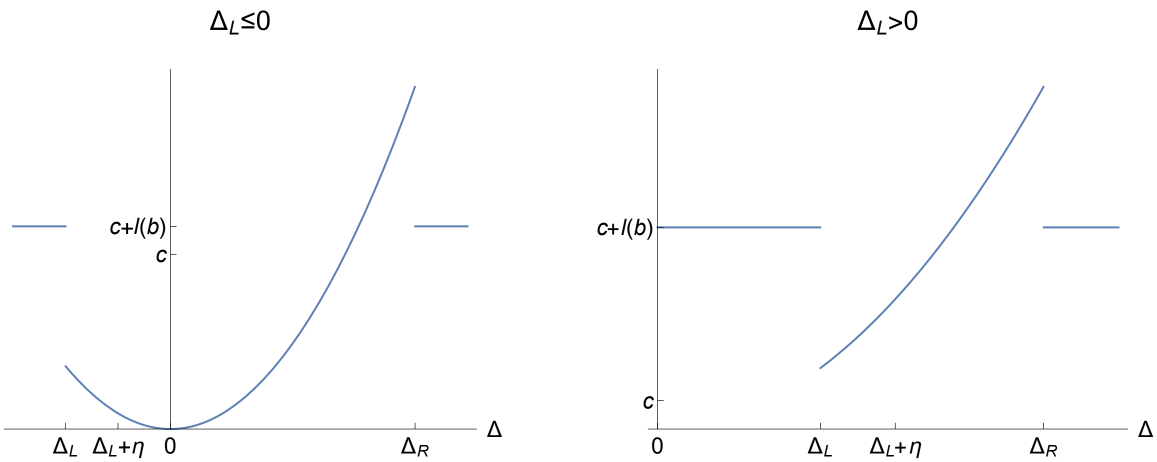

Figure 2: Graph of $L(\cdot)$.

The logic behind statement 1) goes as follows. Assume $\Delta_L \leq 0$ and consider an employee with $\Delta = \Delta_L$. The default is too high for this employee ($\Delta \leq 0$ is equivalent to $x \leq D$), and the employee recognises this (recall that an employee with $\Delta = \Delta_L$ thinks the default is too high). However, given her bias, she exaggerates the extent to which the default is too high and, as a result, exaggerates the loss from staying with it. Given that an employee with $\Delta = \Delta_L$ thinks that the loss from staying with the default is c, it must be that the true loss from staying, $l(\Delta_L)$, is less than c (see the left panel in Figure 3).

Statement 2) holds because, when $\Delta_L > 0$, an employee with $\Delta = \Delta_L$ is closer to the default than an employee with $\Delta = b$, so that $l(\Delta_L) < l(b)$ (see the right panel in Figure 3). Given that $L(\Delta_L) = l(\Delta_L)$, the latter inequality is equivalent to $L(\Delta_L) < l(b)$.

The intuition behind statement 3) is the following. Consider an employee with $\Delta = \Delta_R$ (recall that she stays with the default) and an employee who chooses actively. Each of them thinks she's incurring a loss equal to c, and each of them underestimates her actual loss. The former employee underestimates her loss because she thinks her optimal option is above the default by only $\Delta_R - b$ when in fact it's above

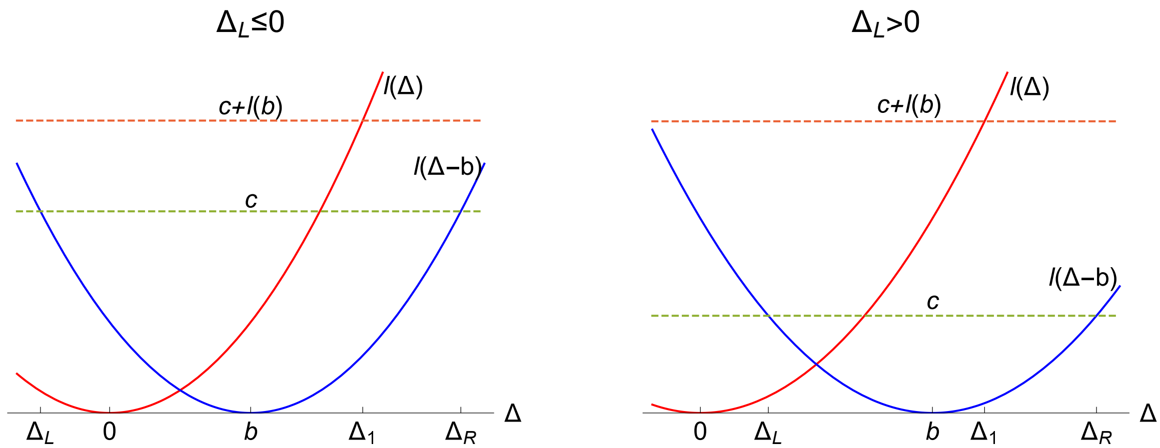

Figure 3: Actual loss, $l(\Delta)$, and perceived loss, $l(\Delta - b)$, from staying with the default.

the default by Δ_R. The latter employee underestimates her loss because she thinks she's choosing her optimal option when in fact she's choosing her optimal option minus b. Although both of these employees misjudge their optimal options by b, the misjudgement of the former employee comes on top of a perceived deviation of $\Delta_R - b$ from her optimal option while the misjudgement of the latter employee comes on top of a perceived deviation of 0 from her optimal option. Given the strict convexity of $l(\cdot)$, the misjudgement of the former employee translates into her underestimating her loss by more, so that she must be incurring the larger loss.

Statements 1) and 2) in Lemma 1 imply that $L(\Delta_L) < c + l(b)$, so that $L(\cdot)$ has a downward jump at Δ_L. Statement 3) says that $L(\cdot)$ has a downward jump at Δ_R. The downward jumps of $L(\cdot)$ at Δ_L and Δ_R will play a key role in determining the optimal default policy.

4 The Optimal Default Policy

It is assumed that the planner's objective it to minimise the average loss in the population. Before solving the planner's problem, let us observe that, given that employees' x's lie in $[\underline{x}, \overline{x}]$, employees' corresponding Δ's given a default D lie in $[\underline{x} - D, \overline{x} - D]$. Thus, by choosing D the planner is effectively shifting around the latter interval on the horizontal axis in each panel in Figure 2. Thus, denoting the upper endpoint of this interval as $\bar{\Delta}$ (i.e., $\bar{\Delta} = \overline{x} - D$), we can think of the planner as directly choosing $\bar{\Delta}$ rather than D. In fact, this turns out to be notationally more convenient. Moreover, given any $\bar{\Delta}$, it is easy to infer the position of the corresponding default, D, relative to $[\underline{x}, \overline{x}]$. In particular, the position of D relative to $[\underline{x}, \overline{x}]$ is the same as the position of $\Delta = 0$ relative to $[\bar{\Delta} - 2\epsilon, \bar{\Delta}]$. 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 $[\underline{x}, \overline{x}]$.

4.1 Active Decision

Before exploring how the planner optimally sets $\bar{\Delta}$, let us consider whether AD could ever be an optimal default policy.

Proposition 1 *AD is never an optimal default policy.*

The logic for Proposition 1 is the following. By the fact that $L(\Delta_L) < c + l(b)$ and the continuity of $l(\cdot)$, there exists $\eta > 0$, such that $L(\Delta_L + \eta) < c + l(b)$. If the planner sets $\bar{\Delta} = \Delta_L + \eta$, then (i) no employee incurs a loss strictly greater than $c + l(b)$ because $L(\Delta) \leq c + l(b)$ for all $\Delta \leq \Delta_L + \eta$ (see Figure 2) and (ii) employees with Δ's between Δ_L and $\Delta_L + \eta$ incur a loss strictly less than $c + l(b)$ (again, see Figure 2). Thus, the average loss in the population with $\bar{\Delta} = \Delta_L + \eta$ is strictly lower than the average loss with AD.

Note that factors omitted from the current model can make AD optimal.[19]

[19] For example, AD can be optimal because (i) employees procrastinate opting out of unsuitable defaults (Carrol et al. (2009)), (ii) the planner disregards the cost of deciding actively (Goldin and Reck (2019)), (iii) defaults are informative and undermine agents' willingness to gather information (Carlin et al. (2013)), or (iv) each default influences employees' beliefs in a blunt way and may make some employees more biased (Ivanov (2019)).

4.2 The Optimal Default

Given Proposition 1, an optimal default is an optimal default policy. We now turn our attention to the question of how the planner sets the default, or equivalently $\bar{\Delta}$, optimally.

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

$$\min_{\overline{x}-1 \leq \bar{\Delta} \leq \overline{x}} \frac{1}{2\epsilon} \int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} L(\Delta) d\Delta. \tag{2}$$

The constraint is equivalent to $0 \leq D \leq 1$, which reflects the natural constraint that the planner cannot set a default contribution rate or a default allocation to stocks outside of $[0, 1]$.

The following proposition characterises the solution to problem (2).

Proposition 2 *Let Δ_1 be the value of $\Delta > 0$, such that $l(\Delta) = c + l(b)$ (see Figure 3). If $l(\cdot)$ is symmetric and strictly convex, the unique solution to problem (2) is:*

$$\bar{\Delta}^*(b, c, \epsilon) = \begin{cases} \epsilon & \textit{if } \Delta_L \leq -\epsilon \\ \Delta_L + 2\epsilon & \textit{if } \Delta_L > -\epsilon, \Delta_1 - \Delta_L \geq 2\epsilon \\ \Delta_1 & \textit{if } \Delta_1 - \Delta_L < 2\epsilon \end{cases}. \tag{3}$$

The first case in expression (3) is illustrated in Figure 4. The shaded area in the figure represents the average loss in the population (scaled by 2ϵ). $\bar{\Delta} = \epsilon$ corresponds to the C default. As is evident from the figure, nobody opts out.

To see the intuition for why $\bar{\Delta} = \epsilon$ solves problem (2) when $\Delta_L \leq -\epsilon$, consider shifting $\bar{\Delta}$ slightly to the right or left away from ϵ in Figure 4. If we shift $\bar{\Delta}$ slightly to the right by a small number $\delta > 0$, the shaded area increases by the area under

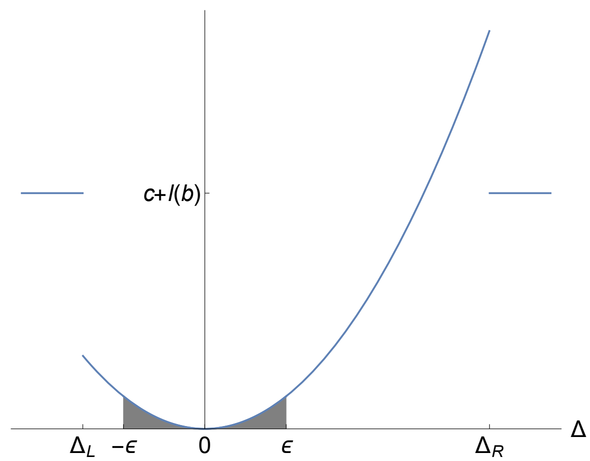

Figure 4: $\bar{\Delta}^*(b, c, \epsilon) = \epsilon$ (C default).

$L(\cdot)$ over the interval $[\epsilon, \epsilon + \delta]$ and decreases by the area under $L(\cdot)$ over the interval $[-\epsilon, -\epsilon + \delta]$. Because $L(\cdot)$ has higher values just to the right of ϵ than just to the right of $-\epsilon$, the shift to the right leads to a net increase in the shaded area and is thus undesirable. A similar logic reveals that shifting $\bar{\Delta}$ slightly to the left is undesirable. An analogous argument involving small shifts to the right or left of $\bar{\Delta}^*(b, c, \epsilon)$ will apply to the second and third cases in expression (3), so that we don't repeat the argument when discussing these cases below.

The second case in expression (3) is illustrated in the two panels of Figure 5 (one panel for $\Delta_L \leq 0$ and one for $\Delta_L > 0$). $\bar{\Delta} = \Delta_L + 2\epsilon$ corresponds to the $\underline{x}$-PEL default. As is evident from the figure, nobody opts out. Note that employees with low x's (i.e., with Δ's near Δ_L) incur a smaller loss than those with high x's (i.e., with Δ's near $\Delta_L + 2\epsilon$). However, increasing the default to better accommodate employees with high x's (by lowering $\bar{\Delta}$ below $\Delta_L + 2\epsilon$) is undesirable because it causes employees with low x's to opt out, which leads to even higher losses.

The third case in expression (3) is illustrated in the two panels of Figure 6 (one panel for $\Delta_L \leq 0$ and one for $\Delta_L > 0$).[20] $\bar{\Delta} = \Delta_1$ corresponds to the $\overline{x}$-EL default.

[20]The right panel is drawn so that $\Delta_1 - 2\epsilon > 0$. However, we could just as well have $\Delta_1 - 2\epsilon \leq 0$.

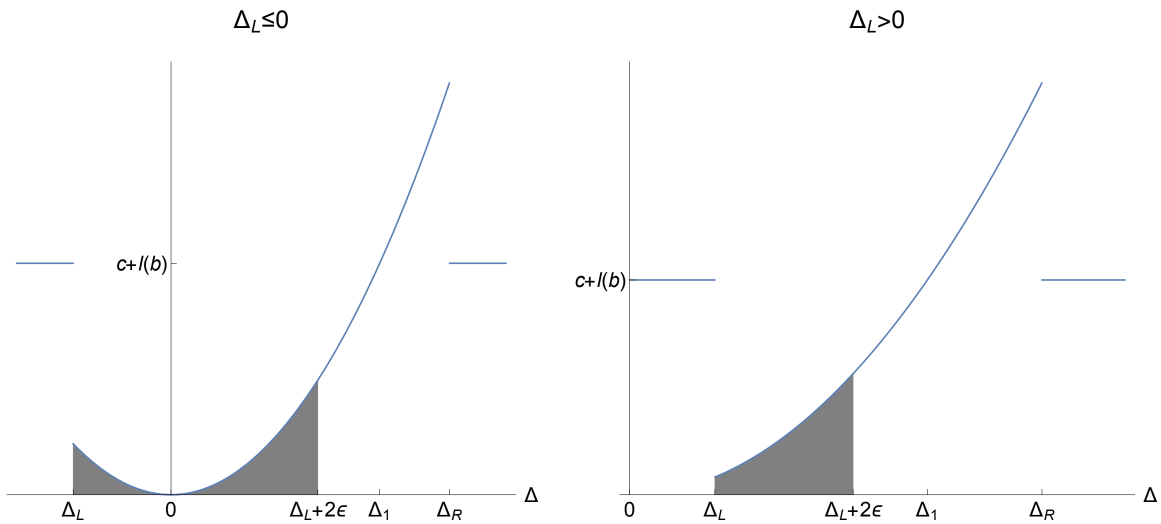

Figure 5: $\bar{\Delta}^*(b,c,\epsilon) = \Delta_L + 2\epsilon$ ($\underline{x}$-PEL default).

As is evident from the figure, employees with x's above a cutoff (i.e., with Δ's above Δ_L) stay while employees with x's below the cutoff (i.e., with Δ's below Δ_L) opt out. Note that the planner could induce more employees to stay with the default if she slightly lowers it, i.e., if she shifts $\bar{\Delta}$ slightly to the right of Δ_1.[21] However, this would lead employees with high x's to incur losses above $c + l(b)$, which would more than undo any gains from getting more employees with lower x's to stay.

5 How Does Each Parameter Affect the Average Loss?

How does each parameter affect the average loss in the population at the optimal default, $\frac{1}{2\epsilon}\int_{\bar{\Delta}^*(b,c,\epsilon)-2\epsilon}^{\bar{\Delta}^*(b,c,\epsilon)} L(\Delta)d\Delta$?

Proposition 3 *Assume that $l(\cdot)$ is differentiable on $(0,\infty)$, symmetric, and strictly*

[21]The proof of Proposition 2 shows that $\Delta_1 < \overline{x}$ in the third case in expression (3), so that shifting $\bar{\Delta}$ slightly to the right of Δ_1 won't violate the constraints in problem (2).

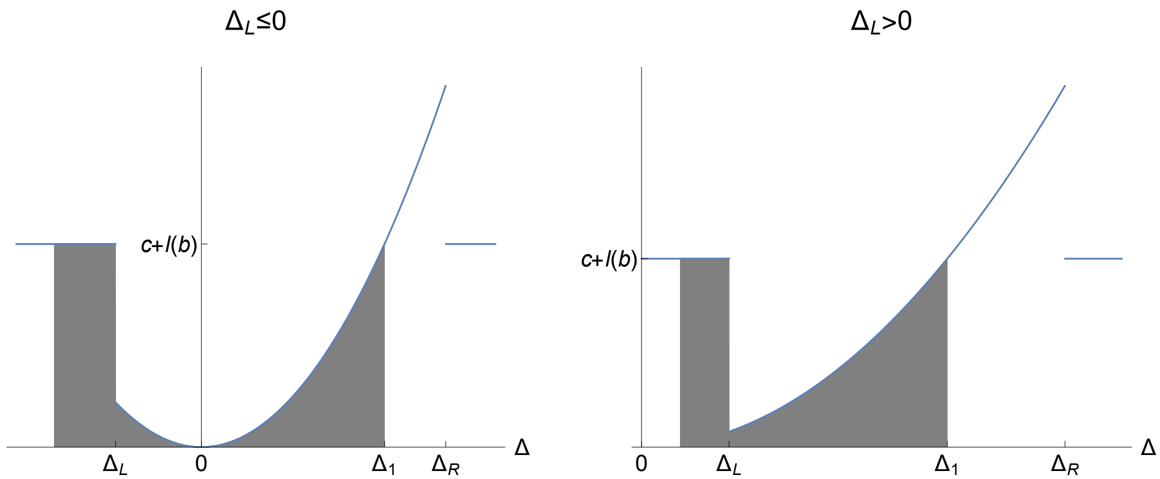

Figure 6: $\bar{\Delta}^*(b, c, \epsilon) = \Delta_1$ ($\overline{x}$-EL default).

convex. Then, $\frac{1}{2\epsilon}\int_{\bar{\Delta}^*(b,c,\epsilon)-2\epsilon}^{\bar{\Delta}^*(b,c,\epsilon)} L(\Delta)d\Delta$ *is weakly increasing in* b*, either weakly decreasing or nonmonotone in* c*, and strictly increasing in* ϵ*.*

The intuition for the comparative statics of $\frac{1}{2\epsilon}\int_{\bar{\Delta}^*(b,c,\epsilon)-2\epsilon}^{\bar{\Delta}^*(b,c,\epsilon)} L(\Delta)d\Delta$ with respect to b or c depends on whether the C, $\underline{x}$-PEL, or $\overline{x}$-EL default is optimal. When the C default is optimal, a small change in one of these parameters does not affect the shaded area in Figure 4.

When the $\underline{x}$-PEL default is optimal, consider Figure 5. A small increase in b shifts up slightly Δ_L (see Figure 1) and, hence, $\bar{\Delta}^*(b, c, \epsilon) = \Delta_L + 2\epsilon$. This results in eliminating a sliver of shaded area just to the right of Δ_L and expanding the shaded area just to the right of $\Delta_L + 2\epsilon$ in Figure 5. Given that $L(\cdot)$ has higher values around $\Delta_L + 2\epsilon$ than around Δ_L, the net effect on the shaded area is positive, so that the average loss increases. Analogously, a small increase in c shifts down slightly Δ_L (see Figure 1) and, hence, $\bar{\Delta}^*(b, c, \epsilon) = \Delta_L + 2\epsilon$, so that the average loss decreases.

When the $\overline{x}$-EL default is optimal, consider Figure 6. A small increase in b has two effects. First, it shifts up slightly Δ_L, so that employees just to the right of Δ_L in the figure opt out. Because $L(\Delta_L) < c + l(b)$, these employees' losses increase.

Second, it shifts the $c+l(b)$ line up, which increases the loss of employees with Δ's in $[\Delta_1 - 2\epsilon, \Delta_L]$, thus reinforcing the first effect.[22] A small increase in c has analogous effects, except that the first effect on the average loss is negative because Δ_L shifts slightly down rather than up. As a result the two effects have opposite signs and the overall effect of an increase in c on the average loss is ambiguous.

To see the intuition for why $\frac{1}{2\epsilon}\int_{\bar{\Delta}^*(b,c,\epsilon)-2\epsilon}^{\bar{\Delta}^*(b,c,\epsilon)} L(\Delta)d\Delta$ is strictly increasing in ϵ, observe the following. First, $\frac{1}{2\epsilon}\int_{\bar{\Delta}^*(b,c,\epsilon)-2\epsilon}^{\bar{\Delta}^*(b,c,\epsilon)} L(\Delta)d\Delta$ equals the average height of $L(\cdot)$ over the interval $[\bar{\Delta}^*(b,c,\epsilon)-2\epsilon, \bar{\Delta}^*(b,c,\epsilon)]$. Second, in Figure 4/5/6 increasing ϵ expands the interval $[\bar{\Delta}^*(b,c,\epsilon)-2\epsilon, \bar{\Delta}^*(b,c,\epsilon)]$ at both of its margins/at its right margin/at its left margin, where the value of $L(\cdot)$ exceeds the average value of this function over $[\bar{\Delta}^*(b,c,\epsilon)-2\epsilon, \bar{\Delta}^*(b,c,\epsilon)]$.

6 A Critical Look at the Model's Assumptions

In this section, we discuss some of the model's assumptions. First, consider the assumptions on $l(\cdot)$. Propositions 2 and 3 assume that $l(\cdot)$ is strictly convex and symmetric. Proposition 3 also assumes that $l(\cdot)$ is differentiable on $(0,\infty)$. We also assume throughout that the loss from deviations depends on x and x' only through the difference $x-x'$. For brevity, let us call this the difference assumption.

Although such assumptions are common in the literature, they are not derived from deeper assumptions about utility functions, intertemporal budget constraints, and how employees might change in future periods the option with which they end up today. Notably, in the context of contribution rates, employer-match caps and tax-bracket thresholds are likely to lead to violations of symmetry, differentiability,

[22] The increase in b also raises Δ_1 (see Figure 3), which results in transferring a sliver of shaded area from just to the right of $\Delta_1 - 2\epsilon$ to just to the right of Δ_1 in Figure 6. Because $L(\Delta_1 - 2\epsilon) = L(\Delta_1)$, this effect is approximately 0.

and the difference assumption.[23]

Second, Propositions 2 and 3 rely on the assumption of a uniform distribution of optimal options, which is also clearly a simplification. Notably, this assumption excludes the realistic possibility that there is a positive mass of employees with (i) optimal contribution rate equal to zero or to the employer-match cap or (ii) optimal allocation to stocks that equals zero or one.[24]

Third, it is assumed that the bias is additive rather than, for example, multiplicative. This assumption is made because it simplifies the analysis, but there is no clear justification for it.

Fourth, the model implicitly assumes away many potentially relevant factors mentioned in the introduction and discussed further in the appendix.

7 Policy Implications

Although the model is highly stylised, the theoretical analysis yields three lessons that might be useful to policy-makers. The first lesson is based on the discontinuity of $L(\cdot)$ at Δ_L and doesn't rely on the strict convexity or symmetry of $l(\cdot)$, or on the uniform distribution of employees' optimal options. Let F denote an arbitrary cumulative distribution function (CDF) on $[\underline{x}, \overline{x}]$. If F doesn't have an atom at $\underline{x}$, the planner's objective function[25] exhibits a kink at $\bar{\Delta} = \Delta_L + 2\epsilon$ with the left derivative

[23]For example, suppose there is an employer-match cap of m and let $\tilde{l}(x, x')$ denote the loss from deviations for an employee with optimal contribution rate x who ends up with contribution rate x'. In that case, $\tilde{l}(x, \cdot)$ probably exhibits a kink at m (so that differentiability fails) without a matching kink at $2x - m$ (so that symmetry around x fails). Moreover, given that the location of the kink does not depend on x, its position relative to x varies (so that the difference assumption fails). (I suspect that the conclusion of Proposition 3 continues to hold if $l(\cdot)$ is not differentiable at finitely many points, so that any failure of differentiability due to employer-match caps and tax-bracket thresholds may not be a problem.)

[24]The possibility of a positive mass of employees with optimal contribution rate equal to zero also strains the assumption that $b \leq \underline{x}$.

[25]For an arbitrary F, the planner's objective function in (2) needs to be rewritten as $\int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} L(\Delta) dF_{\bar{\Delta}}$, where $F_{\bar{\Delta}}$ is the CDF over Δ's induced by F and $\bar{\Delta}$.

being strictly smaller than the right derivative.[26] The intuition is that decreasing $\bar{\Delta}$ slightly below $\Delta_L + 2\epsilon$ induces workers with optimal options near $\underline{x}$ to opt out and discontinuously increases their losses (see Figure 5) while increasing $\bar{\Delta}$ slightly above $\Delta_L + 2\epsilon$ has no comparable effect. If F does have an atom at $\underline{x}$, the planner's objective function jumps upwards when $\bar{\Delta}$ is reduced slightly below $\Delta_L + 2\epsilon$. This occurs because decreasing $\bar{\Delta}$ slightly below $\Delta_L + 2\epsilon$ induces the positive mass of employees at $\underline{x}$ to opt out and discontinuously increases their losses. Whether the planner's objective function exhibits a kink at $\bar{\Delta} = \Delta_L + 2\epsilon$ with the left derivative being strictly smaller than the right derivative or it jumps up to the left of $\bar{\Delta} = \Delta_L + 2\epsilon$, the broad lesson is that the $\underline{x}$-PEL default is a prominent candidate for an optimal default. Note that, to know where the $\underline{x}$-PEL default lies, it isn't necessary to know Δ_L and ϵ. It suffices to have a sense of what is the default for which employees with the lowest optimal options are only just willing to stay with it.

The second lesson is based on the discontinuity of $L(\cdot)$ at Δ_R. This discontinuity means that, for values of $\bar{\Delta}$ for which some employees' Δ's lie between Δ_1 and Δ_R, these employees are staying with the default even though it is so far below their optimal options that each of them incurs a loss exceeding the loss of an employee who opts out. The broad lesson is to be mindful of the possibility that biased employees may stay with the default even though it is far from their optimal options, and to recognise that, in such a situation, it may be better to bring the default closer to these employees' optimal options even if that induces other employees to opt out. Note that this lesson runs counter to the approach, suggested by Thaler and Sunstein (2003), of choosing the default so as minimise the number of opt-outs. Note also that this lesson doesn't rely on the symmetry or differentiability of $l(\cdot)$, or on the uniform distribution of employees' optimal options.

[26] This statement relies on some assumptions (notably, differentiability of $l(\cdot)$ and a positive right derivative of F at $\underline{x}$). See Lemma 4 in the appendix.

The third lesson is based on Proposition 3. This proposition suggests that, if the planner can lower employees' bias (e.g., through education) or decrease heterogeneity in the targeted population (e.g., by customising different defaults for young vs. old employees), doing so is beneficial (or at least doesn't hurt). Although Proposition 3 makes strong assumptions, this lesson seems intuitive and it is reasonable to conjecture that it holds more generally.[27]

References

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): 267–319.

Bernheim, B. Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *Quarterly Journal of Economics* 124 (1): 51–104.

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

[27]However, if employees update their preferred options in the direction of the default as in Ivanov (2019), the average loss in the population at the optimal default policy can be nonmonotone in b and ϵ. The reasons for this are complex and are discussed in Ivanov (2019).

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 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.

Blumenstock, JE, Callen, M, and Ghani, T. 2018. "Why Do Defaults Affect Behavior? Experimental Evidence from Afghanistan." *American Economic Review*, 108 (10), 2868-2901

Bogachev, Vladimir I. "Measure theory. Vol. 1." Springer Science & Business Media, 2007.

Bronchetti, Erin Todd, Thomas S. Dee, David B. Huffman, and Ellen Magenheim. 2013. "When a Nudge Isn't Enough: Defaults and Saving among Low-Income Tax Filers." *National Tax Journal* 66 (3): 609–34.

Caplin, Andrew, and Daniel Martin. 2017. "Defaults and Attention: The Drop Out Effect." *Revue économique* 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 Economics* 124 (4): 1639–74.

Chetty, Raj, John N. Friedman, Søren 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, 67–113. 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, 81–126. 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, 304–51. New York: Russell Sage Foundation.

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

Ericson KM, Laibson D. 2018. “Intertemporal Choice.” National Bureau of Economic Research, working paper 25358

Gabaix, Xavier, and David Laibson. 2017. “Myopia and discounting. No.” National Bureau of Economic Research, working paper 23254.

Goldin, Jacob and Daniel Reck. 2019. “Optimal Defaults with Normative Ambiguity.” Working paper.

Ivanov, Asen. 2019. “Optimal Pension Plan Default Policies when Employees are Biased and Update from the Default.” Working paper.

Laibson, David. 2020. “Nudges are Not Enough: The Case for Price-Based Paternalism.” Presentation at a joint lucheon at the annual meetings of the American Economic Association and the American Finance Association in San Diego, California. A video recording of the presentation is available at https://www.aeaweb.org/webcasts/2020/aea-afa-joint-luncheon-nudges-are-not-enough

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): 1149–87.

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.

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

8 Appendix: How Do the Parameters Affect the Optimal Default?

In this appendix, we explore how the parameters determine whether the C, $\underline{x}$-PEL, or $\overline{x}$-EL default is the optimal one. Let us start with the following lemma.

Lemma 2 *If $l(\cdot)$ is strictly convex, $\Delta_1 - \Delta_L$ is strictly decreasing in b and strictly increasing in c.*

The logic for why $\Delta_1 - \Delta_L$ is strictly decreasing in b is the following. Increasing b by $\sigma > 0$ shifts the $c+l(b)$ line in each panel of Figure 3 up by $l(b+\sigma)-l(b)$. Given the fact that $\Delta_1 > b$[28] and the strict convexity of $l(\cdot)$, $l(\Delta_1+\sigma)-l(\Delta_1) > l(b+\sigma)-l(b)$, so that Δ_1 needs to increase by less than σ to accommodate the increase in b. Thus, Δ_1 increases less than one-for-one with b whereas Δ_L increases one-for-one with b (see Figure 1).

That $\Delta_1 - \Delta_L$ is strictly increasing in c follows from the fact that Δ_1 is strictly increasing in c and the fact that Δ_L is strictly decreasing in c (see Figure 3).

We are ready to state the main result in this section.

Proposition 4 *If $l(\cdot)$ is symmetric and strictly convex, the following hold.*

1) *Increasing b contracts/expands the set of (c, ϵ)-pairs for which the C/$\overline{x}$-EL default is optimal.*

2) *Increasing c expands/contracts the set of (b, ϵ)-pairs for which the C/$\overline{x}$-EL default is optimal.*

3) *Increasing ϵ contracts/expands the set of (b, c)-pairs for which the C/$\overline{x}$-EL default is optimal.*[29]

[28] $\Delta_1 > 0$ and $l(\Delta_1) = c + l(b) > l(b)$ imply $\Delta_1 > b$.

[29] The terms "contracts" and "expands" are used in the sense of strict set inclusion.

Statement 1) follows because increasing b (i) increases Δ_L (see Figure 1), thus making the inequality in the first case in expression (3) harder to satisfy, and (ii) by Lemma 2, decreases $\Delta_1 - \Delta_L$, thus making the inequality in the third case in expression (3) easier to satisfy.

Statement 2) holds because increasing c (i) decreases Δ_L (see Figure 1), thus making the inequality in the first case in expression (3) easier to satisfy, and (ii) by Lemma 2, increases $\Delta_1 - \Delta_L$, thus making the inequality in the third case in expression (3) harder to satisfy.

Statement 3) follows because increasing ϵ makes the inequality in the first/third case in expression (3) harder/easier to satisfy.[30]

Loosely speaking, Proposition 4 says that increasing b or ϵ tilts the planner away from the C default and towards the $\overline{x}$-EL default, whereas increasing c has the opposite effect.

To derive some additional relationships between the parameters and the optimal default, it is convenient to rewrite expression (3) as follows:

$$\bar{\Delta}^*(b,c,\epsilon) = \begin{cases} \epsilon & \text{if } \epsilon \leq -\Delta_L \\ \Delta_L + 2\epsilon & \text{if } -\Delta_L < \epsilon \leq \frac{\Delta_1 - \Delta_L}{2} \\ \Delta_1 & \text{if } \epsilon > \frac{\Delta_1 - \Delta_L}{2} \end{cases} . \tag{4}$$

Assuming that $l(\cdot)$ is symmetric and letting $l^{-1}(\cdot)$ denote the inverse of $l(\cdot)$ on $(0,\infty)$, we have $\Delta_L = -l^{-1}(c)+b$ (see Figure 1). The following lemma, in conjunction with Lemma 2, will be useful in discussing how the parameters affect the conditions in expression (4).

[30]For each parameter, increasing it makes one of the inequalities in the second case in expression (3) easier to satisfy whilst making the other inequality in the second case in expression (3) harder to satisfy. As a result, it is not the case for any parameter that increasing it unambiguously contracts or expands the set of pairs of the other two parameters for which the $\underline{x}$-PEL default is optimal.

Lemma 3 *Assume that $l(\cdot)$ is symmetric and strictly convex. Holding c fixed:*

1) $\lim_{b\downarrow 0} \frac{\Delta_1 - \Delta_L}{2} = l^{-1}(c)$*;*

2) $\lim_{b\to\infty} \frac{\Delta_1 - \Delta_L}{2} \geq \frac{l^{-1}(c)}{2}$.

Figure 7 shows the optimal default as a function of b and ϵ, holding c fixed. In particular, the C default is optimal below the line plotting $-\Delta_L$ as a function of b; the $\overline{x}$-EL default is optimal above the line plotting $\frac{\Delta_1-\Delta_L}{2}$ as a function of b; the $\underline{x}$-PEL default is optimal between the two lines. Based on this figure, the following relationships hold between the parameters and the optimal default.

1. Fixing c and ϵ, such that $\epsilon \leq \lim_{b\to\infty} \frac{\Delta_1-\Delta_L}{2}$, there is a cutoff value of b, such that the C/$\underline{x}$-PEL default is optimal for b below/above the cutoff.

2. Fixing c and ϵ, such that $\lim_{b\to\infty} \frac{\Delta_1-\Delta_L}{2} < \epsilon < l^{-1}(c)$, there are two cutoff values of b, such that the C/$\underline{x}$-PEL/$\overline{x}$-EL default is optimal for b below the lower cutoff/between the two cutoffs/above the higher cutoff.

3. Fixing c and ϵ, such that $\epsilon \geq l^{-1}(c)$, the $\overline{x}$-EL default is optimal for all b.

4. Fixing c and b, such that $b < l^{-1}(c)$, there are two cutoff values of ϵ, such that the C/$\underline{x}$-PEL/$\overline{x}$-EL default is optimal for ϵ below the lower cutoff/between the two cutoffs/above the higher cutoff.

5. Fixing c and b, such that $b \geq l^{-1}(c)$, there is a cutoff value of ϵ, such the $\underline{x}$-PEL/$\overline{x}$-EL default is optimal for ϵ below/above the cutoff.

6. Fixing b and ϵ, there are two cutoff values of c, such that the C/$\underline{x}$-PEL/$\overline{x}$-EL default is optimal for c above the higher cutoff/between the two cutoffs/below the lower cutoff. (To see this, observe that increasing c increases the $l^{-1}(c)$ intercept on each axis in Figure 7.)

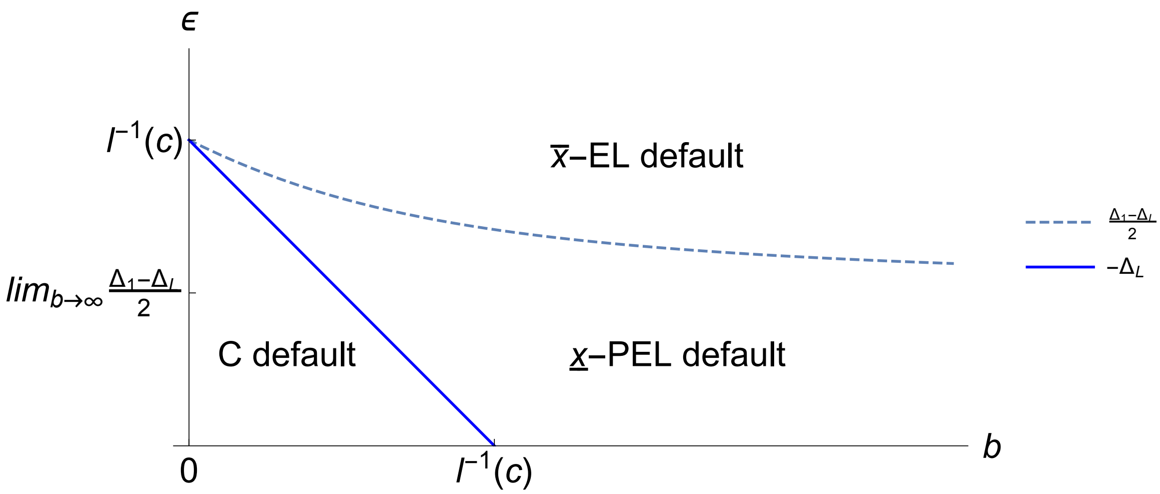

Figure 7: Optimal default as a function of b and ϵ, holding c fixed.

9 Appendix: Omitted Factors

In this section, we discuss some potentially relevant factors that are omitted from the model.

9.1 Normatively Irrelevant Cost of Deciding Actively

Within a standard model, employees' empirical reluctance to switch from default contribution rates implies opt-out costs in the thousands of dollars. Such opt-out costs seem orders of magnitude higher than what is justified based on the value of employees' time.[31] Motivated by these observations, Goldin and Reck (2019) allow the planner to count a part of each employee's cost of deciding actively as normatively irrelevant and to exclude it from the social welfare function.

[31] See Goldin and Reck (2019) and Section IIB in Bernheim et al. (2015).

9.2 Updating from the Default

In the context of asset allocation, there is evidence that, upon seeing a default, employees update their preferred options in the direction of the default. In particular, 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 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.

In the context of contribution rates, the evidence is more mixed. On the one hand, 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. On the other hand, if employees sufficiently update their preferred contribution rates based on the default, this would explain their reluctance to opt out of the default contribution rate without any need to invoke exorbitant opt-out costs (see Section II D in Bernheim et al. (2015)).[32]

Ivanov (2019) modifies the model in the current paper by allowing for updating from the default under the assumption of a linear loss from deviations. In an extension to their baseline model, Goldin and Reck (2019) also allow for updating from the default.

[32] Thus, in the presence of sufficient updating from the default, there is no need to consider a part of the cost of deciding actively as normatively irrelevant.

9.3 Procrastination

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. Given that the model in the current paper is static, it cannot address the issue of procrastination. On the other hand, in Carroll et al. (2009), employees are not biased regarding their optimal options.

9.4 Normative Ambiguity

The model here assumes that, for each employee, there is an option x, which is unambiguously optimal for her. This assumption is nontrivial. For example, suppose employees are dynamically inconsistent, so that they save differently (i) if they decide in each period how much to save for that period vs. (ii) if they can commit in the current period how much to save in the current period and in all future periods. Although many papers assume that the savings decisions under (i) are biased and under (ii) are correct, the justification for this is unclear.[33]

Bernheim and Rangel (2009) develop a framework for conducting welfare analysis in the case of such normative ambiguity. Bernheim et al. (2015) apply this framework to the issue of optimal default policies in the context of contribution rates. Bernheim et al. allow for normative ambiguity regarding whether employees should opt out of the default. However, their analysis doesn't allow us to say much about the optimal default policy if employees misperceive their optimal options or if there is normative ambiguity regarding their optimal options.[34]

[33] See Bernheim and Rangel (2009) and Bernheim (2009).

[34] In their model with partially naive β-δ employees and their model with inattentive employees, each employee has a unique, unbiased preferred contribution rate. In their anchoring model, unless we are willing to assume that each employee correctly perceives her optimal option under AD but misperceives her optimal option when faced with a default, there is normative ambiguity regarding each employee's preferred contribution rate. However, this normative ambiguity prevents us from saying much about the optimal default policy.

9.5 Decision Costs

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,[35] it is tempting to reinterpret c as reflecting both implementation and decision costs. However, this is problematic because of the following features of the 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.

There are papers which explore the implications of decision costs for the optimal default policy. In Carlin et al. (2013) and Wisson (2016), 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.[36] 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.

[35] Based on a field experiment in Afghanistan, Blumenstock et al. (2018) conclude that decision costs are a major reason for why the default savings rate has a large effect on saving in a savings account.

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

9.6 The Presence of Sophisticated Employees

In the model here, all employees share the same parameters b and c. Such homogeneity is probably unrealistic.

One way to incorporate heterogeneity is to assume that a fraction, ρ, of the population is sophisticated in the sense that they have zero bias. Sophisticated employees can also have their own cost of deciding actively, which would be smaller than the cost of deciding actively for unsophisticated employees. Given that sophisticated employees would be 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.

10 Appendix: Kink in the Planner's Objective Function at $\bar{\Delta} = \Delta_L + 2\epsilon$

Lemma 4 below formalises the statement made in section 7 about a kink in the planner's objective function at $\bar{\Delta} = \Delta_L + 2\epsilon$.

Let $F_{\bar{\Delta}}$ denote the CDF over Δ's induced by F and $\bar{\Delta}$. I.e., $F_{\bar{\Delta}}$ is defined by $F_{\bar{\Delta}}(\Delta) = F(\underline{x} + \Delta - \bar{\Delta} + 2\epsilon)$. Let $Z(\bar{\Delta}) = \int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} L(\Delta) dF_{\bar{\Delta}}$ denote the generalisation of the planner's objective function in (2).

Lemma 4 *Assume that F can be written as $F = \sum_{k=1}^{K} \alpha_k G_k + (1 - \sum_{k=1}^{K} \alpha_k) G$, where $\alpha_k \geq 0$ for all k, $\sum_{k=1}^{K} \alpha_k < 1$, each G_k is a CDF that is degenerate on some x in $[\underline{x}, \overline{x}]$, and G is a CDF with a continuously differentiable probability density function, g, such that $g(\underline{x}) > 0$.*[37] *Further assume that, for no k, G_k is degenerate on $\underline{x}$ or on $\underline{x} + \Delta_R - \Delta_L$. Finally, assume that $\Delta_L + 2\epsilon \neq \Delta_R$ and that $l(\cdot)$ is differentiable.*

[37] This formulation allows F to have finitely many atoms.

Then, $Z(\cdot)$ has a left and right derivative at $\bar{\Delta} = \Delta_L + 2\epsilon$, and the left derivative is strictly smaller than the right derivative.

11 Appendix: Proofs

11.1 Proof of Lemma 1

The logic given straight after the lemma in the main text is rigorous. Nevertheless, it might be worth restating in more formal notation the arguments given there.

If $\Delta_L \leq 0$, we have $L(\Delta_L) = l(\Delta_L) < l(\Delta_L - b) = c$. The inequality follows because $\Delta_L - b < \Delta_L \leq 0$ and $l(\cdot)$ is strictly decreasing over $(-\infty, 0]$.

If $\Delta_L > 0$, we have $L(\Delta_L) = l(\Delta_L) < l(b)$. The inequality follows because $0 < \Delta_L < b$ and $l(\cdot)$ is strictly increasing over $[0, \infty)$.

Finally, $L(\Delta_R) = l(\Delta_R) = c + l(\Delta_R) - l(\Delta_R - b) > c + l(b) - l(0) = c + l(b)$. The second equality holds because $l(\Delta_R - b) = c$ (by the definition of Δ_R). The inequality holds because (i) the interval $[\Delta_R - b, \Delta_R]$ represents a rightward translation of the interval $[0, b]$ and (ii) by the strict convexity of $l(\cdot)$, the increase in $l(\cdot)$ must be larger over the former interval. Q.E.D.

11.2 Proof of Proposition 1

The logic given straight after the proposition is rigorous. Q.E.D.

11.3 Proof of Proposition 2

Let us first show that $\bar{\Delta}$, such that $\bar{\Delta} < \Delta_L$ or $\bar{\Delta} > \Delta_R$, cannot be a solution to problem (2) because setting $\bar{\Delta} = \Delta_1$ is strictly better. If $\bar{\Delta} < \Delta_L$ or $\bar{\Delta} > \Delta_R$, we have

$$
\begin{aligned}
&\int_{[\bar{\Delta}-2\epsilon,\bar{\Delta}]} L(\Delta)d\Delta = \\
&\int_{[\bar{\Delta}-2\epsilon,\bar{\Delta}]\cap[\Delta_1-2\epsilon,\Delta_1]} L(\Delta)d\Delta + \int_{[\bar{\Delta}-2\epsilon,\bar{\Delta}]\cap[\Delta_1-2\epsilon,\Delta_1]^c} L(\Delta)d\Delta = \\
&\int_{[\bar{\Delta}-2\epsilon,\bar{\Delta}]\cap[\Delta_1-2\epsilon,\Delta_1]} L(\Delta)d\Delta + \int_{[\bar{\Delta}-2\epsilon,\bar{\Delta}]\cap[\Delta_1-2\epsilon,\Delta_1]^c} L(\Delta)d\Delta + \\
&\int_{[\bar{\Delta}-2\epsilon,\bar{\Delta}]^c\cap[\Delta_1-2\epsilon,\Delta_1]} L(\Delta)d\Delta - \int_{[\bar{\Delta}-2\epsilon,\bar{\Delta}]^c\cap[\Delta_1-2\epsilon,\Delta_1]} L(\Delta)d\Delta = \\
&\int_{[\Delta_1-2\epsilon,\Delta_1]} L(\Delta)d\Delta + \int_{[\bar{\Delta}-2\epsilon,\bar{\Delta}]\cap[\Delta_1-2\epsilon,\Delta_1]^c} L(\Delta)d\Delta - \int_{[\bar{\Delta}-2\epsilon,\bar{\Delta}]^c\cap[\Delta_1-2\epsilon,\Delta_1]} L(\Delta)d\Delta > \\
&\int_{[\Delta_1-2\epsilon,\Delta_1]} L(\Delta)d\Delta + \int_{[\bar{\Delta}-2\epsilon,\bar{\Delta}]\cap[\Delta_1-2\epsilon,\Delta_1]^c} (c+l(b))d\Delta - \int_{[\bar{\Delta}-2\epsilon,\bar{\Delta}]^c\cap[\Delta_1-2\epsilon,\Delta_1]} (c+l(b))d\Delta = \\
&\int_{[\Delta_1-2\epsilon,\Delta_1]} L(\Delta)d\Delta,
\end{aligned}
$$

where the superscript "c" on an interval denotes its complement. The last equality holds because because the two intervals $[\bar{\Delta}-2\epsilon,\bar{\Delta}]\cap[\Delta_1-2\epsilon,\Delta_1]^c$ and $[\bar{\Delta}-2\epsilon,\bar{\Delta}]^c\cap[\Delta_1-2\epsilon,\Delta_1]$ have equal length. To see that the inequality holds, let's consider separately the cases (i) $\bar{\Delta}<\Delta_L$ or $\bar{\Delta}-2\epsilon\geq\Delta_R$ and (ii) $\bar{\Delta}>\Delta_R$ and $\bar{\Delta}-2\epsilon<\Delta_R$. In case (i), $L(\Delta)=c+l(b)$ for all $\Delta\in[\bar{\Delta}-2\epsilon,\bar{\Delta}]$ (and, hence, for all $\Delta\in[\bar{\Delta}-2\epsilon,\bar{\Delta}]\cap[\Delta_1-2\epsilon,\Delta_1]^c$), $L(\Delta)\leq c+l(b)$ for all $[\Delta_1-2\epsilon,\Delta_1]$ (and, hence, for all $\Delta\in[\bar{\Delta}-2\epsilon,\bar{\Delta}]^c\cap[\Delta_1-2\epsilon,\Delta_1]$), and $L(\Delta)<c+l(b)$ for a positive measure of Δ's in $[\bar{\Delta}-2\epsilon,\bar{\Delta}]^c\cap[\Delta_1-2\epsilon,\Delta_1]$. In case (ii), $L(\Delta)\geq c+l(b)$ for all $\Delta\in[\bar{\Delta}-2\epsilon,\bar{\Delta}]\cap[\Delta_1-2\epsilon,\Delta_1]^c$, $L(\Delta)>c+l(b)$ for a positive measure of Δ's in $[\bar{\Delta}-2\epsilon,\bar{\Delta}]\cap[\Delta_1-2\epsilon,\Delta_1]^c$, and $L(\Delta)\leq c+l(b)$ for all $[\Delta_1-2\epsilon,\Delta_1]$ (and, hence, for all $\Delta\in[\bar{\Delta}-2\epsilon,\bar{\Delta}]^c\cap[\Delta_1-2\epsilon,\Delta_1]$). Thus, we can safely assume that any solution to problem (2) is in $[\Delta_L,\Delta_R]$.

Next, we show that $\bar{\Delta}^*(b, c, \epsilon)$ given in expression (3) is the unique solution to problem (2). We have:

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

It is apparent from Figures 4-6 that $L(\Delta) < L(\Delta - 2\epsilon)$ for all $\Delta_L \leq \Delta < \bar{\Delta}^*(b, c, \epsilon)$. Thus, $\int_{-\infty}^{\bar{\Delta}} \big(L(\Delta) - L(\Delta - 2\epsilon)\big)d\Delta$ is strictly decreasing in $\bar{\Delta}$ over $[\Delta_L, \bar{\Delta}^*(b, c, \epsilon)]$. Similarly, it is apparent from Figures 4-6 that $L(\Delta) > L(\Delta - 2\epsilon)$ for all $\bar{\Delta}^*(b, c, \epsilon) < \Delta \leq \Delta_R$. Thus, $\int_{-\infty}^{\bar{\Delta}} \big(L(\Delta) - L(\Delta - 2\epsilon)\big)d\Delta$ is strictly increasing in $\bar{\Delta}$ over $[\bar{\Delta}^*(b, c, \epsilon), \Delta_R]$. Thus, $\int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} L(\Delta)d\Delta$ has a unique minimum over $[\Delta_L, \Delta_R]$ at $\bar{\Delta} = \bar{\Delta}^*(b, c, \epsilon)$.

It remains to verify that $\bar{\Delta}^*(b, c, \epsilon)$ satisfies $\overline{x} - 1 \leq \bar{\Delta}^*(b, c, \epsilon) \leq \overline{x}$. The first inequality holds because $\bar{\Delta}^*(b, c, \epsilon) > 0$ (see Figures 4-6) and $\overline{x} \leq 1$. In the first case in expression (3), $\bar{\Delta}^*(b, c, \epsilon) = \epsilon < \overline{x}$. In the second case in expression (3), $\bar{\Delta}^*(b, c, \epsilon) = \Delta_L + 2\epsilon < b + 2\epsilon \leq \overline{x}$, where the last inequality follows from our assumption $b \leq \underline{x}$. In the third case in expression (3), $\bar{\Delta}^*(b, c, \epsilon) = \Delta_1 = \Delta_1 - \Delta_L + \Delta_L < 2\epsilon + \Delta_L < 2\epsilon + b \leq \overline{x}$, where the last inequality follows from our assumption $b \leq \underline{x}$. Q.E.D.

11.4 Proof of Lemma 2

The logic given straight after the lemma in the main text is rigorous. Nevertheless, it might be worth restating in more formal notation the arguments given there for why

$\Delta_1 - \Delta_L$ is strictly decreasing in b.

Let us write $\Delta_L(b,c)$ and $\Delta_1(b,c)$ to make explicit the dependence of Δ_L and Δ_1 on b and c. Given the fact that $\Delta_1(b,c) > b$ and the convexity of $l(\cdot)$, for any $\sigma > 0$ we have:

$$l(\Delta_1(b,c) + \sigma) - l(\Delta_1(b,c)) > l(b+\sigma) - l(b),$$

so that

$$l(\Delta_1(b,c) + \sigma) > l(\Delta_1(b,c)) + l(b+\sigma) - l(b) = c + l(b+\sigma),$$

where the equality follows from $l(\Delta_1(b,c)) = c + l(b)$.

Given that $l(\Delta_1(b,c)) < c + l(b+\sigma) < l(\Delta_1(b,c) + \sigma)$, it must be that $\Delta_1(b,c) < \Delta_1(b+\sigma,c) < \Delta_1(b,c) + \sigma$. Thus,

$$\Delta_1(b+\sigma,c) - \Delta_L(b+\sigma,c) - (\Delta_1(b,c) - \Delta_L(b,c)) < \sigma - \Delta_L(b+\sigma,c) + \Delta_L(b,c) = 0.$$

The equality follows from the fact that increasing the bias from b to $b+\sigma$ increases Δ_L one-for-one (see Figure 1). Q.E.D.

11.5 Proof of Lemma 3

Statement 1) follows because, as b goes to 0, Δ_1 is arbitrarily close to the positive root of $l(\Delta) = c$ (see Figure 3) and $\Delta_L = -l^{-1}(c) + b$ is arbitrarily close to $-l^{-1}(c)$.

Statement 2) hold because $\Delta_1 - \Delta_L$ is strictly decreasing in b (see Lemma 2) and $\Delta_1 - \Delta_L \geq b - \Delta_L = l^{-1}(c)$. Q.E.D.

11.6 Proof of Proposition 3

The proof makes use of a sequence of claims. Before we launch into these, let us introduce some additional notation.

Let $\Psi(b,c,\epsilon) = \frac{1}{2\epsilon}\int_{\bar{\Delta}^*(b,c,\epsilon)-2\epsilon}^{\bar{\Delta}^*(b,c,\epsilon)} L(\Delta)d\Delta$. Let Ψ_p denote the partial derivative of Ψ with respect to the parameter $p \in \{b, c, \epsilon\}$. Let $-p$ denote the two parameters other than p and, with the usual abuse of notation, we will write $\bar{\Delta}^*(p,-p)$, $\Psi(p,-p)$, and $\Psi_p(p,-p)$. Let $p(-p,i)$ denote the set of values of p that satisfy, given fixed $-p$, the inequalities in the i^{th} case in expressions (3) and (4). Finally, let $l'(\cdot)$ denote the derivative of $l(\cdot)$ and let $l^{-1}(\cdot)$ denote the inverse of $l(\cdot)$ on $(0,\infty)$.

Claim 1

$$\Psi(b,c,\epsilon) = \frac{1}{2\epsilon} \times \begin{cases} \int_{-\epsilon}^{\epsilon} l(\Delta)d\Delta & \text{if } \epsilon \le -\Delta_L \\ \int_{\Delta_L}^{\Delta_L+2\epsilon} l(\Delta)d\Delta & \text{if } -\Delta_L < \epsilon \le \frac{\Delta_1-\Delta_L}{2} \\ (\Delta_L - \Delta_1 + 2\epsilon)(c + l(b)) + \int_{\Delta_L}^{\Delta_1} l(\Delta)d\Delta & \text{if } \epsilon > \frac{\Delta_1-\Delta_L}{2} \end{cases} . \tag{5}$$

Proof:

The claim follows directly from Figures 4-6. Q.E.D.

Claim 2

1) One of the following holds.

(i) $b(-b,1) = b(-b,2) = \emptyset$ *and* $b(-b,3) = (0,\infty)$ *or*

(ii) $b(-b,1) = (0,b_1]$, $b(-b,2) = (b_1,\infty)$, *and* $b(-b,3) = \emptyset$, *where* $b_1 > 0$, *or*

(iii) $b(-b,1) = (0,b_1]$, $b(-b,2) = (b_1,b_2]$, *and* $b(-b,3) = (b_2,\infty)$, *where* $0 < b_1 < b_2$.

2) $c(-c,3) = (0,c_1)$, $c(-c,2) = [c_1,c_2)$, *and* $c(-c,1) = [c_2,\infty)$, *where* $0 < c_1 < c_2$.

3) One of the following holds.

(i) $\epsilon(-\epsilon, 1) = \emptyset$, $\epsilon(-\epsilon, 2) = (0, \epsilon_1]$, *and* $\epsilon(-\epsilon, 3) = (\epsilon_1, \infty)$, *where* $\epsilon_1 > 0$, *or*

(ii) $\epsilon(-\epsilon, 1) = (0, \epsilon_1]$, $\epsilon(-\epsilon, 2) = (\epsilon_1, \epsilon_2]$, *and* $\epsilon(-\epsilon, 3) = (\epsilon_2, \infty)$, *where* $0 < \epsilon_1 < \epsilon_2$.

Proof:

The claim follows from expression (4) and Figure 7, which plots the $-\Delta_L$ and $\frac{\Delta_1 - \Delta_L}{2}$ curves as functions of b, holding c fixed. (The key features of the curves in the figure are based on Lemmas 2 and 3.) In particular, looking at the figure, it is straightforward that statements 1) and 3) hold. To see that statement 2) also holds, observe that increasing c increases the $l^{-1}(c)$ intercept on each axis in Figure 7.[38] Q.E.D.

Claim 3 *Assume that* $l(\cdot)$ *is differentiable on* $(0, \infty)$. *Then, the following hold.*

1) $l'(\cdot) > 0$ *on* $(0, \infty)$.

2) Δ_L *and* Δ_1 *are continuously differentiable in each parameter.*

3) $\frac{\partial \Delta_L}{\partial c} < 0$.

Proof:

Given that $l(\cdot)$ is increasing on $(0, \infty)$, we have $l'(\cdot) \geq 0$ on $(0, \infty)$. Given that $l(\cdot)$ is strictly convex, $l'(\cdot)$ is strictly increasing, so that it cannot equal zero anywhere on $(0, \infty)$.

Next, consider statement 2). Given that $l(\cdot)$ is differentiable on $(0, \infty)$ and convex, it is continuously differentiable on $(0, \infty)$. Given that $l(\cdot)$ is continuously differentiable on $(0, \infty)$ and $l'(\cdot) \neq 0$ on $(0, \infty)$, $l^{-1}(\cdot)$ is continuously differentiable on $(0, \infty)$.

[38] Figure 7 doesn't make clear whether various intervals in Claim 2 should include their endpoints. However, this can easily be ascertained by consulting expression (4).

Given that $\Delta_L = -l^{-1}(c) + b$ and $\Delta_1 = l^{-1}(c + l(b))$, Δ_L and Δ_1 are continuously differentiable in each parameter.

Statement 3) follows because $\Delta_L = -l^{-1}(c) + b$ and the derivative of $l^{-1}(\cdot)$ evaluated at c equals $1/l'(l^{-1}(c)) > 0$. Q.E.D.

Claim 4 *The maximand in problem (2) is continuous in* $(\bar{\Delta}, p)$*.*

Proof:

When $p = \epsilon$, the integrand does not depend on p. The claim follows from the continuity of the definite integral with respect to the bounds of integration.

From here on, we restrict attention to $p \in \{b, c\}$. We write $L(\Delta, p)$, $\Delta_L(p)$, and $\Delta_R(p)$ to make explicit the dependence on p of expression (1), Δ_L, and Δ_R, respectively.

Take $\sigma > 0$. We need to find $\delta > 0$, such that $|\int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} L(\Delta, p)d\Delta - \int_{\bar{\Delta}'-2\epsilon}^{\bar{\Delta}'} L(\Delta, p')d\Delta| < \sigma$ whenever $(\bar{\Delta}', p')$ is within δ Euclidean distance of $(\bar{\Delta}, p)$.

Let $M = 1 + \sup\limits_{\Delta} L(\Delta, p)$ and $r = min(\frac{\sigma}{8M}, 1)$. Also, let I, J and K denote the sets $[\bar{\Delta} - 2\epsilon, \bar{\Delta}]$, $[\bar{\Delta}' - 2\epsilon, \bar{\Delta}']$, and $[\Delta_L(p) - r/4, \Delta_L(p) + r/4] \cup [\Delta_R(p) - r/4, \Delta_R(p) + r/4]$, respectively. We have:

$$
\begin{aligned}
&|\int\limits_I L(\Delta, p)d\Delta - \int\limits_J L(\Delta, p')d\Delta| = \\
&|\int\limits_{I\setminus K} L(\Delta, p)d\Delta + \int\limits_{I\cap K} L(\Delta, p)d\Delta - \int\limits_{J\setminus K} L(\Delta, p')d\Delta - \int\limits_{J\cap K} L(\Delta, p')d\Delta| \leq \\
&|\int\limits_{I\setminus K} L(\Delta, p)d\Delta - \int\limits_{J\setminus K} L(\Delta, p')d\Delta| + \int\limits_{I\cap K} L(\Delta, p)d\Delta + \int\limits_{J\cap K} L(\Delta, p')d\Delta \leq \\
&|\int\limits_{I\setminus K} L(\Delta, p)d\Delta - \int\limits_{J\setminus K} L(\Delta, p')d\Delta| + \int\limits_{K} L(\Delta, p)d\Delta + \int\limits_{K} L(\Delta, p')d\Delta
\end{aligned}
$$

Clearly, $\sup_{\Delta \in K} L(\Delta, p) < M$. Also, we can choose δ small enough that $\sup_{\Delta \in K} L(\Delta, p') < M$. In that case, the last expression is smaller than:

$$
\begin{aligned}
&|\int_{I \setminus K} L(\Delta, p) d\Delta - \int_{J \setminus K} L(\Delta, p') d\Delta| + 2rM \leq \\
&|\int_{I \setminus K} L(\Delta, p) d\Delta - \int_{J \setminus K} L(\Delta, p') d\Delta| + \frac{\sigma}{4} = \\
&|\int_{I \setminus (K \cup J)} L(\Delta, p) d\Delta + \int_{(I \setminus K) \cap J} L(\Delta, p) d\Delta - \int_{J \setminus (K \cup I)} L(\Delta, p') d\Delta - \int_{(J \setminus K) \cap I} L(\Delta, p') d\Delta| + \frac{\sigma}{4} \leq \\
&|\int_{(I \setminus K) \cap J} L(\Delta, p) d\Delta - \int_{(J \setminus K) \cap I} L(\Delta, p') d\Delta| + \int_{I \setminus (K \cup J)} L(\Delta, p) d\Delta + \int_{J \setminus (K \cup I)} L(\Delta, p') d\Delta + \frac{\sigma}{4} = \\
&|\int_{(I \cap J) \setminus K} (L(\Delta, p) - L(\Delta, p')) d\Delta| + \int_{I \setminus (K \cup J)} L(\Delta, p) d\Delta + \int_{J \setminus (K \cup I)} L(\Delta, p') d\Delta + \frac{\sigma}{4} \leq \\
&|\int_{(I \cap J) \setminus K} (L(\Delta, p) - L(\Delta, p')) d\Delta| + \int_{I \setminus J} L(\Delta, p) d\Delta + \int_{J \setminus I} L(\Delta, p') d\Delta + \frac{\sigma}{4}.
\end{aligned}
$$

We can choose δ small enough, so that (i) $L(\Delta, p) - L(\Delta, p') < \frac{\sigma}{8\epsilon}$ on $(I \cap J) \setminus K$, (ii) the length of each of $I \setminus J$ and $J \setminus I$ is less than $\frac{\sigma}{4M}$, and (iii) $\sup_{\Delta \in J \setminus I} L(\Delta, p') < M$. In that case, the last expression is less than

$$\frac{\sigma}{8\epsilon} 2\epsilon + M \frac{\sigma}{4M} + M \frac{\sigma}{4M} + \frac{\sigma}{4} = \sigma.$$

Q.E.D.

Claim 5 $\Psi(\cdot, -p)$ *is continuous.*

Proof:

By Claim 4, the maximand in problem (2) is continuous in $(\bar{\Delta}, p)$. The claim follows from the Maximum Theorem applied to the special case when the optimisation problem has a unique solution and the constraints do not depend on the parameters. (Proposition 2 shows that problem (2) has a unique solution.) Q.E.D.

Claim 6 *Assume that (i)* $f : (0,\infty) \to \mathbb{R}$ *is absolutely continuous and weakly monotone on* $[a,b]$*, and (ii)* $\phi : \mathbb{R} \to \mathbb{R}$ *is absolutely continuous on* $f([a,b])$*.*[39] *Then,* $\phi \circ f$ *is absolutely continuous on* $[a,b]$*.*

Proof: Assume that f is weakly increasing on $[a,b]$.[40] Given $\epsilon > 0$, let $\delta > 0$ be such that, for any finite collection of pairwise disjoint intervals $(c_k, d_k) \subseteq f([a,b])$ satisfying $\sum_k (d_k - c_k) < \delta$, we have $\sum_k |\phi(d_k) - \phi(c_k)| < \epsilon$. Given this δ, let $\tau > 0$ be such that, for any finite collection of pairwise disjoint intervals $(a_k, b_k) \subseteq [a,b]$ satisfying $\sum_k (b_k - a_k) < \tau$, we have $\sum_k |f(b_k) - f(a_k)| < \delta$. We have $\sum_k |\phi(f(b_k)) - \phi(f(a_k))| = \sum_{\{k|f(a_k)\neq f(b_k)\}} |\phi(f(b_k)) - \phi(f(a_k))|$. Given that f is weakly increasing, any two intervals $(f(a_{k'}), f(b_{k'}))$ and $(f(a_{k''}), f(b_{k''}))$, such that $f(a_{k'}) \neq f(b_{k'})$ and $f(a_{k''}) \neq f(b_{k''})$, are disjoint. Thus, the latter sum is less than ϵ.[41] Q.E.D.

Claim 7 *Assume that* $l(\cdot)$ *is differentiable on* $(0,\infty)$*. Fix an arbitrary interval* $[p', p'']$*, where* $0 < p' < p''$*.* $\Psi(\cdot, -p)$ *is absolutely continuous on* $[p', p'']$*.*

Proof:

Given that Δ_L and Δ_1 are continuously differentiable in p (see Claim 3), they are absolutely continuous on $[p', p'']$. Given that they are also weakly monotone in p and the definite integral is absolutely continuous in each bound of integration, it follows from Claim 6 that each integral in expression (5) is absolutely continuous in p on $[p', p'']$. It is straightforward that each piece in expression (5) is absolutely continuous on $[p', p'']$.[42]

[39] $f([a,b])$ denotes the image of $[a,b]$ under f.

[40] The proof is analogous if f is weakly decreasing.

[41] This proof is based on the hint given at the end of problem 5.8.59 in Bogachev (2007).

[42] Given that $l(\cdot)$ is differentiable on $(0,\infty)$ and convex, it is continuously differentiable on $(0,\infty)$. Hence, the term $l(b)$ in the third piece in expression (5) is absolutely continuous in b.

The absolute continuity of $\Psi(\cdot, -p)$ on $[p', p'']$ follows from (i) the continuity of $\Psi(\cdot, -p)$ (see Claim 5), (ii) the absolute continuity of each piece in expression (5) on $[p', p'']$, and (iii) the fact that there are finitely many points in $[p', p'']$ at which there is a switch between the cases in expression (5) (see Claim 2). Q.E.D.

Claim 8 *Assume that $l(\cdot)$ is differentiable. Then, the following hold.*

1) For almost all b, $\Psi_b(b, c, \epsilon)$ exists and

$$\Psi_b(b, c, \epsilon) = \begin{cases} 0 & \textit{if } b \in b(-b, 1) \\ \frac{1}{2\epsilon}(l(\Delta_L + 2\epsilon) - l(\Delta_L)) & \textit{if } b \in b(-b, 2) \\ \frac{1}{2\epsilon}\left((\Delta_L - \Delta_1 + 2\epsilon)l'(b) + (c + l(b) - l(\Delta_L))\right) & \textit{if } b \in b(-b, 3) \end{cases} . \tag{6}$$

2) For almost all c, $\Psi_c(b, c, \epsilon)$ exists and

$$\Psi_c(b, c, \epsilon) = \begin{cases} 0 & \textit{if } c \in c(-c, 1) \\ \frac{1}{2\epsilon}(l(\Delta_L + 2\epsilon) - l(\Delta_L))\frac{\partial \Delta_L}{\partial c} & \textit{if } c \in c(-c, 2) \\ \frac{1}{2\epsilon}\left((\Delta_L - \Delta_1 + 2\epsilon) + (c + l(b) - l(\Delta_L))\frac{\partial \Delta_L}{\partial c}\right) & \textit{if } c \in c(-c, 3) \end{cases} . \tag{7}$$

3) For almost all ϵ, $\Psi_\epsilon(b, c, \epsilon)$ exists and

$$\Psi_\epsilon(b, c, \epsilon) = \frac{1}{\epsilon}\left(l(\bar{\Delta}^*(b, c, \epsilon)) - \Psi(b, c, \epsilon)\right). \tag{8}$$

Proof:

The claim follows by differentiating each piece in expression (5) with respect to p. The differentiation of any integrals is done via the Leibniz integral rule. Given that Δ_L and Δ_1 are continuously differentiable in each parameter (see Claim 3), we can use this rule when the limits of integration involve Δ_L and Δ_1.[43] Q.E.D.

Proof of Proposition 3:

Suppose that p' and p'' are such that $0 < p' < p''$. By the absolute continuity of $\Psi(\cdot, -p)$ (see Claim 7), we can write:

$$\Psi(p'', -p) - \Psi(p', -p) = \int_{p'}^{p''} \Psi_p(\tilde{p}, -p) d\tilde{p}.$$

First, suppose that $p = b$. It is easy to see that each piece in expression (6) is nonnegative. Thus, for any nondegenerate $[b', b''] \subseteq (0, \infty)$, $\Psi(b'', c, \epsilon) - \Psi(b', c, \epsilon) = \int_{b'}^{b''} \Psi_b(\tilde{b}, c, \epsilon) d\tilde{b} \geq 0$.

Next, suppose that $p = c$. Whenever we wish to make the dependence of Δ_L and Δ_1 on c explicit, we will write $\Delta_1(c)$ and $\Delta_L(c)$. Let c_1 and c_2 be as in statement 2) in Claim 2.

For all $c \in [c_2, \infty)$, the first piece of expression (7) applies. Thus, for any nondegenerate $[c', c''] \subseteq [c_2, \infty)$, $\Psi(b, c'', \epsilon) - \Psi(b, c', \epsilon) = \int_{c'}^{c''} \Psi_c(b, \tilde{c}, \epsilon) d\tilde{c} = 0$.

For all $c \in [c_1, c_2)$, the second piece of expression (7) applies. This second piece is negative given that (i) $l(\Delta_L + 2\epsilon) - l(\Delta_L) > 0$ whenever $-\Delta_L < \epsilon$ and (ii) $\frac{\partial \Delta_L}{\partial c} < 0$ (see Claim 3). Thus, for any nondegenerate $[c', c''] \subseteq [c_1, c_2)$, $\Psi(b, c'', \epsilon) - \Psi(b, c', \epsilon) = \int_{c'}^{c''} \Psi_c(b, \tilde{c}, \epsilon) d\tilde{c} < 0$.

We have shown that $\Psi(b, \cdot, \epsilon)$ is weakly decreasing on $[c_1, \infty)$. It follows that $\Psi(b, \cdot, \epsilon)$ is weakly decreasing or nonmonotone on $(0, \infty)$, which is the statement in Proposition 3. However, here we go further by establishing that, depending on the values of b and ϵ, both possibilities can in fact occur if $l(z) = z^2$.[44]

[43] At points p_1 and p_2 in each statement in Claim 2, different pieces of expression (5) apply to the left and right, so that $\Psi_p(p, -p)$ may not exist at $p = p_1$ or $p = p_2$. This does not invalidate Claim 8 given the "for almost all" qualification.

[44] I suspect this is true for any $l(\cdot)$ satisfying the assumptions behind Proposition 3, but I have not proved this.

Assuming that $l(z) = z^2$, we consider the behaviour of $\Psi(b, \cdot, \epsilon)$ on $(0, c_1)$. For all $c \in (0, c_1)$, the third piece of expression (7) applies. With a quadratic loss from deviations, this third piece becomes:

$$\frac{1}{2\epsilon}(\Delta_L(c) - \Delta_1(c) + 2\epsilon - b).$$

If $b \geq 2\epsilon$, we have $\Delta_L(c) - \Delta_1(c) + 2\epsilon - b < 0$. Thus, for any nondegenerate $[c', c''] \subseteq (0, c_1)$, $\Psi(b, c'', \epsilon) - \Psi(b, c', \epsilon) = \int_{c'}^{c''} \Psi_c(b, \tilde{c}, \epsilon) d\tilde{c} < 0$.

If $b < 2\epsilon$, we have $\lim_{c\downarrow 0}(\Delta_L(c) - \Delta_1(c) + 2\epsilon - b) = 2\epsilon - b > 0$. The equality follows because $\lim_{c\downarrow 0}(\Delta_L(c) - \Delta_1(c)) = 0$ (see Figure 3). Thus, there exists c_0, such that $0 < c_0 < c_1$ and $\Delta_L(c) - \Delta_1(c) + 2\epsilon - b > 0$ for all $c < c_0$. Thus, for any nondegenerate $[c', c''] \subseteq (0, c_0)$, $\Psi(b, c'', \epsilon) - \Psi(b, c', \epsilon) = \int_{c'}^{c''} \Psi_c(b, \tilde{c}, \epsilon) d\tilde{c} > 0$.

Finally, let us turn to the case $p = \epsilon$. Observe that (i) $\Psi(b, c, \epsilon)$ equals the average height of $L(\cdot)$ over the interval $[\bar{\Delta}^*(b, c, \epsilon) - 2\epsilon, \bar{\Delta}^*(b, c, \epsilon)]$ and (ii) $l(\bar{\Delta}^*(b, c, \epsilon))$ is strictly larger than this average height (see Figures 4-6). Hence, expression (8) is strictly positive. Thus, for any nondegenerate $[\epsilon', \epsilon''] \subseteq (0, \infty)$, $\Psi(b, c, \epsilon'') - \Psi(b, c, \epsilon') = \int_{\epsilon'}^{\epsilon''} \Psi_\epsilon(b, c, \tilde{\epsilon}) d\tilde{\epsilon} > 0$. Q.E.D.

11.7 Proof of Proposition 4

Let us write $\Delta_L(b, c)$ and $\Delta_1(b, c)$ to make explicit the dependence of Δ_L and Δ_1 on b and c.

Given Lemma 2 and the fact that $\Delta_L(b, c)$ is strictly increasing in b and strictly decreasing in c (see Figure 1), the statements in the proposition follow if we replace "contracts" and "expands" with "weakly contracts" and "weakly expands", respectively. It remains to show that the statements continue to hold without the "weakly" qualifier.[45]

[45] The terms "weakly contracts" and "weakly expands" are used in the sense of weak set inclusion.

First, consider statement 1) and suppose that $b' > b$. If c is such that $\Delta_L(b,c) < 0$[46] and ϵ is such that $\Delta_L(b,c) < -\epsilon < \Delta_L(b',c)$, the C default is optimal given (b,c,ϵ), but not given (b',c,ϵ). If c is arbitrary and ϵ is such that $\Delta_1(b',c) - \Delta_L(b',c) < 2\epsilon < \Delta_1(b,c) - \Delta_L(b,c)$, the $\overline{x}$-EL default is optimal given (b',c,ϵ), but not given (b,c,ϵ). This establishes statement 1) without the "weakly" qualifier. The proof of statement 2) without the "weakly" qualifier is analogous.

Next, consider statement 3) and suppose that $\epsilon' > \epsilon$. Let b be arbitrary. If c is such that $-\epsilon' < \Delta_L(b,c) < -\epsilon$,[47] the C default is optimal given (b,c,ϵ), but not given (b,c,ϵ'). If c is such that $2\epsilon < \Delta_1(b,c) - \Delta_L(b,c) < 2\epsilon'$,[48] the $\overline{x}$-EL default is optimal given (b,c,ϵ'), but not given (b,c,ϵ). This establishes statement 3) without the "weakly" qualifier. Q.E.D.

11.8 Proof of Lemma 4

Let x_k denote the value of x on which G_k is degenerate. Let $G_{k,\bar{\Delta}}$ denote the CDF over Δ's induced by G_k and $\bar{\Delta}$. Note that $G_{k,\bar{\Delta}}$ is degenerate on $\Delta = \bar{\Delta} - 2\epsilon + x_k - \underline{x}$. Let $G_{\bar{\Delta}}$ denote the CDF over Δ's induced by G and $\bar{\Delta}$. I.e., $G_{\bar{\Delta}}$ is defined by $G_{\bar{\Delta}}(\Delta) = G(\underline{x} + \Delta - \bar{\Delta} + 2\epsilon)$. We have:

[46] We can always find a value of c that is large enough, so that $\Delta_L(b,c) < 0$ (see Figure 1).

[47] Given that $\Delta_L(b,\cdot)$ is continuous and its range is $(-\infty, b)$ (see Figure 1), we can find for any b, ϵ, and $\epsilon' > \epsilon$ a value of c, such that $-\epsilon' < \Delta_L(b,c) \leq -\epsilon$.

[48] Given that $\Delta_1(b,\cdot) - \Delta_L(b,\cdot)$ is continuous and its range is $(0,\infty)$ (see Figure 3), we can find for any b, ϵ, and $\epsilon' > \epsilon$ a value of c, such that $2\epsilon \leq \Delta_1(b,c) - \Delta_L(b,c) < 2\epsilon'$.

$$
\begin{aligned}
&\int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} L(\Delta)dF_{\bar{\Delta}} = \\
&\int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} L(\Delta)d\left(\sum_{k=1}^{K}\alpha_k G_{k,\bar{\Delta}} + (1-\sum_{k=1}^{K}\alpha_k)G_{\bar{\Delta}}\right) = \\
&\sum_{k=1}^{K}\alpha_k \int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} L(\Delta)dG_{k,\bar{\Delta}} + (1-\sum_{k=1}^{K}\alpha_k)\int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} L(\Delta)dG_{\bar{\Delta}} = \\
&\sum_{k=1}^{K}\alpha_k L(\bar{\Delta}-2\epsilon+x_k-\underline{x}) + (1-\sum_{k=1}^{K}\alpha_k)\int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} L(\Delta)g(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta
\end{aligned}
$$

Given the requirement that G_k cannot be degenerate on $\underline{x}$ or on $\underline{x}+\Delta_R-\Delta_L$, it follows that $G_{k,\bar{\Delta}}$ cannot be degenerate on $\bar{\Delta}-2\epsilon$ or on $\bar{\Delta}-2\epsilon+\Delta_R-\Delta_L$. Hence, at $\bar{\Delta}=\Delta_L+2\epsilon$, $G_{k,\bar{\Delta}}$ cannot be degenerate on Δ_L or on Δ_R, so that $\bar{\Delta}-2\epsilon+x_k-\underline{x}\neq\Delta_L$ and $\bar{\Delta}-2\epsilon+x_k-\underline{x}\neq\Delta_R$. This is true for any k. Thus, given the differentiability of $l(\cdot)$, the first term in the last line above is differentiable at $\bar{\Delta}=\Delta_L+2\epsilon$. Thus, it suffices to restrict attention to the last integral, which we denote by $z(\bar{\Delta})$.

We need to consider two cases: $\Delta_L+2\epsilon>\Delta_R$ and $\Delta_L+2\epsilon<\Delta_R$.

<u>The Case when $\Delta_L+2\epsilon>\Delta_R$</u>

Let us write out $z(\bar{\Delta})$ for values of $\bar{\Delta}$, such that $\Delta_R<\bar{\Delta}\leq\Delta_L+2\epsilon$.

$$
\begin{aligned}
&z(\bar{\Delta}) = \int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} L(\Delta)g(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta = \\
&(c+l(b))\left(\int_{\bar{\Delta}-2\epsilon}^{\Delta_L} g(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta + \int_{\Delta_R}^{\bar{\Delta}} g(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta\right) + \\
&\qquad\qquad\int_{\Delta_L}^{\Delta_R} l(\Delta)g(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta
\end{aligned}
$$

Each integrand in the last expression is such that it and its partial derivative with respect to $\bar{\Delta}$ are continuous in Δ and $\bar{\Delta}$, so that this expression is differentiable and its derivative is given via the Leibniz integral rule:

$$(c+l(b))\left(-g(\underline{x})-\int_{\bar{\Delta}-2\epsilon}^{\Delta_L} g'(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta+g(\underline{x}+2\epsilon)-\int_{\Delta_R}^{\bar{\Delta}} g'(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta\right)-\int_{\Delta_L}^{\Delta_R} l(\Delta)g'(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta$$

The left derivative of $z(\cdot)$ at $\bar{\Delta}=\Delta_L+2\epsilon$ equals the last expression evaluated at $\bar{\Delta}=\Delta_L+2\epsilon$:

$$(c+l(b))\left(-g(\underline{x})+g(\underline{x}+2\epsilon)-\int_{\Delta_R}^{\Delta_L+2\epsilon} g'(\underline{x}+\Delta-\Delta_L)d\Delta\right)-\int_{\Delta_L}^{\Delta_R} l(\Delta)g'(\underline{x}+\Delta-\Delta_L)d\Delta \tag{9}$$

Next, let us write out $z(\bar{\Delta})$ for values of $\bar{\Delta}$, such that $\Delta_L+2\epsilon\leq\bar{\Delta}<\Delta_R+2\epsilon$.

$$z(\bar{\Delta})=\int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} L(\Delta)g(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta=$$
$$\int_{\bar{\Delta}-2\epsilon}^{\Delta_R} l(\Delta)g(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta+(c+l(b))\int_{\Delta_R}^{\bar{\Delta}} g(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta$$

Each integrand in the last expression is such that it and its partial derivative with respect to $\bar{\Delta}$ are continuous in Δ and $\bar{\Delta}$, so that this expression is differentiable and its derivative is given via the Leibniz integral rule:

$$(c+l(b))\left(g(\underline{x}+2\epsilon)-\int_{\Delta_R}^{\bar{\Delta}} g'(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta\right)-l(\bar{\Delta}-2\epsilon)g(\underline{x})-$$
$$\int_{\bar{\Delta}-2\epsilon}^{\Delta_R} l(\Delta)g'(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta$$

The right derivative of $z(\cdot)$ at $\bar{\Delta}=\Delta_L+2\epsilon$ equals the last expression evaluated at $\bar{\Delta}=\Delta_L+2\epsilon$:

$$(c+l(b))\left(g(\underline{x}+2\epsilon)-\int_{\Delta_R}^{\Delta_L+2\epsilon} g'(\underline{x}+\Delta-\Delta_L)d\Delta\right)-l(\Delta_L)g(\underline{x})-\int_{\Delta_L}^{\Delta_R} l(\Delta)g'(\underline{x}+\Delta-\Delta_L)d\Delta \tag{10}$$

Subtracting expression (10) from expression (9) yields $-(c+l(b)-l(\Delta_L))g(\underline{x})=-(c+l(b)-L(\Delta_L))g(\underline{x})$, which by Lemma 1 and the assumption $g(\underline{x})>0$ is strictly less than zero.

The Case when $\Delta_L+2\epsilon<\Delta_R$

Let us write out $z(\bar{\Delta})$ for values of $\bar{\Delta}$, such that $\Delta_L<\bar{\Delta}\leq\Delta_L+2\epsilon$.

$$z(\bar{\Delta})=\int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} L(\Delta)g(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta=$$
$$(c+l(b))\int_{\bar{\Delta}-2\epsilon}^{\Delta_L} g(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta+\int_{\Delta_L}^{\bar{\Delta}} l(\Delta)g(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta$$

Each integrand in the last expression is such that it and its partial derivative with respect to $\bar{\Delta}$ are continuous in Δ and $\bar{\Delta}$, so that this expression is differentiable and its derivative is given via the Leibnitz integral rule:

$$(c+l(b))\left(-g(\underline{x})-\int_{\bar{\Delta}-2\epsilon}^{\Delta_L} g'(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta\right)+$$
$$l(\bar{\Delta})g'(\underline{x}+2\epsilon)-\int_{\Delta_L}^{\bar{\Delta}} l(\Delta)g'(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta$$

The left derivative of $z(\cdot)$ at $\bar{\Delta}=\Delta_L+2\epsilon$ equals the last expression evaluated at $\bar{\Delta}=\Delta_L+2\epsilon$:

$$-(c+l(b))g(\underline{x})+l(\Delta_L+2\epsilon)g(\underline{x}+2\epsilon)-\int_{\Delta_L}^{\Delta_L+2\epsilon} l(\Delta)g'(\underline{x}+\Delta-\Delta_L)d\Delta \tag{11}$$

Next, let us write out $z(\bar{\Delta})$ for values of $\bar{\Delta}$, such that $\Delta_L+2\epsilon\leq\bar{\Delta}<\Delta_R$.

$$z(\bar{\Delta})=\int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} L(\Delta)g(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta=$$
$$\int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} l(\Delta)g(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta$$

Each integrand in the last expression is such that it and its partial derivative with respect to $\bar{\Delta}$ are continuous in Δ and $\bar{\Delta}$, so that this expression is differentiable and its derivative is given via the Leibnitz integral rule:

$$l(\bar{\Delta})g(\underline{x}+2\epsilon)-l(\bar{\Delta}-2\epsilon)g(\underline{x})-\int_{\bar{\Delta}-2\epsilon}^{\bar{\Delta}} l(\Delta)g'(\underline{x}+\Delta-\bar{\Delta}+2\epsilon)d\Delta$$

The right derivative of $z(\cdot)$ at $\bar{\Delta}=\Delta_L+2\epsilon$ equals the last expression evaluated at $\bar{\Delta}=\Delta_L+2\epsilon$:

$$l(\Delta_L + 2\epsilon)g(\underline{x} + 2\epsilon) - l(\Delta_L)g(\underline{x}) - \int_{\Delta_L}^{\Delta_L+2\epsilon} l(\Delta)g'(\underline{x} + \Delta - \Delta_L)d\Delta \qquad (12)$$

Subtracting expression (12) from expression (11) yields $-(c+l(b)-l(\Delta_L))g(\underline{x}) = -(c+l(b)-L(\Delta_L))g(\underline{x})$, which by Lemma 1 and the assumption $g(\underline{x}) > 0$ is strictly less than zero. Q.E.D.