NBER WORKING PAPER SERIES

OPTIMAL DEFAULT OPTIONS: THE CASE FOR OPT-OUT MINIMIZATION

B. Douglas Bernheim
Jonas Mueller Gastell

Working Paper 28254
http://www.nber.org/papers/w28254

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2020

We gratefully acknowledge general research support from Stanford University. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Optimal Default Options: The Case for Opt-Out Minimization
B. Douglas Bernheim and Jonas Mueller Gastell
NBER Working Paper No. 28254
December 2020
JEL No. D10,D11,D14

ABSTRACT

We examine the desirability of opt-out minimization, a well-known and simple rule of thumb for setting default options such as passively selected contribution rates in employee-directed pension plans. Existing results suggest that this strategy is welfare-optimal only under highly restrictive assumptions. In this paper, we dispense with those assumptions and demonstrate far more generally that opt-out minimization is approximately optimal. Our main results require only a small number of weak regularity conditions. We also conduct simulations to evaluate the accuracy of the approximation, as well as the robustness of our conclusions with respect to additional dimensions of heterogeneity. We conclude that opt-out minimization is not only practical, but also has a solid and general normative foundation.

B. Douglas Bernheim
Department of Economics
Stanford University
Stanford, CA 94305-6072
and NBER
bernheim@stanford.edu

Jonas Mueller Gastell
jonasmg@stanford.edu

Optimal Default Options: The Case for Opt-Out Minimization

B. Douglas Bernheim and Jonas Mueller-Gastell*

December 13, 2020

Abstract

We examine the desirability of opt-out minimization, a well-known and simple rule of thumb for setting default options such as passively selected contribution rates in employee-directed pension plans. Existing results suggest that this strategy is welfare-optimal only under highly restrictive assumptions. In this paper, we dispense with those assumptions and demonstrate far more generally that opt-out minimization is *approximately* optimal. Our main results require only a small number of weak regularity conditions. We also conduct simulations to evaluate the accuracy of the approximation, as well as the robustness of our conclusions with respect to additional dimensions of heterogeneity. We conclude that opt-out minimization is not only practical, but also has a solid and general normative foundation.

1 Introduction

In standard consumer theory, a decision problem consists of a menu of alternatives. The consumer's choice depends on the contents of the menu and nothing else. In practice, there is always some item on the menu that serves as a default option, in the following sense: if the consumer fails to make a choice, whether intentionally or by neglect, the default option will prevail. Standard theory ignores default options because it presumes they have no bearing on the consumer's opportunities or preferences. But if the implementation of a choice requires the expenditure of effort, the identity of the default option materially impacts the contents of the opportunity set. In addition, default options may create psychological framing effects that trigger behavioral responses.

The ubiquity of default options gives rise to important normative questions about the optimal design of "choice architectures." Indeed, Thaler and Sunstein (2008) point to the manipulation of default options as a core strategy for

*Bernheim: Department of Economics Stanford University Stanford, CA 94305-6072 and NBER, bernheim@stanford.edu. Mueller-Gastell: Department of Economics Stanford University Stanford, CA 94305-6072, jonasmg@stanford.edu

"nudging" people towards better choices.[1] The literature has addressed these questions primarily in the context of setting default contribution rates for 401(k) plans, where a collection of empirical studies have revealed that changing the default option has a powerful effect on employees' contributions (see Madrian and Shea (2001), or Beshears et al. (2018) for a summary of the subsequent literature). It is worth emphasizing, however, that the same conceptual considerations arise in other contexts, including widely studied topics such as asset allocation in investment portfolios (Agnew and Szykman 2005) and employee health insurance plan choice (Handel and Kolstad 2015).[2]

Within the economic literature, discussions of optimal default options begin with Thaler and Sunstein (2003), who propose a simple rule of thumb: minimize the fraction of consumers who opt out of the default. They do not attempt to justify the criterion formally, and the ensuing literature establishes that opt-out minimization is welfare-optimal only under special conditions. Carroll, Choi, Laibson, Madrian, and Metrick (2009) consider a model in which present focus, which they interpret as present bias, causes workers to place excessive weight on opt-out costs. Under restrictive assumptions about the distribution of ideal points (uniform in a given interval) and the utility function (losses are a fixed quadratic function of the distance from the ideal point), they show that the optimal default is either opt-out minimizing (for small bias and low dispersion of preferences), an "offset-default" that forces some of the distribution into active choice (for small bias but wide preference dispersion), or an extreme choice that forces active decision making on all agents (for large bias). Goldin and Reck (2019) consider a related model that admits a more general interpretation of the bias parameter and its normative significance. Under similarly restrictive assumptions about the distribution of ideal points (symmetry and single-peakedness) and the utility function (loss is a fixed symmetric and concave function of the distance from the ideal point), they prove a similar result: there exist both parameterizations under which forcing active choice is optimal and ones under which opt-out minimization is optimal. In both cases, the opt-out minimizing default option is also the mean, median, and mode of the ideal-point distribution. Thus, neither study reveals whether opt-out minimization is desirable per se in these settings, or merely because it coincides with these other distributional features.

A somewhat different message emerges from Bernheim, Fradkin, and Popov (2015). Instead of proving formal characterization results under specialized assumptions, they examine the welfare effects of default options in empirically

[1] Bernheim and Taubinsky (2018) question the classification of variations in defaults as nudges. In their taxonomy, a nudge is a change in the decision frame that does not change opportunities. As noted above, if opting out is costly, changing the default does alter the opportunity set, and hence it is not a nudge.

[2] While the existing literature has examined these issues primarily in contexts involving government-regulated employee benefits, it is worth emphasizing that they arise in many other contexts. A few examples illustrate the diversity of potential applications: the default of equal division governs the allocation of assets for those who die intestate, "boilerplate" legal contracts serve as defaults for many business transactions, and airlines sometimes make default seat assignments when processing new reservations.

parametrized models. In many instances, they find that the welfare-maximizing and opt-out-minimizing default rates coincide. In other instances, the two diverge meaningfully. On the surface, these findings appear to corroborate the impression one draws from Carroll et al. and Goldin and Reck, that the optimality of opt-out minimization is a special and fragile result. However, Bernheim, Fradkin, and Popov also find that "the Thaler-Sunstein opt-out-minimization criterion yields small welfare losses even when it is suboptimal; hence it is a reasonable rule of thumb" (Bernheim, Fradkin, and Popov 2015, p. 2800). In addition, Bernheim, Fradkin, and Popov argue that the optimality of extreme unattractive defaults in settings with large biases may be artifactual, because it ignores the possibility of using complementary policy instruments. Their simulations encompass the possibility that the employer can also impose a dissipative penalty for passive choice, such as "red tape" requirements. In their simulations, the employer never uses the default to incentivize active choice when such penalities are available.

These studies leave two critical questions unanswered. First, is the *approximate* optimality of the opt-out minimizing default noted in Bernheim, Fradkin, and Popov a general property, or does it too depend on highly specialized assumptions? Second, how is the approximate optimality of opt-out minimization affected by the availability of penalties for passive default? With respect to the second question, we depart from Bernheim, Fradkin, and Popov by considering the natural possibility that the employer can impose non-dissipative penalties for passive choice – in other words, the employer can collect fees from those who fail to choose actively, and distribute the proceeds equally among all workers in the form of higher wages, thereby leaving profits unchanged.[3]

Our analysis yields a surprisingly general case for the opt-out-minimization criterion. We consider a model closely related to those studied in Carroll et al. (2009), Bernheim, Fradkin, and Popov (2015), and Goldin and Reck (2019), but we dispense with restrictive assumptions involving symmetry, single-peakedness, and the like. Instead, we impose only a limited set of technical regularity conditions. We start by characterizing the limit of opt-out-minimizing default options as opt-out costs shrink to zero.[4] Then we characterize the limit of welfare-maximizing default options, and show that the two are the same. For settings in which the employer believes biases infect the worker's opt-out choices, we characterize the optimal fine, and then demonstrate that, subject to the imposition of the fine, the same limiting result obtains. We then use numerical simulations to address two limitations of our analysis. First, we show that our characterization of optimal policy for the limiting case provides a decent approximation for settings with meaningful social stakes. Second, we examine the robustness of our conclusions with respect to the introduction of

[3] As we explain in Section 3, dissipative and non-dissipative penalties are feasible in settings where opting out involves implementation costs, but not in settings where it involves deliberation costs. It is therefore important to emphasize that the pertinent literature studies the first type of settings, not the second.

[4] Choukhmane (2019)argues that prior literature on the size of opt-out costs has systematically overestimated these costs and finds "as-if" opt-out costs of around $250.

additional and potentially important dimensions of worker heterogeneity that are not in our basic model. The main lesson from the simulations is that the limiting result generally serves as a good guide, and that opt-out-minimization is approximately optimal.

Opt-out minimization has the advantage of being significantly easier to achieve in practice than explicit welfare maximization. Employers can determine the former through "model free" experimentation or by using relatively simple surveys, while the latter requires analytic sophistication. The approximate coincidence of opt-out-minimizing and welfare-maximizing defaults therefore enhances the feasibility of optimizing policy.

The remainder of the paper proceeds as follows. Section 2 details the model. Section 3 develops our formal results, and Section 4 describes our simulations. We close in Section 5 with some brief thoughts about directions for subsequent research.

2 The model

For concreteness and to promote interpretability, we depict the problem of interest as one of selecting a default contribution rate for workers participating in an employer-base retirement savings plan. However, the model is sufficiently general to apply in a wide range of contexts involving default options.

2.1 Workers

We use x to stand for the contribution rate of a worker ("he") newly eligible to participate in a plan sponsored by his employer ("she"). The worker chooses x from a compact interval X. The plan's provisions specify a default contribution rate of D. We focus on the worker's initial choice between accepting the default and opting out to some $x \neq D$.

We assume the worker's utility is additively separable in the contribution rate (x), income (m), and the level of effort exerted to effectuate opt-out (c). For the sake of analytic tractability, utility is linear in income and additively separable in effort. Thus:

$$u(x, x^*, m, c) = \beta \left[V(x, x^*) + m \right] - \Gamma(c). \tag{1}$$

Several remarks concerning equation (1) are in order.

First, notice that the function V depends not only on x, but also on a parameter x^*, which we interpret as the contribution rate the worker regards as ideal, in the sense that $x = x^*$ uniquely maximizes $V(x, x^*)$. The ideal contribution rate varies across the population, and its distribution is given by F, a CDF, with density f.

Second, our model presupposes that opt-out is costly because the worker must expend effort to *implement* any selection other than D. For example, he must inform himself about selection procedures, fill out forms, visit his employer's personnel office, and the like. Consistent with other theoretical work

on this topic (Bernheim, Fradkin, and Popov (2015), Carroll et al. (2009), and Goldin and Reck (2019)), we abstract from the interesting possibility that the worker must expend cognitive effort to understand his own preferences (the function $V(\cdot, x^*)$). Accordingly, the effort level, c, is a binary variable, where $c = 0$ indicates that the worker has accepted the default, and $c = 1$ indicates that he has taken the necessary steps to opt out. We adopt the normalization that $\Gamma(0) = 0$, and we define $\gamma = \Gamma(1)$, which represents the utility penalty associated with the effort of opting out.

Third, we apply a weighting factor, β, to the utility derived from retirement contributions and money. We use this parameter to introduce inclinations that the employer views as biases. We elaborate on the interpretation of this parameter below when discussing the employer's objectives.

In addition to specifying a default contribution rate D, the plan may also provide workers with a lump-sum bonus, B, and specify a fixed fine, K, that falls on those who make passive choices (i.e., accept the default). The purpose of the fine will be to incentivize active choice; the purpose of the bonus will be to maintain budget balance for the employer. To be clear, in a setting where workers must expend effort to understand their own preferences, an incentive of this type might simply induce them to go through the motions of opting out, for example by selecting an option that differs only slightly from D without giving serious consideration to his choice. It is therefore worth emphasizing that our results on optimal fines, like other results in this literature, are applicable only in settings where implementation rather than deliberation is costly.

For simplicity, the employer levies fines and disburses bonuses at the same point in time. Each worker is infinitessimal, and therefore ignores any effect of his own choice on the magnitude of the bonus through the budget balance condition. These transfers flow to and from the worker's income. Because utility is linear in income, the level of the worker's baseline income (before fines and bonuses) is immaterial, so we take it to be zero.

The worker chooses x to maximize $u(x, x^*, B - (1 - c)K, c)$, subject to the constraint that $c = 0$ if $x = D$ and $c = 1$ otherwise. When the worker opts out ($x \neq D$), it is obviously in his interest to select $x = x^*$. Accordingly, we can also treat him as choosing $c \in \{0, 1\}$, where these options lead to the following payoffs:

$$\beta\left[(1-c)V(D, x^*) + cV(x^*, x^*)\right] - c\gamma - (1-c)K + B$$

The worker therefore opts out of the default whenever

$$\beta \underbrace{\left(V(x^*, x^*) - V(D, x^*)\right)}_{:=\Delta(D, x^*)} \geq \gamma - K. \tag{2}$$

Thus, the mass of agents who opt-out is given by $\Pr\left[\Delta(D, x^*) \geq \frac{\gamma - K}{\beta}\right]$. We define the optimal opt-out function as follows: $C(D, x^*) = 1$ when equation (2) is satisfied, and $C(D, x^*) = 0$ otherwise. The worker's optimized utility is then

$$U(D,x^*) = \beta\left[(1-C(D,x^*))V(D,x^*) + C(D,x^*)V(x^*,x^*)\right] \\ - C(D,x^*)\gamma - (1-C(D,x^*))K + B,$$

To prove our analytical results, we require V and F to satisfy a handful of regularity properties. For V, we invoke the following assumption:

Assumption 1 *The following properties hold for V: (i) (Differentiability) V has well-defined and continuous first through third derivatives; (ii) (Concavity) $V(y,x)$ is strictly concave in y, and there exists $v^{min} > 0$ such that $V_{11}(y,x) < -v^{min}$ for all $y, x \in X$; (iii) (Single Crossing) $V_{12}(y,x) > 0$.*

We also assume that F, the CDF for x^*, possesses the following regularity properties:

Assumption 2 *The following properties hold for F: (i) (Full Support) there exists $f^{min} > 0$ such that for $f(x)$, the density function of F, $f(x) > f^{min}$ holds for all $x \in X$; (ii) (Differentiability) F has well-defined and continuous first and second derivatives.*

2.2 The employer

The employer (or planner) cannot distinguish among workers based on x^*, their ideals. Instead, she must select values of the default D, the bonus B, and the fine K, that apply uniformly to everyone. She makes this choice subject to budget balance:

$$B = K \Pr\left[\Delta(D,x^*) < \frac{\gamma - K}{\beta}\right]. \tag{3}$$

Because utility is quasi-linear in income, varying B leaves the right-hand side unchanged. Thus, we can think of the employer as choosing D and K, where the resulting value of B is given by equation (3).

The employer is a utilitarian: she seeks to maximize the aggregate value of workers' utilities. However, she may disagree with the workers concerning the assessment of their well-being. In particular, she evaluates each worker's utility based on the assumption that the normatively correct value of β is unity. Thus, to the extent that $\beta \neq 1$, she is of the opinion that decision bias infects opt-out decisions.

One potential interpretation of $\beta < 1$ is that the employer believes workers are subject to "present bias:" she thinks they place "too much" weight on effort costs, which are immediate, compared with retirement income, fines, and bonuses, which are all delayed.[5] Other interpretations are also possible; see Bernheim, Fradkin, and Popov (2015) for an extended discussion. A key feature of our framework is that the employer sees the bias as pertaining to the opt-out decision, rather than to the choice of x conditional on opting out. In other words,

[5] See Bernheim and Taubinsky (2018) for a critical discussion of this normative perspective.

she agrees that x^* is the worker's ideal choice. Whether this assumption is reasonable depends on the context. For retirement savings accounts, companies implement changes in contribution rates with a delay, so all consequences of contribution elections aside from effort are in the future. Thus, to the extent the employer believes workers are quasi-hyperbolic discounters and interprets β as "present bias," that bias would affect the opt-out decision, but not the chosen contribution rate, precisely as we assume.

Under the preceding assumptions, the employer evaluates the worker's well-being according the following function:

$$\begin{aligned}\tilde{U}(D, x^*) = &[(1 - C(D, x^*))V(D, x^*) + C(D, x^*)V(x^*, x^*)] \\ &- C(D, x^*)\gamma - (1 - C(D, x^*))K + B.\end{aligned}$$

In other words, she recognizes that bias (potentially) governs workers' opt-out choices through $C(D, x^*)$, but she ignores the bias parameter β when evaluating welfare. Aggregate utility for all workers is then given by $E_{x^*}\left[\tilde{U}(D, x^*)\right]$. That expression serves as the employer's objective function.

3 Analytic characterization of optimal defaults

Our analysis proceeds in three steps, all of which focus on the limit as opt-out costs become small. First we characterize the opt-out minimizing default option. Second, under the assumptions that the employer shares workers' normative judgments ($\beta = 1$) and is unable to impose fines for passive choice ($K = 0$), we prove that the optimal policy entails approximate opt-out minimization. Third, allowing for the possibility that the employer believes bias infects the workers' opt-out decisions ($\beta \neq 1$), we prove that the optimal policy entials approximate opt-out minimization along with positive fines and bonuses, and that the optimal fine is in fact zero when there is no normative disagreement ($\beta = 1$). Because these results all pertain to a limiting case, we undertake numerical simulations in Section 4 to assess the generalizability of our conclusions to settings with substantial (unbiased) effort costs, and also to address complexities not included in our basic model.

In addition to stating and discussing our main results, we also illuminate their logic by providing partial proofs in the text. The proofs are partial in the sense that they rely on four lemmas for which we provide intuition in the text, but (due to their technical nature) prove in the appendix.

We begin with a lemma that simplifies our analysis by guaranteeing that the set of ideal points for which workers opt in (that is, accept the default) is an interval. This property is a simple consequence of Single Crossing (Assumption 1, part (iii)). For the purpose of this lemma, we define $\tau \equiv \frac{\gamma - K}{\beta}$ (recalling that the worker opts out iff $\Delta(D, x^*) \geq \tau$).

Lemma 1 *For any given D, there is a unique interval $[x_l(D, \tau), x_h(D, \tau)]$ containing D such that the worker weakly prefers the default to opt-out if and*

only if $x^* \in [x_l(D,\tau), x_h(D,\tau)]$. *This preference is strict on the interior of the interval, and the worker is indifferent at any boundary of the interval that is interior to* X.

3.1 Opt-out minimization

As a preliminary matter, we establish existence of an opt-out minimizing default option. To be clear, we interpret opt-out minimization as pertaining to settings with no bias or fines for passive choice ($\beta = 1$, $K = 0$). Under our assumptions, it is easy to check that the opt-in frequency, $\Pr[\Delta(D,x) < \gamma]$, varies continuously with D given any γ. Accordingly, there exists a (possibly non-unique) default option, $D_P(\gamma)$, that maximizes opt-in (and minimizes opt-out).

Our aim is to study the limiting behavior of the opt-out-minimizing default option. A natural strategy would be to examine the limit of the functions $\Pr[\Delta(D,x) < \gamma]$ as $\gamma \to 0$, and to characterize the maximum of the limiting function. That strategy is problematic because $\Pr[\Delta(D,x) < \gamma]$ coverges to zero for all D. Therefore, to study the limiting behavior of the opt-out-minimizing defaults, it is helpful to rescale the objective function. As γ shrinks, we need to progressively scale it up just enough so that the resulting function neither collapses to 0 nor explodes to infinity. As it turns out, we accomplish this objective through the following normalization:

$$Q(D,\gamma) \equiv \frac{\Pr[\Delta(D,x) < \gamma]}{2\gamma^{\frac{1}{2}}}$$

To characterize opt-out-minimizing defaults for small γ, we study how $Q(D,\gamma)$ behaves in the limit as $\gamma \to 0$. The mass of workers who fall within the opt-in interval depends on two features of the model: (i) the width of that interval, which reflects the curvature of V, and (ii) the densities of the points in the window. Taking a second-order approximation of $\triangle(D,x)$ around $x = D$ (and noting that the first-order term is identically zero), we have

$$\triangle(D,x) \approx -\frac{1}{2} V_{11}(D,D)\,(D-x)^2$$

Because workers at the boundaries of the opt-in interval are indifferent between opting in and opting out (as described in Lemma 1), we know that $-\frac{1}{2}V_{11}(D,D)\,(D-x_i)^2 \approx \gamma$ for $i = l, h$. We can use this relationship to approximate the width of each half-interval:

$$|D - x_i| \approx \left(\frac{\gamma}{-\frac{1}{2}V_{11}(D,D)}\right)^{\frac{1}{2}}$$

The full length of the opt-in interval is approximately twice the preceding term, while the density within the interval, for small γ, is roughly constant at $f(D)$. To approximate the opt-in frequency, we multiply these terms together. Dividing

once again by $2\gamma^{\frac{1}{2}}$ to match the scale of $Q(D,\gamma)$, we arrive at the following normalized approximation of the opt-in frequency:

$$\tilde{Q}(D) \equiv f(D)\left(\frac{1}{-\frac{1}{2}V_{11}(D,D)}\right)^{\frac{1}{2}}$$

The preceding intuitive derivation motivates the conjecture that $Q(D,\gamma)$ converges to $\tilde{Q}(D)$ for small γ. The following lemma proves this conjecture and establishes that convergence is in fact uniform in D:

Lemma 2 *$Q(D,\gamma)$ converges uniformly to $\tilde{Q}(D)$ as $\gamma \to 0$.*

Now we define the opt-out-minimizing default rate, D^*, according to the limiting opt-in probability function, $\tilde{Q}(D)$:

$$D^* \equiv \arg\max_{D\in X} \tilde{Q}(D)$$

Given the compactness of X along with our continuity assumptions, the maximum exists. Cases with multiple maxima are non-generic and therefore of little interest.[6] To avoid some technical complications, we will therefore rule those cases out by assumption.[7]

Assumption 3 *D^* is unique.*

It is worth emphasizing that D^* is not necessarily the point of maximal density. If the curvature of V is the same at all ideal points – in other words, if $V_{11}(x,x)$ does not vary with x – then plainly D^* maximizes f. However, we will not impose this curvature restriction.

In light of the fact that $Q(D,\gamma) \rightrightarrows \tilde{Q}(D)$ (Lemma 2) and the fact that $\tilde{Q}(D)$ is bounded on X,[8] it follows immediately that the maximizers of $Q(D,\gamma)$ converge to the maximizer of $\tilde{Q}(D)$. Thus we obtain our characterization of the limit of opt-out-minimizing default options:

Proposition 1 *The opt-out-minimizing default option $D_P(\gamma)$ converges to D^* as $\gamma \to 0$.*

[6] Starting from settings with multiple maxima, there are always small perturbations of f or V that yield uniqueness. Starting from settings with unique equilibria, sufficiently small perturbations of f and V preserve uniqueness. Precise formulations of these assertions require considerable technical detail and are largely orthogonal to our main line of analysis.

[7] Non-uniqueness of D^* would raise the possibility that opt-out-minimizing and welfare-maximizing defaults might converge to different maximizers of $\tilde{Q}(D)$, and hence not to each other. We conjecture that suitably defined sets of ε-optima would nevertheless coincide in the limit.

[8] This claim follows from the fact that f is bounded above and V_{11} is bounded away from zero.

3.2 Optimal default options with no bias and no fines

Next we characterize optimal defaults for small γ. For the purpose of this section, we assume the employer treats the worker's opt-out choices as unbiased ($\beta = 1$), and we exclude the use of fines for passive choice ($K = 0$). In the next section, we show that with $\beta = 1$, the employer would not use fines even if they were available.

The employer's problem, setting D to maximize $E_{x^*}[U(D, x^*)]$, is obviously equivalent to maximizing

$$L(D, \gamma) \equiv E_{x^*}\left[U(D, x^*) - V(x^*, x^*)\right],$$

which we interpret as the total welfare loss relative to the ideal outcome. For any given x^*, the term in brackets is either $-\gamma$ (if the worker incurs the opt-out cost and selects his optimal contribution rate) or $V(D, x^*) - V(x^*, x^*) = -\Delta(D, x^*)$ (if he accepts the default). It follows that we can rewrite $L(D, \gamma)$ as follows:

$$\begin{aligned} L(D, \gamma) = &- \left(1 - \Pr\left[\Delta(D, x) < \gamma\right]\right)\gamma \\ &- \Pr\left[\Delta(D, x) < \gamma\right] E_{x^*}\left[\Delta(D, x) \mid \Delta(D, x) < \gamma\right] \end{aligned} \tag{4}$$

Under our assumptions, it is easy to check that this objective function varies continuously with D. Accordingly, there exists a (possibly non-unique) default rate $D_L(\gamma)$ that minimizes the welfare loss on the compact set X.

Our aim is to study the limiting behavior of the welfare-maximizing default option. Once again, a natural strategy would be to examine the limit of the functions $L(D, \gamma)$ as $\gamma \to 0$, and to characterize the maximum of the limiting function. The problem, as with the case of $\Pr[\Delta(D, x) < \gamma]$, is that $L(D, \gamma)$ coverges to zero for all D. Therefore, to study the limiting behavior of the welfare-maximizing defaults, it is helpful to rescale the objective function. In this instance, to ensure it neither collapses to 0 nor explodes to infinity, we need to both translate and rescale it. In particular, we define:

$$W(D, \gamma) \equiv \frac{L(D, \gamma) + \gamma}{\gamma^{\frac{3}{2}}}$$

Obviously, for any given γ, the maximizers of L and W coincide.

To make progress with our characterization, we employ the following decomposition:

$$W(D, \gamma) \equiv Q(D, \gamma)\left[1 - Z(D, \gamma)\right]$$

where

$$Z(D, \gamma) \equiv \frac{E_{x^*}\left[\Delta(D, x) \mid \Delta(D, x) < \gamma\right]}{\gamma}$$

We have already studied the limiting behavior of $Q(D, \gamma)$, so here we examine the limiting behavior of $Z(D, \gamma)$. The following lemma provides the key step for proving our main result:

Lemma 3 *$Z(D,\gamma)$ converges uniformly to $\frac{1}{3}$ as $\gamma \to 0$.*

Because we have made no parametric assumptions, the uniform convergence of Z to a constant may seem surprising. To build intuition, consider the case where $\Delta(D,x)$ is quadratic, so that $\Delta(D,x) = -v_0(D-x)^2$ for some scalar v_0, and F is uniform, so that the density is some constant, f_0. For convenience, normalize the action so that $D = 0$. In that case,

$$\frac{E_{x^*}\left[\Delta(D,x) \mid \Delta(D,x) < \gamma\right]}{\gamma} = \frac{\int_{-\sqrt{\gamma/v_0}}^{\sqrt{\gamma/v_0}} v_0 x^2 f_0 dx}{\gamma \int_{-\sqrt{\gamma/v_0}}^{\sqrt{\gamma/v_0}} f_0 dx} = \frac{\frac{2}{3} v_0 f_0 \left(\frac{\gamma}{v_0}\right)^{\frac{3}{2}}}{2\gamma f_0 \left(\frac{\gamma}{v_0}\right)^{\frac{1}{2}}} = \frac{1}{3}$$

To visualize this property, picture a parabola that achieves a minimum at the origin and that passes through the points $\left(\sqrt{\gamma/v_0}, \gamma\right)$ and $\left(-\sqrt{\gamma/v_0}, \gamma\right)$. Then the area under the parabola on the interval $\left[-\sqrt{\gamma/v_0}, \sqrt{\gamma/v_0}\right]$ constitutes one-third of the area of the rectangle $\left[-\sqrt{\gamma/v_0}, \sqrt{\gamma/v_0}\right] \times [0, \gamma]$. The intuition for the lemma, which asserts that this property is general for small γ, is simply that utility is approximately quadratic and density is approximately constant within a small neighborhood of D.

At this point, we know that $Q(D,\gamma) \rightrightarrows \tilde{Q}(D)$ (Lemma 2) and $1 - Z(D,\gamma) \rightrightarrows \frac{2}{3}$ (Lemma 3) as $\gamma \to 0$. With $\tilde{Q}(D)$ and $Z(D,\gamma)$ bounded, we therefore have $W(D,\gamma) \equiv Q(D,\gamma)\left[1 - Z(D,\gamma)\right] \rightrightarrows \frac{2}{3}\tilde{Q}(D)$ on the compact set X. It follows that the maximizers of $W(D,\gamma)$, and hence of $L(D,\gamma)$, converge to the maximizer of $\tilde{Q}(D)$. Thus we obtain our characterization of the limit of welfare-maximizing default options:

Proposition 2 *Assuming $\beta = 1$ and restricting $K = B = 0$, the welfare-maximizing default option $D_L(\gamma)$ converges to D^* as $\gamma \to 0$.*

Combining Propositions 1 and 2, we reach our main conclusion: the difference between the opt-out-minimizing and welfare-maximizing default options vanishes as $\gamma \to 0$. As noted in Section 4.2, it is straightforward to extend this result to settings with heterogeneity in the opt-out cost γ, provided x^* and γ are uncorrelated.

3.3 Optimal default options with bias and fines

The final step in our theoretical characterization of optimal defaults is to extend the analysis to cases with arbitrary β while also allowing for fines and bonuses. With $\beta < 1$, the employer believes workers are excessively reluctant to make active choices. One way to incentivize active choice is to set an unattractive default. Carroll et al. (2009) show that this alternative is in fact optimal when the decision bias is sufficiently severe, but their analysis does not contemplate a role for fines. In our setting, the employer can in principle incentivize active

choice through various combinations of unattractive defaults and fines. We prove that the problem is separable, in the sense that the optimal policy addresses bias exclusively through fines, and then sets the default to balance the costs to opt-outs and opt-ins exactly as in settings with no bias.

Formally, we decompose the general problem into two parts: first, we determine the optimal fine and bonus for arbitrary D; then we optimize over D. The following lemma characterize the optimal fine and, by implication, the optimal bonus for any fixed default option.

Lemma 4 *Fixing D, the optimal fine is $K^* = (1-\beta)\gamma$.*

The intuition for Lemma 4 is that, by establishing a fine equal to the portion of costs that the worker ignores ($(1-\beta)\,\gamma$), the employer corrects the "internality" that would otherwise give rise to a welfare loss. The literature on Behavioral Public Economics contains a collection of parallel results; see Bernheim and Taubinsky (2018).

Conditional on setting the optimal fine and bonus for each D, the objective function reduces to the same one analyzed in the proof of Proposition 2. Indeed, this property emerges directly from the arguments in the proof of Lemma 4. Consequently, solving for the optimal default with arbitrary β conditional on the optimal fine, call it $D_L^{\beta}(\gamma)$, is mathematically equivalent to solving for the optimal default with $\beta = 1$ and no fine. Combining Lemma 4 with Proposition 3, we therefore arrive at the following extension of our main analytic result:

Proposition 3 *For arbitrary β and optimal fines, the welfare-maximizing default option $D_L^{\beta}(\gamma)$ converges to D^* as $\gamma \to 0$.*

It is important to emphasize that, in this setting, we interpret D^* as the default rate that minimizes opt out conditional on setting the optimal fine, K^*, rather than the opt-out-minimizing default rate with $K = 0$. With that interpretation in mind, we reach the general conclusion that, when fines are feasible, the difference between the opt-out-minimizing and welfare-maximizing default options vanishes as $\gamma \to 0$. Notice that Prop 3 extends Prop 2 for the case of no bias, in that it shows the robustness of the latter result to the availability of fines, which are not used in the optimum.

4 Numerical simulations

In this section, we use numerical simulations to address two limitations of the preceding analysis. First, we have characterized optimal policy for small γ. As γ converges to zero, the stakes implicated by the policy question also become vanishingly small. It is therefore important to determine whether our characterization of optimal policy for the limiting case provides a decent approximation for settings with meaningful social stakes. Second, for the sake of tractability, we have assumed that heterogeneity among workers is limited to their ideal points. It is also likely that workers differ in terms of the cost of opt-out, γ,

and the bias parameter, β. Of greatest concern is the possibility that heterogeneity in β could overturn Lemma 4 in ways that undermine the implications of Proposition 3. The main lesson from the simulations is that the limiting result generally serves as a good approximation, and that opt-out-minimization is approximately optimal.

4.1 Parametrizations

To conduct numerical simulations, we must replace the general functions V and F with parametric specifications. For V, we consider two alternatives: a quadratic utility function, which exhibits the symmetry property imposed in the prior literature, and an asymmetric linear-exponential utility function.[9] For F, the CDF for the distribution of ideal points, we examine three alternatives: 1) a truncated Normal distribution, which exhibits the symmetry and single peakedness properties imposed in the prior literature, 2) a highly asymmetric distribution with a unique mode at a boundary value, and 3) an asymmetric bimodal distribution. Across all simulations, the support of the ideal-point distribution is the interval $[0, 5]$.

Some of our simulations add heterogeneity with respect to β and γ. Extending the model to settings with heterogeneous γ requires us to revisit the technical definition of vanishing opt-out costs. Specifically, we define a worker-specific opt-out proclivity parameter, η_i, which we assume is distributed according to some CDF, G, with bounded support. We then introduce a scaling parameter γ, such that the opt-out cost for worker i, call it γ_i, equals $\gamma\eta_i$. This formulation allows us to examine the case of vanishing heterogeneous opt-out costs by letting the common scaling parameter shrink to zero. In effect, this formulation preserves the pattern of heterogeneity as opt-out costs decline. For the sake of numerical tractability, we model heterogeneity in β and η by restricting attention to settings with three potential present-bias factors and three potential opt-out proclivities: $\beta_i \in \{0.5, 0.8, 1\}$ and $\eta_i \in \{0.5, 1, 2\}$.

In our basic simulations with heterogeneity in present bias and/or opt-out costs, we assume that β_i, η_i, and x_i^* are distributed independently, and that the marginal distributions of β_i and η_i are uniform. However, in some simulations, we allow for correlations among these parameters. Because the possibilities are virtually limitless, we employ a simple correlational structure that allows us to explore the impact of directional relationships between the variables. Interpreting a higher ideal point x^* as a higher savings rate, it is natural to assume that workers with smaller ideal points, on average, are more present biased and have higher opt-out proclivities. Specfically, conditional on $x^* < 1.5$, half the workers have $\beta_i = 0.5$ and/or $\eta_i = 2$ (depending on which parameters exhibit heterogeneity); conditional on $x \in [1.5, 3.5]$, half have $\beta_i = 0.8$ and/or $\eta_i = 1$; and conditional on $x > 3.5$, half have $\beta_i = 1$ and/or $\eta_i = 0.5$. In each case, the rest of the workers divide equally between the remaining parameter values.

[9] For a discussion of analytical properties and illustrative uses of linear-exponential utility functions, see Martinez-Mora and Puy (2012).

Table 1 summarizes the various parametric specifications and Figure 1 illustrates the distributions and utility functions used in the analysis.

4.2 Simulation results

Table 2 summarizes our simulation results.[10] Each row represents a separate simulation. Columns (1) through (5) provide details concerning the parametrization; Columns (6) through (13) present pertinent simulation results for different values of the cost-scaling parameter γ. For each simulation, we choose the value of the scaling-parameter to achieve the opt-out frequencies listed at the top of the columns: 95%, 90%, 75%, and 40%.[11] Converting values of γ into their implied opt-out frequencies renders the size of the parameter more easily interpretable.[12]

For each specification and opt-out frequency, the table reports the distance between the welfare-maximizing default option $D_L(\gamma)$ and the opt-out-minimizing default option $D_P(\gamma)$, as well as the fraction of the potential welfare gain, $\Omega(\gamma)$, achieved by the opt-out-minimizing default option relative to a zero-default policy. Both of these metrics require some explanation.

First we clarify the interpretation and measurement of $D_P(\gamma)$. In Section 3, we interpreted D^* in settings with bias ($\beta \neq 1$) as the limiting opt-out-minimizing default rate conditional on setting the optimal fine, K^*. It is therefore appropriate to use a parallel definition for $D_P(\gamma)$. Notably, Lemma 4 assures us that the optimal fine is independent of D. However, with heterogeneous biases, the optimal fine can vary with D. Accordingly, we interpret $D_P(\gamma)$ more generally as the opt-out-minimizing default rate conditional on setting the optimal fine for each D. Thus, for each simulation, we first find the default $D_L(\gamma)$ and fine K^* that maximize welfare, then we minimize opt-out for the same γ using the value of K^* obtained in the previous step. In Table 2, we then report the absolute value of the difference between the two defaults thus computed, $|D_L(\gamma) - D_P(\gamma)|$.

To compute $\Omega(\gamma)$, we first evaluate the welfare gain achieved by the welfare-optimal policy relative to a baseline scenario in which the default is non-participation ($D = 0$): $L\left(D_L(\gamma), \gamma\right) - L(0, \gamma)$. Next we calculate the welfare gain achieved by the opt-out minimizing policy relative to the same baseline: $L\left(D_P(\gamma), \gamma\right) - L(0, \gamma)$. We then define $\Omega(\gamma)$ as the ratio of the second welfare gain to the first,

[10] We performed all simulations using Python3 and Scipy. We employ the Limited-Memory approximation to the Broyden–Fletcher–Goldfarb–Shanno algorithm with Simplex Box constraints. We employ a grid-search over multiple starting points to ensure we reach a global maximum rather than one of potentially many local maxima. We calculated all integrals numerically using quadrature. We cross-checked all results in Julia using the COBYLA implemented in the NLopt package, with multiple starting points. We employed a maximal function value tolerance of $1e-11$ and maximal absolute quadrature error of $1e-12$.

[11] We select γ so that the opt-out rate under the welfare-maximizing default matches the stated target rate. For the same γ, the opt-out minimizing default necessarily leads to lower opt-out rates.

[12] By way of comparison, in the sample studied by Choukhmane (2019), opt-out rates in a 401(k) pension plan vary by tenure from about 20% to about 75%.

Table 1: Utility Functions and Distribution Functions Used in Numerical Simulations

Name	**Function**	**Parameterization**
Quadratic	$V(x, D) = -\alpha(x - D)^2$	$\alpha = 0.5$
Linear-Exponential	$V(x, D) = -\exp(\alpha(x - D)) + \alpha(x - D) + 1$	$\alpha = 0.75$

Table 1a): Utility functions used in the numerical simulations.

Distribution	**Mean**	**Median**	**Variance**	**Maximand(s)**
Truncated Normal $f(x) = H * \phi\,(x - 2.5)$	2.5	2.5	≈ 0.911	2.5
Right-peaked $f(x) = H * x$	$3.\overline{3}$	≈ 3.538	≈ 1.389	5
Bimodal $f(x) = H * \left(\frac{1}{(x-3)^2+\frac{1}{10}} + \frac{1}{(x-2)^2+\frac{1}{20}}\right)$	≈ 2.408	≈ 2.245	≈ 0.583	$\{2, 3\}$

Table 1b): Probability density functions $f(x)$ for the distributions used in the numerical simulations. For all distributions, the range is $x \sim [0, 5]$ and H is a normalization constant that ensures the density sums to 1.

Heterogeneity?	**Distribution of β**	**Distribution of η**
No Heterogeneity	$Pr[\beta = 0.8] = 1$	$Pr[\eta = 1] = 1$
Independence	$Pr[\beta = 0.5] = 1/3$ $Pr[\beta = 0.8] = 1/3$ $Pr[\beta = 1] = 1/3$	$Pr[\eta = 0.5] = 1/3$ $Pr[\eta = 1] = 1/3$ $Pr[\eta = 2] = 1/3$
Non-independence	$Pr[\beta = 0.5] = \begin{cases} 0.5 & x < 1.5 \\ 0.25 & x \geq 1.5 \end{cases}$ $Pr[\beta = 0.8] = \begin{cases} 0.5 & x \in [1.5, 3.5] \\ 0.25 & \text{otherwise} \end{cases}$ $Pr[\beta = 1] = \begin{cases} 0.5 & x > 3.5 \\ 0.25 & x \leq 3.5 \end{cases}$	$Pr[\eta = 0.5] = \begin{cases} 0.5 & x < 1.5 \\ 0.25 & x \geq 1.5 \end{cases}$ $Pr[\eta = 1] = \begin{cases} 0.5 & x \in [1.5, 3.5] \\ 0.25 & \text{otherwise} \end{cases}$ $Pr[\eta = 2] = \begin{cases} 0.5 & x > 3.5 \\ 0.25 & x \leq 3.5 \end{cases}$

Table 1c): Types of heterogeneity studied in the numerical simulations: 1) no heterogeneity in β and η, 2) independent random heterogeneity in one or both of β and η, and 3) heterogeneity in one or both of β and η, with dependence on x.

Figure 1: Utility Functions and Distribution Functions for Numerical Simulations: Illustrations

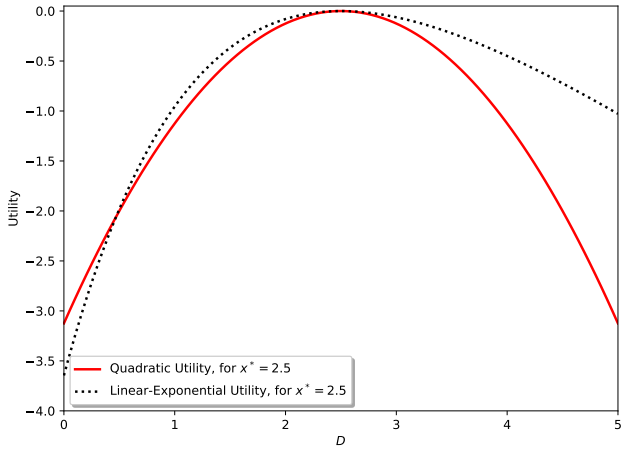

Figure 1a): Quadratic utility (in solid red) and linear-exponential asymmetric utility (in dotted black) for defaults $D \in [0, 5]$ given ideal point $x^* = 2.5$.

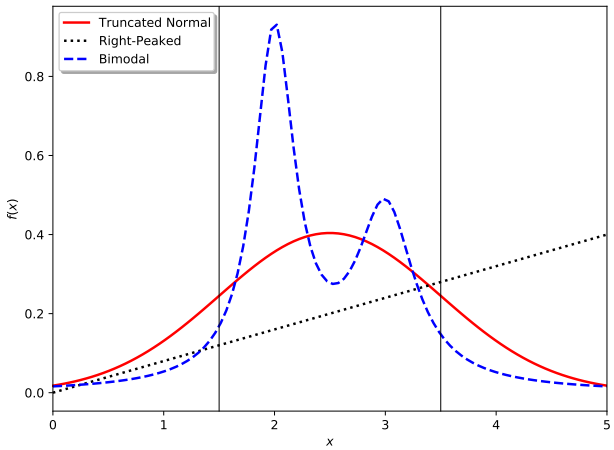

Figure 1b): Density of ideal point x^* over support $x \in [0, 5]$ for the three distributions studied: in solid red, the truncated Normal distribution, in dotted black, the right-peaked distribution, and in dashed blue the bimodal distribution.

expressed as a percentage: $\Omega(\gamma) = 100\% \frac{L(D_P(\gamma),\gamma) - L(\mathbf{0},\gamma)}{L(D_L(\gamma),\gamma) - L(\mathbf{0},\gamma)}$.

The accuracy of the limiting approximation

The first step in our simulation analysis is to ask whether the theoretical characterization of optimal policy (from Section 3), which describes the limiting case, provides a reasonable approximation for settings with meaningful social stakes. Pertinent results appear in Part A of Table 2, which encompasses simulations in which heterogeneity is limited to ideal points. Several notable patterns emerge. First, we see numerical corroboration of Proposition 2: in each case, when γ is low enough to produce an opt-out frequency of 95%, $D_P(\gamma)$ and $D_L(\gamma)$ are nearly identical. The maximal difference between the two is 0.0005 (which occurs for right-peaked preference distribution and quadratic utility), which corresponds to only 0.04% of the standard deviation of the ideal points x^*, and only 0.01% of its total range. The percent of the potential welfare gain achieved through opt-out minimization, $\Omega(\gamma)$, is larger than 99.99 in all but one case.

Second, for higher opt-out costs (lower opt-out rates), the correspondence between the two defaults remains close. With 75% opt-out, the maximal distance between $D_P(\gamma)$ and $D_L(\gamma)$ (which again occurs for right-peaked preference distribution and quadratic utility) is 0.0161, which corresponds to only 1.4% of the standard deviation of x^* and only 0.27% of its range. In that simulation, the opt-out minimizing default achieves 99.83% of the total attainable welfare improvement. In other distribution/utility specifications the percentage of welfare gain achieved by opt-out minimization is even greater. For the smallest opt-out percentage considered in the table, 40%, the approximations remain surprisingly good. The largest difference between $D_P(\gamma)$ and $D_L(\gamma)$ is just under 0.4, which represents less than a third of a standard deviation, and in every case opt-out minimization achieves at least 87.5% of the potential welfare gain; in fact, it achieves more than 98% of the potential gain for two of the five specifications.

Figure 2, which focuses on the specification with an asymmetric linear-exponential utility function along with a bimodal ideal-point distribution, shows the relationship between $D_P(\gamma)$ and $D_L(\gamma)$ for $\gamma \in [5e-4, 0.25]$, which yields opt-out frequencies between roughly 18% and 92%. Even with relatively low opt-out frequencies (high opt-out costs), the divergence between the two default rates is modest, and the percentage of the total potential welfare gain achieved through opt-out minimization is high for parameters that produce opt-out rates above 50%. An interesting feature of the figure is that neither the percentage welfare gain achieved nor the absolute distance between the two defaults is monotonic in the welfare-maximizing opt-out percentage. As that percentage rises from 20% and 45% (due to declining opt-out costs), we observe divergence between the two defaults and a reduction in the percentage welfare gain achieved. However, as we increase the welfare-maximizing opt-out rate further, the two metrics behave as predicted by the theorem: for small cost parameters (i.e., large opt-out rates), the two defaults converge, and opt-out minimization leaves only a trivial portion of the potential welfare gains unrealized.

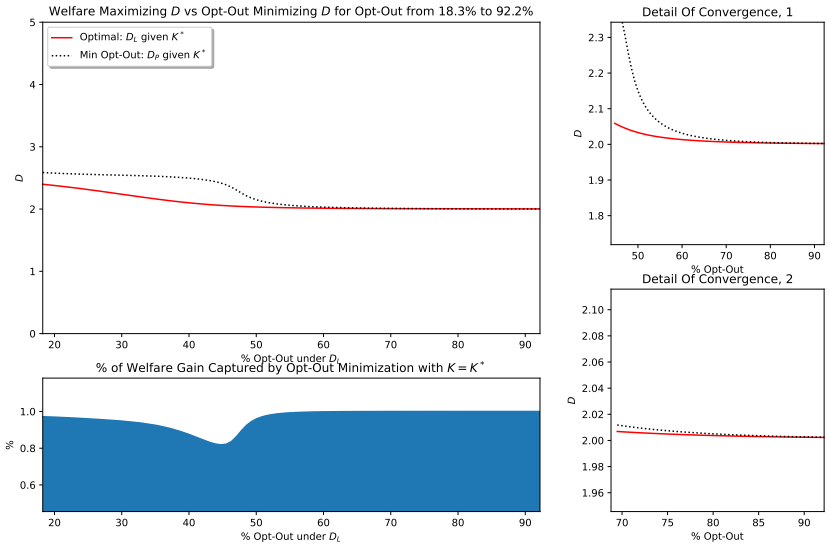

Figure 2: Illustration of welfare-maximizing and opt-out-minimizing defaults. The main panel shows the welfare-maximizing default $D_L(\gamma)$ and the opt-out minimizing default $D_P(\gamma)$ (where the latter is conditional on the welfare-maximizing fine K^*), plotted for $\gamma \in [5e-4, 0.25]$, which yields opt-out frequencies between 18% and 92% for the welfare-maximizing default. The utility function is linear-exponential, with asymmetry factor $\alpha = 0.75$, the preference density is bimodal with peaks at 2 and 3. The present bias parameter is $\beta = 0.8$ and opt-out costs are $\eta = 1$ for all agents. Detail Panels 1 and 2 reproduce the main panel at higher resolution for opt-out frequencies closer to unity (γ close to zero) and with a zoomed-in y-axis. The bottom panel displays the percentage of the potential welfare gain achieved through the opt-out-minimization policy $(D_P(\gamma), K^*)$.

Additional dimensions of heterogeneity

The second step in our simulation analysis is to ask whether our main findings are robust with respect to the additional dimensions of heterogeneity.

We begin with heterogeneity in the opt-out cost parameter η. To the extent η is unrelated to x^*, the population consists of a set of η-indexed subgroups that are otherwise identical. If every group's opt-out cost converges to zero via the common scaling factor γ, then the group-specific optimal default options all converge to D^*, and hence the population-wide optimal default rate should also converge to D^*. Accordingly, the cases of greatest concern are those in which η and x^* are correlated.

Part B of Table 2 exhibits results for simulations in which the η terms are heterogeneous but uncorrelated with x^*, while Part E displays results for simulations in which these terms are correlated. For some specifications, small discrepancies between $D_P(\gamma)$ and $D_L(\gamma)$ remain when opt-out costs are low (so that the welfare-maximizing opt-out rate is 95%), but those differences are no greater than 0.0075 assuming independence, and 0.0092 assuming correlation (respectively, less than 0.6% and 0.8% of the standard deviation of x^*). In all of these simulations, the opt-out minimizing default captures more than 99% of the potential welfare gain. Even for the smallest opt-out percentage considered in the table, 40%, the largest difference between $D_P(\gamma)$ and $D_L(\gamma)$ is just under 0.3, which represents roughly one-quarter of a standard deviation, and in every case opt-out minimization achieves more than 94% of the potential welfare gain.

Next we turn to heterogeneity in β. Assuming the fine is optimized for workers with some intermediate value of β, those with high β will opt out too much and those with low β will opt out too little. Increasing (resp. reducing) D will tend to increase (resp. decrease) opt-out for those with low values of x^*, and decrease (resp. increase) opt-out for those with high values of x^*. Consequently, it is reasonable to conjecture that the desirability of shifting D in either direction from D^* depends on the existence of correlation between x^* and β, a possibility that our simulations encompass.

Part C of Table 2 exhibits simulation results for specifications with heterogeneity in β but without correlation between β and x^*, while part F studies cases with correlation. Once again, we see only modest divergence between $D_P(\gamma)$ and $D_L(\gamma)$ for large opt-out costs (i.e., with a welfare-maximizing opt-out rate of 40%). The largest such distance, 0.370, represents less than one-third of the standard deviation of x^*. Moreover, opt-out minimization achieves at least 89% of the potential welfare gain, and in five of the ten simulations that fraction exceeds 98%. The modest discrepancies between $D_P(\gamma)$ and $D_L(\gamma)$ shrink rapidly as opt-out costs fall (and opt-out frequencies rise). For example, in simulations with a 75% opt-out rate, the difference is always less than 0.13 (roughly 11% of a standard deviation), and opt-out minimization achieves more than 96% of the potential welfare gains in every simulation.

Parts D and G of Table 2 exhibit simulation results for specifications with heterogeneity in both β and η, respectively without and with correlation between β and η on the one hand and x^* on the other. In terms of the magnitudes of the

discrepancies between $D_P(\gamma)$ and $D_L(\gamma)$ and the values of $\Omega(\gamma)$, the results are similar to those displayed in Parts B, C, E, and F of the table, which pertain to simulations with heterogeneity in either η or β, but not both.

5 Conclusion

In this paper, we have shown that, in addition to providing a practically implementable criterion for setting default options, opt-out minimization also has a solid and general normative foundation. In this concluding section, we briefly mention some potential avenues for future work.

Further explorations of generality could usefully test the limits of our conclusions. The following two issues merit additional scrutiny. First, while the framework used here potentially accomodates many types of decision-making biases (Goldin and Reck 2019), other important classes of bias may require different formulations. As an example, the model of mechanistic (as opposed to optimal) inattention in Bernheim, Fradkin, and Popov (2015) involves a different formulation. Second, as noted in Section 2, the literature has conceptualized opt-out costs as arising from the mechanics of implementation, rather than from deliberation. Because the latter mechanism seems plausbile in many settings, it merits further study. One can imagine a class of models in which the worker starts with a diffuse prior over the best default rate and can refine that prior by acquiring a costly signal. A worker whose prior aligns insufficiently with the default will incur the cost of signal aquisition, and then potentially opt out depending on what the signal reveals. It would be of interest to examine the robustness of our conclusions to these types of possibilities.

Finally, because default options are ubiquitous features of real-world choices, it is worth examining applications other than contribution rates in employee-directed pension plans. Some applications may raise issues that call for new modeling wrinkles and lead to additional insights.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	Heterogeneity					95% opt-out		90% opt-out		75% opt-out		40% opt-out	
	η	β	Interdep.	Utility Func.	Distribution	$\|\|D_L - D_P\|\|$	Δ Welfare %	$\|\|D_L - D_P\|\|$	Δ Welfare %	$\|\|D_L - D_P\|\|$	Δ Welfare %	$\|\|D_L - D_P\|\|$	Δ Welfare %
				Quadratic	Right-peaked	0.00053	99.99	0.00227	99.98	0.01609	99.83	0.06963	99.18
				Quadratic	Bimodal	2e-05	>99.99	0.00011	>99.99	0.00123	>99.99	0.35162	87.56
A	X	X	X	Lin.-Exp.	Trunc. Normal	0.00019	>99.99	0.00077	>99.99	0.00499	>99.99	0.03606	99.94
				Lin.-Exp.	Right-peaked	0.00053	99.99	0.00222	99.98	0.01469	99.84	0.103	98.74
				Lin.-Exp.	Bimodal	7e-05	>99.99	0.00026	>99.99	0.00251	99.99	0.39662	87.50
				Quadratic	Right-peaked	0.00753	99.60	0.0188	99.41	0.08014	98.45	0.28705	95.21
				Quadratic	Bimodal	3e-05	>99.99	0.00013	>99.99	0.00127	>99.99	0.14811	97.81
B	✓	X	X	Lin.-Exp.	Trunc. Normal	0.00023	>99.99	0.00093	>99.99	0.00594	>99.99	0.03643	99.94
				Lin.-Exp.	Right-peaked	0.00739	99.61	0.0178	99.44	0.07066	98.65	0.02461	99.96
				Lin.-Exp.	Bimodal	0.0001	>99.99	0.0003	>99.99	0.00234	99.99	0.19257	97.26
				Quadratic	Right-peaked	0.01591	98.36	0.03369	98.02	0.09991	96.72	0.25222	93.65
				Quadratic	Bimodal	3e-05	>99.99	0.00013	>99.99	0.00142	>99.99	0.32371	89.91
C	X	✓	X	Lin.-Exp.	Trunc. Normal	0.00024	>99.99	0.00098	>99.99	0.00631	>99.99	0.04336	99.91
				Lin.-Exp.	Right-peaked	0.01563	98.41	0.03245	98.13	0.08832	97.16	0.07672	99.17
				Lin.-Exp.	Bimodal	9e-05	>99.99	0.00032	>99.99	0.0028	99.99	0.36944	89.57
				Quadratic	Right-peaked	0.03077	97.35	0.06754	96.61	0.06748	98.87	0.17024	98.31
				Quadratic	Bimodal	6e-05	>99.99	0.00024	>99.99	0.00184	>99.99	0.1103	98.70
D	✓	✓	X	Lin.-Exp.	Trunc. Normal	0.00047	>99.99	0.00183	>99.99	0.01104	99.99	0.0487	99.90
				Lin.-Exp.	Right-peaked	0.03013	97.43	0.06404	96.80	0.05921	99.02	0.17318	98.44
				Lin.-Exp.	Bimodal	0.00014	>99.99	0.00052	>99.99	0.00309	99.99	0.15427	98.16
				Quadratic	Right-peaked	0.00923	99.54	0.02293	99.31	0.10098	98.07	0.26364	97.71
				Quadratic	Bimodal	3e-05	>99.99	0.00012	>99.99	0.00124	>99.99	0.20033	95.73
E	✓	X	✓	Lin.-Exp.	Trunc. Normal	0.00021	>99.99	0.00087	>99.99	0.0055	>99.99	0.0439	99.93
				Lin.-Exp.	Right-peaked	0.00905	99.55	0.02179	99.34	0.08801	98.34	0.04756	99.92
				Lin.-Exp.	Bimodal	8e-05	>99.99	0.00028	>99.99	0.00231	99.99	0.25712	94.46
				Quadratic	Right-peaked	0.01814	98.20	0.03942	97.76	0.12721	96.02	0.0734	98.91
				Quadratic	Bimodal	3e-05	>99.99	0.00013	>99.99	0.00137	>99.99	0.32594	89.82
F	X	✓	✓	Lin.-Exp.	Trunc. Normal	0.00024	>99.99	0.00094	>99.99	0.00608	>99.99	0.11102	99.64
				Lin.-Exp.	Right-peaked	0.01781	98.25	0.03758	97.88	0.11281	96.55	0.08121	98.72
				Lin.-Exp.	Bimodal	8e-05	>99.99	0.00032	>99.99	0.00274	99.99	0.36971	89.43
				Quadratic	Right-peaked	0.03364	97.48	0.07393	96.67	0.10226	98.48	0.18126	98.41
				Quadratic	Bimodal	5e-05	>99.99	0.00021	>99.99	0.0017	>99.99	0.13973	97.88
G	✓	✓	✓	Lin.-Exp.	Trunc. Normal	0.00041	>99.99	0.00162	>99.99	0.00982	>99.99	0.07126	99.79
				Lin.-Exp.	Right-peaked	0.03291	97.55	0.07049	96.84	0.08961	98.71	0.07863	99.71
				Lin.-Exp.	Bimodal	0.00014	>99.99	0.00048	>99.99	0.00294	99.99	0.19049	97.01

Table 2: Overview of Simulation Results. Each row is a separate simulation. The cost-scaling parameter γ is chosen to achieve the opt-out frequency detailed in columns (6) through (13). Columns (1) through (3) specify the presence and type of heterogeneity: a checkmark indicates heterogeneity in η (Column (1)), β (Column (2)), and that any present heterogeneity is dependent on the preference distribution as detailed in Table 1.c (Column (3)). Column (4) indicates the utility function; Column (5) specifies the preference distribution. Columns (6), (8), (10), and (12) report the absolute distance between the welfare maximizing default and the opt-out minimizing default for the given opt-out level; Columns (7), (9), (11), and (13) display the percentage of the welfare increase of having an optimal policy versus no policy that is captured by the opt-out minimizing default. A value of ≥ 99.99 indicates that the percentage would round to 100.00. We omit the combination of no heterogeneity, truncated Normal preference distribution, and quadratic utility, as the two defaults coincide for any γ.

References

Agnew, Julie R., and Lisa R. Szykman. 2005. Asset allocation and information overload: the influence of information display, asset choice, and investor experience. *Journal of Behavioral Finace* 6 (2): 57–70.

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. doi:10.1257/aer.20130907. https://www.aeaweb.org/articles?id=10.1257/aer.20130907.

Bernheim, B Douglas, and Dmitry Taubinsky. 2018. Behavioral public economics. In *Handbook of behavioral economics: applications and foundations 1*, 1:381–516. Elsevier.

Beshears, John, James J Choi, David Laibson, and Brigitte C Madrian. 2018. Behavioral household finance. In *Handbook of behavioral economics: applications and foundations 1*, 1:177–276. Elsevier.

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): 1639–1674.

Choukhmane, Taha. 2019. Default options and retirement saving dynamics. In *112th annual conference on taxation*. NTA.

Goldin, Jacob, and Daniel Reck. 2019. Optimal defaults with normative ambiguity. https://ssrn.com/abstract=2893302.

Handel, Benjamin R., and Jonathan T. Kolstad. 2015. Health insurance for "humans": information frictions, plan choice, and consumer welfare. *American Economic Review* 105 (8): 2449–2500.

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): 1149–1187.

Martinez-Mora, Francisco, and M Socorro Puy. 2012. Asymmetric single-peaked preferences. *The BE Journal of Theoretical Economics* 12 (1).

Thaler, Richard H, and Cass R Sunstein. 2003. Libertarian paternalism. *American Economic Review* 93 (2): 175–179.

Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: improving decisions about health, wealth, and happiness*.

A Mathematical Appendix

In this appendix, we prove the four lemmas stated in the main text. The proofs of the three proposition follow directly from the lemmas by the arguments given in the text.

Proof of Lemma 1

Consider any $x_1 < D$ and $x_2 \in (x_1, D)$. Then

$$\Delta(D, x_1) = V(x_1, x_1) - V(D, x_1) = -\int_{x_1}^{D} V_1(z, x_1)dz$$

$$> -\int_{x_2}^{D} V_1(z, x_1)dz > -\int_{x_2}^{D} V_1(z, x_2)dz = V(x_2, x_2) - V(D, x_2) = \Delta(D, x_2)$$

where the first inequality follows from the concavity of V (Assumption 1, (ii)) (which ensures $V_1(z, x_1) < 0$ for $z \in (x_1, D)$), and the second follows from single crossing ($V_{12} > 0$) (Assumption 1, (iii)). It follows that opt-out at x_2 implies opt-out at x_1, and opt-in at x_1 implies opt-in at x_2. An analogous argument establishes that a symmetric property holds for $x_1 > D$ and $x_2 \in (D, x_1)$. Furthermore, $\triangle(D, x)$ inherits continuity from V. Thus, the opt-in set is a closed interval with indifference at the boundaries (whenever they are interior to X) and strict preference on the interior. □

Proof of Lemma 2

The proof proceeds in a series of steps. The first step references the opt-in window, $S(D, \gamma) \equiv [x_l(D, \gamma), x_h(D, \gamma)]$. Throughout, we use the symbol $\rightrightarrows$ to denote uniform convergence.

Step 1: $|S(D, \gamma)| \rightrightarrows 0$ *as* $\gamma \to 0$.

Using Taylor's theorem, we know there is some $\tilde{x}(D, x) \in [D, x]$ such that[13]

$$\Delta(D, x) = -\frac{1}{2} V_{11}(\tilde{x}(D, x), x)(D - x)^2$$

It will be convenient to define

$$d(D, x) \equiv -\frac{1}{2} V_{11}(\tilde{x}(D, x), x)$$

so that $\Delta(D, x) = d(D, x)\,(D - x)^2$. Under part (ii) of Assumption 1, we have $d(D, x) > \frac{v^{min}}{2} > 0$ for all D, x.

Next define

$$S^0(D, \gamma) \equiv [D - \omega(\gamma), D + \omega(\gamma)]$$

where

$$\omega(\gamma) \equiv \left(\frac{2\gamma}{v^{min}}\right)^{\frac{1}{2}}$$

[13] In applying the theorem, we have used the fact that $V_1(x, x) = 0$.

We claim that $S(D,\gamma) \subset S^0(D,\gamma)$. Consider any $x \in S(D,\gamma)$. Then, using our exact second-order approximation, we have $d(D,x)(D-x)^2 < \gamma$. Using the fact that $d(D,x) > \frac{v^{min}}{2}$ for all D, x, we see that $\frac{v^{min}}{2}(D-x)^2 < \gamma$. It then follows immediately that $x \in S^0(D,\gamma)$.

Now observe that

$$|S(D,\gamma)| \leq \left|S^0(D,\gamma)\right| \leq 2\omega(\gamma).$$

Notice that this term vanishes uniformly over D as $\gamma \to 0$.

Step 2: There exists a function $\delta(\gamma)$ with $\lim_{\gamma\to 0}\delta(\gamma) = 0$ such that, for all $D \in X$ and $x \in S(D,\gamma)$, we have

$$|f(D) - f(x)| < \delta(\gamma) \tag{5}$$

and

$$|d(D,D) - d(D,x)| < \delta(\gamma) \tag{6}$$

.

First consider f. Because F is twice-continuously differentiable and X is compact, f is Lipschitz-continuous on X. Accordingly, there exists $K_f > 0$ such that $|f(D) - f(x)| < K_f |D - x|$. In Step 1, we showed that $|D - x| \leq \omega(\gamma)$ for $x \in S(D,\gamma)$. Therefore $|f(D) - f(x)| < K_f\omega(\gamma)$ for $x \in S(D,\gamma)$.

Now consider d. Because V has continuous third derivatives and X is compact, V_{11} is Lipschitz-continuous on X^2. Accordingly, there exists $K_v > 0$ such that

$$|d(D,D) - d(D,x)| = \frac{1}{2}|V_{11}(\tilde{x}(D,x),x) - V_{11}(D,D)| < K_v |D - \tilde{x}(D,x)| \tag{7}$$

For $x \in S(D,\gamma)$, we have $\tilde{x}(D,x) \in [D,x] \subseteq S(D,\gamma)$, where the set inclusion follows from Lemma 1. In Step 1, we showed that $|D - x'| \leq \omega(\gamma)$ for $x' \in S(D,\gamma)$. Setting $x' = \widetilde{x}(D,x)$ and substituting into (7), we obtain $|d(D,D) - d(D,x)| < K_v\omega(\gamma)$ for $x \in S(D,\gamma)$.

To complete Step 2, we simply define $\delta(\gamma) \equiv \max\{K_f, K_v\} \cdot \omega(\gamma)$.

Step 3: Proof of the lemma.

From Step 2, we know that for all $x \in S(D,\gamma)$, we have $d(D,D) - \delta(\gamma) < d(D,x) < d(D,D) + \delta(\gamma)$. It follows that, for such x,

$$(d(D,D) - \delta(\gamma))(D-x)^2 < \Delta(D,x) < (d(D,D) + \delta(\gamma))(D-x)^2$$

Accordingly, $\Delta(D,x) < \gamma$ implies $(D-x)^2 < \frac{\gamma}{d(D,D)-\delta(\gamma)}$, and $(D-x)^2 > \frac{\gamma}{d(D,D)-\delta(\gamma)}$ implies $\Delta(D,x) > \gamma$. Thus,

$$S(D,\gamma) \subset \left(D - \left(\frac{\gamma}{d(D,D)-\delta(\gamma)}\right)^{\frac{1}{2}}, D + \left(\frac{\gamma}{d(D,D)-\delta(\gamma)}\right)^{\frac{1}{2}}\right) \tag{8}$$

$$S(D,\gamma) \supset \left(D - \left(\frac{\gamma}{d(D,D)+\delta(\gamma)}\right)^{\frac{1}{2}}, D + \left(\frac{\gamma}{d(D,D)+\delta(\gamma)}\right)^{\frac{1}{2}}\right) \tag{9}$$

Using these inclusion relations and along with the fact that $f(D) - \delta(\gamma) < f(x) < f(D) + \delta(\gamma)$ for all $x \in S(D, \gamma)$, we then have

$$2\left(f(D)+\delta(\gamma)\right)\left(\frac{\gamma}{d(D,D)-\delta(\gamma)}\right)^{\frac{1}{2}} > \Pr[\Delta(D,x)<\gamma] > 2\left(f(D)-\delta(\gamma)\right)\left(\frac{\gamma}{d(D,D)+\delta(\gamma)}\right)^{\frac{1}{2}}$$

It thus follows that

$$\left(f(D)+\delta(\gamma)\right)\left(\frac{1}{-V_{11}(D,D)-2\delta(\gamma)}\right)^{\frac{1}{2}} > Q(D,\gamma) > \left(f(D)-\delta(\gamma)\right)\left(\frac{1}{-V_{11}(D,D)+2\delta(\gamma)}\right)^{\frac{1}{2}}.$$

As $\gamma \to 0$, both sides converge to the same value: $f(D)\left(\frac{1}{d(D,D)}\right)^{\frac{1}{2}} = \widetilde{Q}(D)$. Therefore we know that $Q(D,\gamma)$ converges pointwise to $\widetilde{Q}(D)$.

To show that convergence is uniform, notice first that $\widetilde{Q}(D)$ lies within the same bounds. We consider the difference between the upper and lower bounds on $Q(D,\gamma)$ and $\widetilde{Q}(D)$:

$$\xi(D,\gamma) = \left(f(D)+\delta(\gamma)\right)\left(\frac{1}{-V_{11}(D,D)-2\delta(\gamma)}\right)^{\frac{1}{2}} - \left(f(D)-\delta(\gamma)\right)\left(\frac{1}{-V_{11}(D,D)+2\delta(\gamma)}\right)^{\frac{1}{2}} > 0$$

Notice that this expression is increasing in $f(D)$ and decreasing in $-V_{11}(D,D)$. Because we have assumed that f is continuous, it obtains a maximum, f^{max}, on the compact set X. Thus,

$$\xi(D,\gamma) < \left(f^{max}+\delta(\gamma)\right)\left(\frac{1}{v^{min}-2\delta(\gamma)}\right)^{\frac{1}{2}} - \left(f^{max}-\delta(\gamma)\right)\left(\frac{1}{v^{min}+2\delta(\gamma)}\right)^{\frac{1}{2}}$$

The right-hand side of this expression converges to 0 as $\gamma \to 0$, and does not depend upon D. Therefore, we have $Q(D,\gamma) \rightrightarrows \widetilde{Q}(D)$. □

Proof of Lemma 3

Because $0 < \mathbb{E}\left[\Delta(D,x)|\Delta(D,x)<\gamma\right] < \gamma$ for all γ, we know that $Z(D,\gamma)$ is bounded between 0 and 1. Observe that:

$$Z(D,\gamma) = \frac{\mathbb{E}\left[\Delta(D,x)|\Delta(D,x)<\gamma\right]}{\gamma} = \frac{\mathbb{E}\left[\Delta(D,x)\mathbf{1}_{\Delta(D,x)<\gamma}\right]}{\gamma \Pr\left[\Delta(D,x)<\gamma\right]} \tag{10}$$

The denominator equals $Q(D)\gamma^{\frac{3}{2}}$.

Defining $\delta(D)$ as in the proof of Lemma 2, as long as γ is sufficiently small to ensure $\delta(\gamma) < v^{min}$, the numerator of (10) is bounded above by:

$$\mathbb{E}\left[\Delta(D,x)\mathbf{1}_{\Delta(D,x)<\gamma}\right] \leq \int_{D-\left(\frac{\gamma}{d(D,D)-\delta(\gamma)}\right)^{\frac{1}{2}}}^{D+\left(\frac{\gamma}{d(D,D)-\delta(\gamma)}\right)^{\frac{1}{2}}} \left(d(D,D)+\delta(\gamma)\right)(D-x)^2\left(f(D)+\delta(\gamma)\right)dx$$

$$= \frac{1}{3}(f(D)+\delta(\gamma))(d(D,D)+\delta(\gamma))\gamma^{\frac{3}{2}}\left(\frac{1}{(d(D,D)-\delta(\gamma))^{\frac{3}{2}}}+\frac{1}{(d(D,D)+\delta(\gamma))^{\frac{3}{2}}}\right)$$

$$= \frac{1}{3}\tilde{Q}(D)\left(1+\frac{\delta(\gamma)}{f(D)}\right)\left(1+\frac{\delta(\gamma)}{d(D,D)}\right)\gamma^{\frac{3}{2}}\left(\left(1-\frac{\delta(\gamma)}{d(D,D)}\right)^{-\frac{3}{2}}+\left(1+\frac{\delta(\gamma)}{d(D,D)}\right)^{-\frac{3}{2}}\right)$$

where the inequality in the first line follows from (5), (6), and (8) (given that the integrand is strictly positive). It then follows from (10) that

$$Z(D,\gamma) \leq \frac{1}{3}\left(\frac{\tilde{Q}(D)}{Q(D,\gamma)}\right)\left(1+\frac{\delta(\gamma)}{f(D)}\right)\left(1+\frac{\delta(\gamma)}{d(D,D)}\right)\left(\frac{\left(1-\frac{\delta(\gamma)}{d(D,D)}\right)^{-\frac{3}{2}}+\left(1+\frac{\delta(\gamma)}{d(D,D)}\right)^{-\frac{3}{2}}}{2}\right)$$

$$\equiv \overline{Z}(D,\gamma)$$

With $f(D)$ and $d(D,D)$ bounded below by $f^{min} > 0$ and $v^{min} > 0$, respectively, it is straightforward to check that $1+\frac{\delta(\gamma)}{f(D)} \rightrightarrows 1$, $1+\frac{\delta(\gamma)}{d(D,D)} \rightrightarrows 1$, and $1-\frac{\delta(\gamma)}{d(D,D)} \rightrightarrows 1$ as $\delta \to 0$. From Lemma 2, we also know that $Q(D,\gamma) \rightrightarrows \tilde{Q}(D)$. Because $V_{11}$is continuous, $-V_{11}(D,D)$ achieves a maximum, call it v^{max}, on the compact set X. Thus, $0 < f^{min}\left(\frac{1}{v^{max}}\right)^{\frac{1}{2}} \leq \tilde{Q}(D) \leq f^{max}\left(\frac{1}{v^{min}}\right)^{\frac{1}{2}}$. In light of these bounds, it is straightforward to check that $\frac{\tilde{Q}(D)}{Q(D,\gamma)} \rightrightarrows 1$ as $\delta \to 0$. Putting these observations together, we have $\overline{Z}(D,\gamma) \rightrightarrows \frac{1}{3}$ as $\gamma \to 0$.

Similarly, as long as γ is sufficiently small to ensure $\delta(\gamma) < \min\left\{v^{min}, f^{min}\right\}$, the numerator of (10) is bounded below by:

$$\mathbb{E}\left[\Delta(D,x)\mathbf{1}_{\Delta(D,x)<\gamma}\right] \geq \int_{D-\left(\frac{\gamma}{d(D,D)+\delta(\gamma)}\right)^{\frac{1}{2}}}^{D+\left(\frac{\gamma}{d(D,D)+\delta(\gamma)}\right)^{\frac{1}{2}}} (d(D,D)-\delta(\gamma))\,(D-x)^2\,(f(D)-\delta(\gamma))\,dx$$

A parallel argument then implies that

$$Z(D,\gamma) \geq \frac{1}{3}\left(\frac{\tilde{Q}(D)}{Q(D,\gamma)}\right)\left(1-\frac{\delta(\gamma)}{f(D)}\right)\left(1-\frac{\delta(\gamma)}{d(D,D)}\right)\left(\frac{\left(1-\frac{\delta(\gamma)}{d(D,D)}\right)^{-\frac{3}{2}}+\left(1+\frac{\delta(\gamma)}{d(D,D)}\right)^{-\frac{3}{2}}}{2}\right)$$

$$\equiv \underline{Z}(D,\gamma)$$

Reasoning as for the upper bound, we have $\underline{Z}(D,\gamma) \rightrightarrows \frac{1}{3}$ as $\gamma \to 0$.

Because the upper and lower bounds both converge uniformly to $\frac{1}{2}$, we can infer that $Z(D,\gamma) \rightrightarrows \frac{1}{3}$ as $\gamma \to 0$. □

Proof of Lemma 4

In light of (4), we can write the total loss associated with any value of γ and policy (D, K, B) as follows:

$$L(D,\gamma,K,B) = \int_{x_l\left(D,\frac{\gamma-K}{\beta}\right)}^{x_h\left(D,\frac{\gamma-K}{\beta}\right)} [\Delta(D,x) - B + K]\, dF(x) + \int_{x\notin\left(x_l\left(D,\frac{\gamma-K}{\beta}\right),x_h\left(D,\frac{\gamma-K}{\beta}\right)\right)} [\gamma - B]\, dF(x). \tag{11}$$

From equation (3), we know that $B = \int_{x_l}^{x_u} K dF(x)$. It follows immediately that

$$L(D;\gamma) = \int_{x_l\left(D,\frac{\gamma-K}{\beta}\right)}^{x_h\left(D,\frac{\gamma-K}{\beta}\right)} [\Delta(D,x) - \gamma]\, dF(x) + \gamma.$$

Notice that the integrand is strictly negative for $x \in (x_l(D,\gamma), x_h(D,\gamma))$ and strictly positive for $x \notin [x_l(D,\gamma), x_h(D,\gamma)]$. It follows immediately that the optimum for any D involves setting $K = (1-\beta)\gamma$, as claimed. □