NBER WORKING PAPER SERIES

OPTIMAL DEFAULT OPTIONS: THE CASE FOR (WEIGHTED) 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, Revised December 2021

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 (Weighted) Opt-Out Minimization
B. Douglas Bernheim and Jonas Mueller Gastell
NBER Working Paper No. 28254
December 2020, Revised December 2021
JEL No. D10,D11,D14

ABSTRACT

We examine the problem of setting optimal default options such as passively selected contribution rates in employee-directed pension plans. Existing results suggest that a simple rule of thumb, opt-out minimization, is optimal under special conditions, but this result is fragile, and the literature does not provide a general analytic solution. We demonstrate with considerable generality that weighted opt-out minimization is approximately optimal, and we provide clear mathematical intuition for the robustness of this result. We also identify surprisingly broad conditions under which unweighted opt-out minimization is approximately optimal. We conduct simulations to evaluate the accuracy of the approximation.

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

Jonas Mueller Gastell
Stanford University
jonasmg@stanford.edu

Optimal Default Options: The Case for (Weighted) Opt-Out Minimization

B. Douglas Bernheim and Jonas Mueller-Gastell*

December 22, 2021

Abstract

We examine the problem of setting optimal default options such as passively selected contribution rates in employee-directed pension plans. Existing results suggest that a simple rule of thumb, opt-out minimization, is optimal under special conditions, but this result is fragile, and the literature does not provide a general analytic solution. We demonstrate with considerable generality that *weighted* opt-out minimization is *approximately* optimal, and we provide clear mathematical intuition for the robustness of this result. We also identify surprisingly broad conditions under which unweighted opt-out minimization is approximately optimal. We conduct simulations to evaluate the accuracy of the approximation.

1 Introduction

Most decision problems implicitly or explicitly specify an option that serves as a default, in the following sense: if the consumer fails to make a choice, whether intentionally or by neglect, the default option will prevail. Default options may impact outcomes either because active choice requires the expenditure of effort, or because the identity of the default alters the psychology of choice. The ubiquity of default options gives rise to important normative questions about the optimal design of "choice architectures" (Thaler and Sunstein 2008). 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). 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).

*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

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. Their justification for the criterion is informal. The ensuing literature establishes that opt-out minimization can indeed be welfare-optimal under highly specialized conditions. For a setting in which workers differ with respect to ideal points and choice-rationalizing ("as if") opt-out costs, Goldin and Reck (2019) establish the optimality of opt-out minimization under the following sufficient conditions: (i) as-if opt-out costs are "sufficiently normative" (meaning that biases impacting opt-out decisions are not too large, so that differences between as-if and normative costs are limited), (ii) the utility derived from the action is a convex, single-peaked, symmetric function that depends only on the difference between the action and the worker's ideal point, (iii) the distribution of ideal points is single-peaked and symmetric, and (iv) ideal points are distributed independently of opt-out costs. Carroll et al. (2009) consider a specialized dynamic model in which present focus, which they interpret as a bias, causes workers to place excessive weight on opt-out costs. Their sufficient conditions for the optimality of opt-out minimization are similar those in Goldin and Reck (2020). Neither paper offers a general analytical characterization of optimal defaults for settings that violate these conditions, and one is left with the impression that the result may be fragile. Indeed, both studies find that opt-out maximization can also be optimal.

Curiously, analyses of empirically parametrized models suggests that opt-out minimization may be a more generally attractive policy than these theoretical results appear to imply. Bernheim, Fradkin, and Popov (2015) and Choukhmane (2019) find that the welfare-maximizing and opt-out-minimizing default rates often coincide, even when the Goldin-Reck assumptions are violated. Although the two can also diverge, "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).

The current paper makes two main contributions. The first is to provide a general characterization of approximately optimal defaults that links welfare maximization to *weighted* opt-out minimization. Our notion of approximate optimization entails an extrapolation from the limiting properties of the welfare-maximizing default options as the overall scale of as-if opt-out costs becomes small. Focusing for the sake of concreteness on the problem of setting a default contribution rate for a 401(k) pension plan, we consider enivronments with multiple dimensions of worker heterogeneity: workers differ not only with respect to their ideal points and as-if opt-out costs (as in Goldin and Reck 2019), but also with respect to the magnitudes of their biases (i.e., the differences between their as-if and normative opt-out costs) and the shapes of their continuation valuation functions. We impose no restrictions on correlations between these characteristics. Our main result demonstrates that opt-out minimization yields approximately optimal outcomes when opt-out frequencies are weighted according to the workers' characteristics, with weights given by the simple formula $\omega(\eta, \beta) \equiv \eta \left(1 - \frac{1}{3\beta}\right)$, where η measures the worker's level of opt-out costs relative to the

population mean and β represents her bias. To obtain this general result, we impose only a few mild technical restrictions along with a single-crossing requirement. We also provide straightforward mathematical intuition for the robustness of this result. Notably, our general characterization applies regardless of whether as-if opt-out costs are "sufficiently normative": if pervasive biases are large ($\beta < \frac{1}{3}$), our formula yields negative weights, in which case opt-out minimization can turn into opt-out maximization. Consequently, our analysis allows us to interpret the two polar cases identified in the literature – those for which opt-out minimization is optimal, and those for which opt-out maximization is optimal – as two sides of the same coin.

Given the focus of the previous literature, it is also important to investigate the circumstances under which *unweighted* opt-out minimization is approximately optimal. Our second main contribution is to demonstrate that these circumstances are much broader than previously thought. An immediate implication of our general characterization is that unweighted opt-out minimization is approximately optimal if and only if it coincides asymptotically with weighted opt-out minimization using the weights indicated above. We show that the solutions to these two problems coincide when ideal points are distributed independently of opt-out costs and biases. Thus, apart from our mild technical conditions and the single-crossing requirement, we dispense with the highly restrictive Goldin-Reck assumptions concerning continuation utility and distributions (labelled (ii) and (iii) above). Because we consider additional dimensions of worker heterogeneity, we also demonstrate that the approximate optimality of unweighted opt-out minimization does *not* require ideal points to be distributed independently of parameters governing the properties of continuation value functions, thereby limiting the scope of the assumption labeled (iv) above. Moreover, we show that, under specified conditions, if the employer can impose budget-neutral penalities for passive choice, then opt-out minimization, rather than opt-out maximization, is approximately welfare-optimal regardless of the degree of biases impacting the agent's decision, i.e., irrespective of whether opt-out costs are "sufficiently normative" (the Goldin-Reck assumption labeled (i) above).[1] We therefore conclude that the approximate optimality of unweighted opt-out minimization depends primarily on whether ideal points are distributed independently of opt-out costs and biases.

The preceding results pertain to settings in which the action choice is continuous and

[1]Bernheim, Fradkin, and Popov make a related point in arguing 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. We depart from Bernheim, Fradkin, and Popov by considering the natural possibility that the employer can establish non-dissipative fines; for example, 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. 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.

there are no atoms in the distribution of ideal points. In some contexts, atoms may appear at the boundaries of the opportunity set, or at special points on the interior of that set, such as caps on 401(k) contributions eligible for an employer match. In other contexts, the choice set may involve a small number of alternatives (possibly just two). We show that our general characterization extends to these settings, with a small adjustment: the weighting function that ensures the asymptotic equivalence of weighted opt-out minimization and welfare-maximization is simply $\omega(\eta) = \eta$. We explain this difference intuitively and discuss its implications. Most notably, because the weight is always strictly positive, the optimal strategy necessarily has the flavor of opt-out minimization rather than maximization, irrespective of whether as-if opt-out costs are "sufficiently normative" (the Goldin-Reck assumption labeled (i) above).

We illustrate our main convergence results by simulating optimal defaults in settings that violate specific assumptions imposed in Carroll et al. (2009) and Goldin and Reck (2019). These simulations also validate the asymptotic approximation by showing that the limiting case provides a decent guide for a range of reasonable settings with larger opt-out costs and, consequently, meaningful social stakes. We also use simulations to evaluate the cost of pursuing unweighted opt-out minimization in settings where weighted and unweighted opt-out minimization do not coincide.

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. Weighted opt-out minimization is also easy to implement but requires information on the joint distribution of the pertinent characteristics.

The remainder of the paper proceeds as follows. Section 2 details the model. Section 3 demonstrates the asymptotic optimality of weighted opt-out minimization. Section 4 identifies conditions under which unweighted opt-out minimization also coincides asymptotically with welfare maximization. Section 5 explains how our analysis applies to settings with normative ambiguity (as in the welfare framework of Bernheim and Rangel 2009). Section 6 examines extensions to settings with bunching (arising, for example, from boundary constraints or caps on matching provisions), and to decisions with finite opportunity sets. Section 7 describes our simulations. We close in Section 8 with some brief thoughts about directions for subsequent research. Abbreviated proofs appear in the Appendix.

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.

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 $x \neq D$.

For the sake of tractability, we assume the worker's utility is linear in income (m), and is additively separable in income, the contribution rate (x), and the effort exerted to effectuate opt-out ($\gamma I(x \neq D)$, where $I(x \neq D) = 1$ if $x \neq D$ and 0 otherwise):

$$u(x, m; x^*, \rho, \beta, \gamma) = \beta V(x, x^*, \rho) + m - \gamma I(x \neq D). \tag{1}$$

Notice that the function V depends not only on x, but also on a parameter $x^* \in X$, which we interpret as the contribution rate the worker regards as ideal, in the sense that $x = x^*$ uniquely maximizes $V(x, x^*,\rho)$. We also write V as a function of a parameter ρ that governs properties such as curvature. Another important feature of equation (1) is that we apply a weighting factor, β, to the utility derived from retirement contributions. 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.[2] We allow workers to differ with respect to x^*, ρ, β, and γ. We will write γ as the product of a relative opt-out cost parameter, η, that differs across workers, and a common scaling parameter, λ; thus, $\gamma = \lambda\eta$. This formulation allows us to hold the distribution of relative opt-out costs fixed while shrinking the average opt-out cost toward zero. To keep our notation as compact as possible, we will write the worker's characteristics, other than his ideal point, as $\theta = (\rho, \beta, \eta)$.

We assume that the effort cost of opting out, $\lambda\eta$, is independent of the option selected. Consistent with other work on this topic (Bernheim, Fradkin, and Popov (2015), Carroll et al. (2009), and Goldin and Reck (2019)), our analysis presupposes that opt-out costs reflect effort the worker must expend 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. We abstract from the interesting possibility that the worker must expend cognitive effort to understand his own preferences (the function $V(\cdot, x^*, \rho)$).

As explained in Bernheim, Fradkin, and Popov (2015), this formulation accommodates dynamics, in that we can interpret V as a reduced form representing the worker's perceived continuation value. Because the original default, D, affects the continuation value only

[2]With this formulation, the bias applies to V but not to m, which may be appropriate if, for example, β captures present bias and m is an immediate payment. Applying β to $V(x, x^*, \rho) + m$, rather than merely $V(x, x^*, \rho)$, would alter the formula for the optimal fine in Proposition 4, but would otherwise leave our results unchanged.

through the initial contribution, x, D does not appear as an argument of V. Accordingly, when optimizing the default D, we do not have to contemplate direct effects through V.[3]

For one of our extensions, we assume that, 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 hence chooses x to maximize $u(x, B - I(x = D)K; x^*, \rho, \beta, \lambda\eta)$. 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:

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

The worker therefore opts out of the default whenever

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

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

$$\begin{aligned} U_\lambda(D, x^*, \theta) =& \beta\left[(1 - C_\lambda(D, x^*, \theta))V(D, x^*, \rho) + C_\lambda(D, x^*, \theta)V(x^*, x^*, \rho)\right] \\ & - C_\lambda(D, x^*, \theta)\lambda\eta - (1 - C_\lambda(D, x^*, \theta))K + B, \end{aligned}$$

We assume that $\theta \in \left[\underline{\rho}, \overline{\rho}\right] \times \left[\underline{\beta}, \overline{\beta}\right] \times \left[\underline{\eta}, \overline{\eta}\right] \equiv \Theta$, where all of these bounds are finite, and where $\underline{\beta}, \underline{\eta} > 0$. Let $G(\theta)$ denote the CDF governing the marginal distribution of θ across

[3]For settings in which V is a state evaluation function for some dynamic process, it is worth emphasizing that our approximation involves taking the limit as the *current* opt-out cost approaches zero, holding *future* opt-out costs constant. This construction allows us to treat V as a fixed function as we change λ.

workers, and $F(x^* \mid \theta)$ denote the CDF governing the distribution of x^* conditional on θ.

We impose minimal restrictions on V:

Assumption 1. *For all $(x, x^*\rho) \in X^2 \times \left[\underline{\rho}, \overline{\rho}\right]$, (i) $V(x, x^*, \rho)$ is real-valued and continuous, and has continuous first through third derivatives with $V_{11}(x^*, x^*, \rho) < 0$, and (ii) $V_{12}(x, x^*, \rho) > 0$.*

Part (i) of Assumption 1 is a mild regularity condition. Because $x = x^*$ maximizes $V(x, x^*, \rho)$, we know that $V_{11}(x^*, x^*, \rho) \leq 0$, so the final portion simply rules out the possibility that V is "too flat" at any optimum. Part (ii) is a single-crossing requirement. This property is useful because it implies that the set of types who accept the default is an interval. However, the arguments we use to prove our results appear to rely on this implication only as an analytic convenience. We therefore suspect that an even less restrictive assumption would suffice.

We also impose the following restrictions on F and G:

Assumption 2. *F and G are atomless distributions with well-defined densities. The following properties hold for F: (i) (Full Support) there exists $f^{min} > 0$ such that for f, the density function of F, $f(x^* \mid \theta) > f^{min}$ holds for all $x^* \in X$, $\theta \in \Theta$; (ii) (Differentiability) F is twice continuously differentiable with respect to x^* and θ.*

For notational convenience, we write probabilities and expectations over x^* conditional on θ as $\Pr_{x^*|\theta}$ and $E_{x^*|\theta}$. Without loss of generality, we normalize the total population size to unity ($\int_\Theta dG(\theta) = 1$).

2.2 The employer

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

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

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, attributing the same value to a dollar regardless of who receives it. However, she may disagree with the workers concerning the assessment of their well-being at the moment they decide whether to opt out of the default.[4] We will assume that this disagreement is limited to the normatively appropriate value of β, which she takes to be unity. Thus, to

[4]If workers are time-inconsistent, they may agree with the employer's assessment of their opt-out decisions at other points in time.

the extent the worker's β diverges from unity, the employer is of the opinion that cognitive 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 their utility from retirement income which is delayed.[5] For other interpretations of β, see Bernheim, Fradkin, and Popov (2015) and Goldin and Reck (2019). A key feature of our framework is that the employer sees bias as pertaining to the opt-out decision, rather than to the choice of x conditional on opting out. In other words, she agrees that x^* is the worker's ideal choice.[6] 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 lie in the future. Thus, to the extent the employer believes workers are quasi-hyperbolic discounters and interprets β as "present bias," that bias would infect the opt-out decision, but not the worker's perceived continuation value (V) nor the chosen contribution rate, precisely as we assume.

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

$$\begin{aligned}\tilde{U}_\lambda(D, x^*, \theta) =& (1 - C_\lambda(D, x^*, \theta))V(D, x^*, \rho) + C_\lambda(D, x^*, \theta)V(x^*, x^*, \rho) \\ & - C_\lambda(D, x^*, \theta)\lambda\eta - (1 - C_\lambda(D, x^*, \theta))K + B.\end{aligned}$$

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

3 The approximate optimality of weighted opt-out minimization

In this section, we provide a general characterization of approximately optimal defaults that establishes a connection between welfare maximization and weighted opt-out minimization. Our main results show that, as $\lambda \to 0$, for appropriate weights, rescaled versions of the weighted opt-out frequency and of aggregate welfare both converge uniformly to the same function, and consequently the defaults that maximize those functions also converge. For the time being, we confine attention to settings in which the employer does not have the ability to impose fines for passive choice (i.e., we fix $K = B = 0$).

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

[6] More specifically, she does not take issue with V, which reflects the worker's understanding and assessment of future consequences.

3.1 Weighted opt-out minimization

Because the opt-out frequency for workers with characteristics θ, $\Pr_{x^*|\theta}\left[\Delta(D,x,\rho) \leq \frac{\lambda\eta}{\beta}\middle|\theta\right]$, converges to zero for all D and θ as $\lambda \to 0$, we study the limiting properties of the weighted-opt-out-minimizing defaults by progressively adjusting the scale of the objective function. For this purpose, we define

$$Q_\lambda(D,\theta) \equiv \frac{\Pr\left[\Delta(D,x,\rho) \leq \frac{\lambda\eta}{\beta}\middle|\theta\right]}{2\lambda^{\frac{1}{2}}}$$

For any weighting formula $\omega(\theta)$, the overall opt-out frequency, rescaled by $(2\lambda)^{-\frac{1}{2}}$, is then

$$\Omega_\lambda(D) = \int_\theta \omega(\theta) Q_\lambda(D,\theta) dG(\theta).$$

Our analysis focuses on a specific weighting formula:

$$\omega(\eta,\beta) = \eta\left(1 - \frac{1}{3\beta}\right).$$

Under our assumptions concerning continuity, bounds, and atomless distributions, it is easy to check that, fixing λ, the (rescaled) opt-in frequency, $\Omega_\lambda(D)$, varies continuously with D. Accordingly, because X is compact, there exists a (possibly non-unique) default option, $D_\Omega(\lambda)$, that maximizes weighted opt-in (and minimizes weighted opt-out).

To characterize the limiting case as $\lambda \to 0$, we define the approximate (rescaled) opt-out frequency for workers with characteristic θ,

$$Q(D,\theta) \equiv \left(\frac{\eta}{\beta}\right)^{\frac{1}{2}} f(D \mid \theta)\left(\frac{1}{-\frac{1}{2}V_{11}(D,D,\rho)}\right)^{\frac{1}{2}}.$$

To understand why we interpret this expression as the approximate opt-out frequency for workers with characteristics θ, notice that a worker's perceived net benefit to opting out is $-\frac{1}{2}V_{11}(D,D,\rho)\left(D - x^*\right)^2 - \frac{\lambda\eta}{\beta}$ to a second-order approximation. This expression implies that workers will opt in as long as they fall within an interval with an approximate length of $2\left(\frac{\lambda\eta}{\beta}\right)^{\frac{1}{2}}\left(\frac{1}{-\frac{1}{2}V_{11}(D,D,\rho)}\right)^{\frac{1}{2}}$. If this interval is small, then the density within it is roughly constant at $f(D \mid \theta)$. Consequently, the product of these two terms approximates the opt-in frequency. The function $Q(D,\theta)$ simply equals this product divided by a scaling factor, $2\lambda^{\frac{1}{2}}$.

To the extent $Q(D,\theta)$ approximates the (rescaled) weighted opt-out frequency for workers with characteristics θ, the following function approximates the overall (rescaled) weighted opt-out frequency:

$$\Omega(D) \equiv \int_\theta \eta\left(1 - \frac{1}{3\beta}\right) Q(D,\theta) dG(\theta).$$

Our analysis identifies a special role for the default rate D^* that maximizes the approximate rescaled (weighted) opt-in frequency:

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

As with $D_\Omega(\lambda)$, existence follows directly from our assumptions. Cases with multiple maxima are non-generic and therefore of little interest. To avoid some technical complications, we will therefore rule those cases out by assumption.

Assumption 3. *D^* is unique.*

It is worth emphasizing that, even when Θ is degenerate, D^* generally differs from the point of maximal density, except in special cases (e.g., when the curvature of V is the same at all ideal points).

Our first main result tells us that D^* approximates the opt-out minimizing default rate for small λ. The proof consists of establishing the intuitive property that the actual weighted opt-out frequency, divided by the scaling factor $2\gamma^{\frac{1}{2}}$, converges uniformly to $\Omega(D)$ as $\lambda \to 0$.

Proposition 1. *The weighted opt-out-minimizing default option $D_\Omega(\lambda)$ converges to D^* as $\lambda \to 0$.*

3.2 Welfare-maximization

Our second main result tells us that D^* also approximates the welfare-maximizing default rate for small λ.

Proposition 2. *The welfare-maximizing default option $D_L(\lambda)$ converges to D^* as $\lambda \to 0$.*

Combining Propositions 1 and 2, we reach our central conclusion: the difference between the weighted-opt-out-minimizing and welfare-maximizing default options vanishes as $\lambda \to 0$.

To build intuition for this result, note that the employer's problem – setting D to maximize $E\left[\tilde{U}(D, x^*, \theta)\right]$ – is equivalent to maximizing

$$L_\lambda(D) \equiv E\left[\tilde{U}_\lambda(D, x^*, \theta) - V(x^*, x^*, \rho)\right],$$

which we interpret as the (negative of) total welfare loss relative to the ideal retirement saving choice. For any given x^*, the term in brackets is either $-\lambda\eta$ (if the worker incurs the opt-out cost and selects his optimal contribution rate) or $V(D, x^*, \rho) - V(x^*, x^*, \rho) = -\Delta(D, x^*, \rho)$ (if he accepts the default). We can therefore rewrite the objective function as follows:

$$\begin{aligned} L_\lambda(D) \quad = \quad & -\int_\theta \lambda\eta dG(\theta) + \int_\theta \Pr\left(\Delta(D, x^*, \rho) \leq \tfrac{\lambda\eta}{\beta} \mid \theta\right) \\ & \times \left[\lambda\eta - E\left(\Delta(D, x^*, \rho) \mid \theta, \Delta(D, x^*, \rho) \leq \tfrac{\lambda\eta}{\beta}\right)\right] dG(\theta) \end{aligned} \tag{4}$$

Intuitively, inducing opt-out potentially creates a total welfare loss of $-\int_\theta \lambda\eta dG(\theta)$. Those who opt in mitigate this loss, but only by the difference between $\lambda\eta$ and the average loss associated with the utility difference between choosing x^* and choosing D.

Now let's think about maximizing $L_\lambda(D)$ over D for a fixed value of λ. Plainly, the $-\int_\theta \lambda\eta dG(\theta)$ term does not affect the argmax. The rest of the expression integrates the product of the conditional opt-in frequency and another term involving D. It is therefore not immediately obvious why maximization of the weighted opt-in frequency (or minimization of opt-out) should coincide with maximization of the entire expression. However, if it turned out that the term in brackets was independent of D, then the welfare maximization problem would be equivalent to weighted opt-out minimization. In point of fact, this property holds generally *in the limit* as $\lambda \to 0$.

Here we encounter a small technical complication: as $\lambda \to 0$, the function L_λ converges to the function L_0, which maps all default rates to the same value. That property renders the limiting optimization problem unenlightening. To learn about optimization with small λ from the limiting case, we have to translate and rescale L_λ so that the objective function neither collapses to a constant nor explodes to infinity as $\lambda \to 0$. We therefore define the following objective function:

$$W_\lambda(D) \equiv \frac{L_\lambda(D) + \int_\theta \lambda\eta dG(\theta)}{2\lambda^{\frac{3}{2}}}.$$

For any given λ, the maximizers of L_λ and W_λ obviously coincide. Moreover, when we use W_λ, the optimal defaults for the limiting case approximate the optimal defaults for small λ.

To visualize the limiting optimization problem, we can consider a second-order approximation in x^* for $\Delta(D, x^*, \rho)$. Recalling that $\Delta(x^*, x^*, \rho) = 0$ for all $x \in X$, we see that $\Delta(D, x^*, \rho)$ is (approximately) a parabola with a minimized value of 0 at $x^* = D$. Truncating that parabola at the boundaries of the opt-in interval and taking the density to be constant (to an approximation) over this interval, we see that the $E_{x^*|\theta}$ term is approximately the area beneath this truncated parabola divided by its width. It turns out that this ratio is always $\frac{\lambda\eta}{3\beta}$. Figure 1 illustrates the underlying mathematical principle. It shows the parabola $y = v\,(x^* - D)^2$ (where v is an arbitrary constant) for x^* in the interval $[D - t, D + t]$, which reaches a height of $h = vt^2$ at the interval's endpoints. A straightforward computation shows that the ratio of the shaded area B to the length of the interval A equals $\frac{h}{3}$, regardless of v. Returning to the objective function (equation (4)), and using $h = \frac{\lambda\eta}{\beta}$, we see that the bracketed term is approximately $\lambda\eta - \frac{\lambda\eta}{3\beta} = \lambda\eta\left(1 - \frac{1}{3\beta}\right)$ – in other words, a positive constant – regardless of the second-order coefficient, which may vary with D. It follows that, when λ is small, the welfare-maximization problem is approximately the same as maximizing the weighted opt-in frequency with weights $\omega(\eta, \beta) = \eta\left(1 - \frac{1}{3\beta}\right)$. The formal proof uses the fact that the conditional opt-out probability divided by the scaling factor $2\gamma^{\frac{1}{2}}$ converges uniformly to $Q(D, \theta)$ as $\lambda \to 0$ (as discussed above), as well as the fact

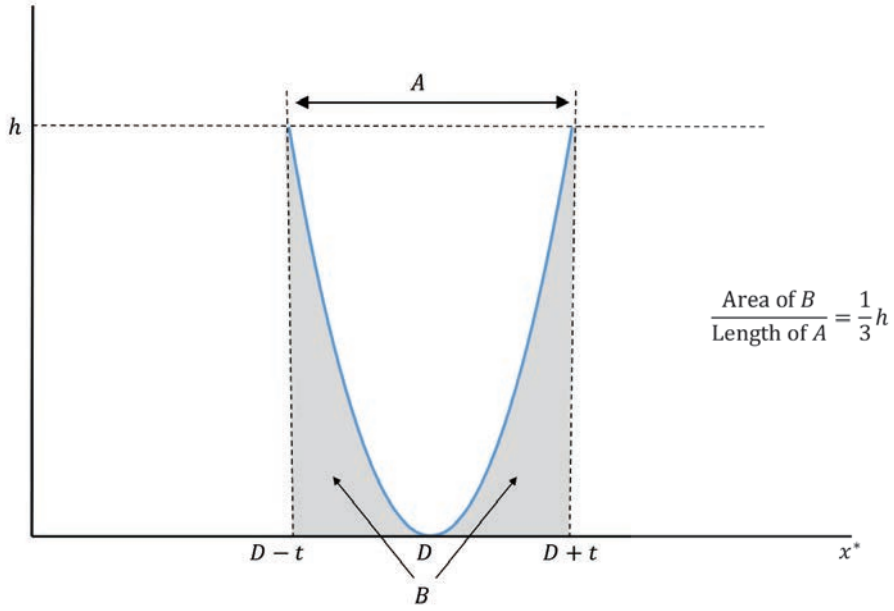

Figure 1: Second-order approximation to the conditional expectation

that the bracketed term divided by the scaling factor λ converges uniformly to $\eta\left(1-\frac{1}{3\beta}\right)$.

Notably, our general characterization applies regardless of whether the weights implied by the formula $\omega(\eta,\beta)=\eta\left(1-\frac{1}{3\beta}\right)$ are positive or negative. If $\underline{\beta}>\frac{1}{3}$, then all the weights are positive, which means the employer tries to achieve low opt-out frequencies. If $\overline{\beta}<\frac{1}{3}$, then all the weights are negative, which means the employer sets the default to achieve high opt-out frequencies. If $\underline{\beta}<\frac{1}{3}<\overline{\beta}$, then some weights are positive while others are negative, which means the employer tries to set the default to achieve low opt-out frequencies for some groups of workers and high opt out frequencies for others. Consequently, our analysis allows us to interpret the two polar cases identified in the literature – those for which opt-out minimization is optimal, and those for which opt-out maximization is optimal – as two sides of the same coin.

4 The approximate optimality of unweighted opt-out minimization

Given the focus of the previous literature, it is also important to investigate the circumstances under which *unweighted* opt-out minimization is approximately optimal. The unweighted opt-in frequency is given by $(2\lambda)^{-\frac{1}{2}}\,\Omega_\lambda^U(D)$, where

$$\Omega_\lambda^U(D)=\int_\theta Q_\lambda(D,\theta)dG(\theta).$$

Our analysis references both the opt-out minimizing default option, $D^U_\Omega(\lambda) \equiv \min_{D \in X} \Omega^U_\lambda(D)$, and the opt-out maximizing default option, $d^U_\Omega(\lambda) \equiv \max_{D \in X} \Omega^U_\lambda(D)$. As with the weighted-opt-out minimizing default option, $D_\Omega(\lambda)$, existence follows directly from our assumptions.

In light of Propositions 1 and 2, we know that unweighted opt-out minimization is asymptotically welfare-maximizing in settings where weighted and unweighted opt-out minimization coincide. A sufficient condition for this coincidence is that the weighted and unweighted opt-in frequencies are related by a fixed constant of proportionality. There are two potential routes to ensuring that this proportionality requirement holds. The first is to identify conditions under which the weight is the same for all workers. Unfortunately, this route does not lead to useful insights, because $\omega(\eta, \beta) = \eta\left(1 - \frac{1}{3\beta}\right)$ is constant only if there is no heterogeneity in η or β or if there is a fortuitous deterministic relationship between them (specifically, $\eta \propto \left(1 - \frac{1}{3\beta}\right)^{-1}$). The second route is to identify conditions under which the weighted opt-in frequency is separable into a component involving the characteristics that determine the weights and a component involving the default D. Unfortunately, $\Pr_{x^*|\theta}\left[\Delta(D, x, \rho) \leq \frac{\lambda\eta}{\beta}\right]$ does not generally factor in this way.

The second route becomes more promising when we focus on the limiting case. Recall that the (rescaled) weighted opt-in frequency, $\Omega_\lambda(D)$, converges uniformly to $\Omega(D)$, which we can write as follows:

$$\Omega(D) = \int_\theta \eta\left(1 - \frac{1}{3\beta}\right) Q(D, \theta) dG(\theta)$$

$$= \int_\theta \left[\eta\left(1 - \frac{1}{3\beta}\right)\left(\frac{\eta}{\beta}\right)^{\frac{1}{2}}\right]\left[f(D \mid \theta)\left(\frac{1}{-\frac{1}{2}V_{11}(D, D, \rho)}\right)^{\frac{1}{2}}\right] dG(\theta)$$

We have divided the integrand into two bracketed components. The first depends only on the parameters (η and β) that determine the weight, while the second depends on D. We do not yet have the desired separability property, however, because $f(D \mid \theta)$ may depend on η and β. Additionally, η and β may be stochastically related to ρ, which appears in the V_{11} term. However, both of these potential dependencies disappear if x^* and ρ are distributed independently of η and β. In that case, using $h(\eta, \beta)$ to denote the density of the marginal distribution of η and β, and using $k(\rho)$ to denote the density for the marginal distribution of ρ, we can write:

$$\Omega(D) = \Phi \times \int_\rho f(D \mid \rho)\left(\frac{1}{-\frac{1}{2}V_{11}(D, D, \rho)}\right)^{\frac{1}{2}} k(\rho) d\rho$$

where

$$\Phi = \int_{\eta,\beta} \eta\left(1 - \frac{1}{3\beta}\right)\left(\frac{\eta}{\beta}\right)^{\frac{1}{2}} h(\eta, \beta) d\eta d\beta. \tag{5}$$

To demonstrate formally the asymptotic equivalence of weighted and unweighted opt-out minimization under the stated independence assumption, we need to provide a characterization of the limiting unweighted opt-in frequency function. Using essentially the same arguments as in the proof of Proposition 1, one can show that $\Omega_\lambda^U(D)$ converges uniformly to

$$\Omega^U(D) = \int_\theta Q(D,\theta) dG(\theta)$$

A calculation analogous to the one provided above for $\Omega(D)$ then implies

$$\Omega^U(D) = \left[\int_{\eta,\beta} \left(\frac{\eta}{\beta}\right)^{\frac{1}{2}} h(\eta,\beta) d\eta d\beta\right] \int_\rho f(D \mid \rho) \left(\frac{1}{-\frac{1}{2}V_{11}(D,D,\rho)}\right)^{\frac{1}{2}} dk(\rho)$$

Accordingly, we have

$$\pi \times \Omega^U(D) = \Omega(D), \tag{6}$$

where

$$\pi = \Phi \left[\int_{\eta,\beta} \left(\frac{\eta}{\beta}\right)^{\frac{1}{2}} h(\eta,\beta) d\eta d\beta\right]^{-1}.$$

Thus, the asymptotic weighted and unweighted opt-in frequencies are related by a positive fixed factor of proportionality when $\Phi > 0$, and by a negative fixed factor of proportionality when $\Phi < 0$. Applying Propositions 1 and 2, we see that unweighted opt-out minimization is asymptotically welfare-optimal when $\Phi > 0$. Alternatively, when $\Phi < 0$, weighted opt-out minimization involves negative weights for some or all workers, and as a result coincides with unweighted opt-out *maximization*. In that case, unweighted opt-out maximization is asymptotically welfare-optimal. The following proposition summarizes these observations:

Proposition 3. *Assume x^* and ρ are distributed independently of η and β. If $\Phi > 0$, then the unweighted opt-out-minimizing default option $D_P^U(\lambda)$ converges to D^* as $\lambda \to 0$. If $\Phi < 0$, then the unweighted opt-out-maximizing default option $d_P^U(\lambda)$ converges to D^* as $\lambda \to 0$.*

Goldin and Reck (2020) also present a result on the optimality of opt-out minimization (their Proposition 4). The environments they consider are more restricted, in that worker heterogeneity is limited to x^* and η. One of their sufficient conditions requires stochastic independence of those two characteristics. Our result reveals that, when worker heterogeneity extends to β and ρ as well as x^* and η, unweighted opt-out minimization may be asymptotically suboptimal even when Goldin and Reck's independence requirement holds. We arrive at a more general condition that requires stochastic independence between x^* and ρ on the one hand, and η and β on the other.

Applying Proposition 3 to settings with homogeneous β and ρ, we see that our focus on approximate optimality allows us to establish the desirability of unweighted opt-out minimization and maximization under conditions that are considerably more general than

Goldin and Reck's. In particular, our result dispenses with a collection of highly specialized assumptions (specifically, that the utility derived from the action is a convex, single-peaked, symmetric function that depends only on difference between the action and the worker's ideal point, and that the distribution of ideal points is single-peaked and symmetric), which we replace with some relatively weak regularity requirements along with a single-crossing property.

Notice that the integrand in the definition of Φ is positive when $\beta > \frac{1}{3}$. Thus, consistent with Goldin and Reck's (2020) Proposition 4, our Proposition 3 indicates that opt-out minimization is optimal when opt-out costs are sufficiently normative; otherwise, opt-out maximization is optimal. However, to the extent the employer can impose non-disappative fines for passive choice, opt-out minimization becomes relatively more attractive.

To understand the preceding claim, we introduce a fine $K(\lambda)$, which we allow to shrink with the scale of opt-out costs. The opt-in condition becomes

$$\Delta(D, x^*, \rho) \leq \frac{\lambda\eta - K(\lambda)}{\beta}$$

Visualizing $E\left(\Delta(D, x^*, \rho) \mid \theta, \Delta(D, x^*, \rho) \leq \frac{\lambda\eta - K(\lambda)}{\beta}\right)$ as the area beneath a truncated parabola divided by the parabola's width (and recalling that $\Delta(D, x^*, \rho) \geq 0$), we see that it equals $\max\left\{0, \frac{\lambda\eta - K(\lambda)}{3\beta}\right\}$ to a second-order approximation. It follows that

$$\frac{1}{\lambda}\left[\lambda\eta - E\left(\Delta(D, x^*, \rho) \mid \theta, \Delta(D, x^*, \rho) < \frac{\lambda\eta - K(\lambda)}{\beta}\right)\right]$$

converges uniformly to $\eta - \max\left\{0, \frac{\eta - \kappa}{3\beta}\right\}$, where $\kappa \equiv \lim_{\lambda \to 0} \frac{K(\lambda)}{\lambda}$. Accordingly, for fixed fines $K(\lambda)$, welfare maximization coincides asymptotically with weighted opt-out minimization using weights given by the formula

$$\omega^{\kappa}(\eta, \beta) \equiv \eta - \max\left\{0, \frac{\eta - \kappa}{3\beta}\right\}.$$

Now notice that, for any given η and β, the weight is increasing in κ. Indeed, if $\kappa > \overline{\eta}(1 - 3\underline{\beta})$, then all of the weights are positive, which implies $\Phi > 0$, and hence that unweighted opt-out minimization is asymptotically optimal when x^* and ρ are distributed independently of η and β.

To illustrate why it may be optimal for the employer to set a positive fine for passive choice, we specialize to settings in which worker heterogeneity is confined to x^*. The following proposition characterizes the optimal fine and, by implication, the optimal bonus for any fixed default option.

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

The intuition for Proposition 4 is that, by establishing a fine for passive choice equal to the portion of active-choice 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 induced objective function coincides with one for a setting in which $\beta = 1$. Consequently, solving for the optimal default with arbitrary β conditional on the optimal fine is mathematically equivalent to solving for the optimal default with $\beta = 1$ and no fine. The optimality of unweighted opt-out minimization then follows immediately under the independence condition stated in Proposition 3, regardless of whether as-if opt-out costs are "sufficiently normative."

As a final observation, we note that if x^* and ρ are distributed independently of η and β, then asymptotically the unweighted opt-out minimizing and maximizing defaults do not depend on the size of the fine. With the introduction of a fine, the asymptotic unweighted opt-in frequency becomes

$$\Omega^U(D) = \left[\int_{\eta,\beta} \left(\max\left\{0, \frac{\eta-\kappa}{\beta}\right\}\right)^{\frac{1}{2}} h(\eta,\beta)d\eta d\beta\right] \int_{\rho} f(D \mid \rho) \left(\frac{1}{-\frac{1}{2}V_{11}(D,D,\rho)}\right)^{\frac{1}{2}} dk(\rho)$$

As long as $\kappa < \overline{\eta}$, the bracketed term is a positive constant, which means that the same default D maximizes this expression regardless of κ. This property is logistically convenient, because it implies that the employer can optimize the default by minimizing (or alternatively maximizing) the opt-out frequency based on data from a regime in which it imposed no fine. Because the size of the fine can determine whether minimization or maximization is appropriate, the employer must still optimize the default and the fine simultaneously, but the problem reduces to consideration of just two fine-invariant default alternatives.

5 Accommodating normative ambiguity

Depending on which psychological mechanisms β purportedly captures, there may be controversy as to whether it constitutes a bias. Imagine, for example, that β parametrizes time-inconsistency. Some studies advocate evaluating welfare based solely on forward-looking choices, on the grounds that people suffer from "self-control problems" when making decisions contemporaneously (see, e.g., O'Donoghue and Rabin 1999). However, this language may reflect normative preconceptions rather than objective inferences. If people fully appreciate experiences only in the moment and overintellectualize at arms length, their in-the-moment choices, rather than the forward-looking choices, would be the ones that merit deference. Absent an objective basis for adjudicating between these perspectives, there is an argument for remaining agnostic and respecting both.

To accommodate normative ambiguity, Bernheim, Fradkin, and Popov (2015) deployed the welfare framework developed in Bernheim and Rangel (2009) and elaborated in Bern-

heim (2009; 2016; forthcoming) and Bernheim and Taubinsky (2018). Within that paradigm, one can evaluate a change from policy p to p' by computing two versions of equivalent variation: EV_A, which is the smallest increment to income with p (that is, the smallest increase or the largest reduction) such that the bundle obtained with p is unambuously chosen over the bundle obtained with p' (i.e., the individual would choose the first bundle over the second in all decision frames), and EV_B, which is the largest increment to income with p (that is, the largest increase or smallest reduction) such that the bundle obtained with p' is unambiguously chosen over the bundle obtained with p. Despite the ambiguities implied by inconsistent choices, one can say that the change is unambiguously worth at least EV_A and no more than EV_B. Bernheim, Fradkin, and Popov (2015) provided a formal justification for aggregating these welfare measures over populations of decision makers. For default-setting problems with sophisticated present focus, they also showed that one calculates EV_B by treating β as a bias, as above. To determine EV_A, one instead evaluates welfare according to a slightly modified objective function that respects β:

$$\begin{aligned}\tilde{U}_\lambda(D,x^*,\theta) =&(1-C_\lambda(D,x^*,\theta))\beta V(D,x^*,\rho)+C_\lambda(D,x^*,\theta)\beta V(x^*,x^*,\rho)\\ &-C_\lambda(D,x^*,\theta)\lambda\eta-(1-C_\lambda(D,x^*,\theta))K+B.\end{aligned}$$

Surprisingly, for empirically parametrized opt-out models, Bernheim, Fradkin, and Popov (2015) find that the same default option maximizes both EV_A and EV_B.

Our analysis provides insight into the mechanisms that drive this conclusion, and also allows us to state precisely the conditions under which it holds (asymptotically). For EV_A, the welfare loss function becomes

$$\begin{aligned}L_\lambda(D) \;\;=\;\; &-\textstyle\int_\theta \lambda\eta dG(\theta)+\int_\theta \Pr\left(\Delta(D,x^*,\rho)\leq\frac{\lambda\eta}{\beta}\mid\theta\right)\\ &\times\left[\lambda\eta-\beta E\left(\Delta(D,x^*,\rho)\mid\theta,\Delta(D,x^*,\rho)\leq\tfrac{\lambda\eta}{\beta}\right)\right]dG(\theta)\end{aligned}$$

Replicating our earlier reasoning, we see that maximization of this objective function is asympotically equivalent to minimizing weighted opt-out, using weights $\omega(\eta,\beta)=\eta\left(1-\frac{1}{3}\right)$, rather than $\omega(\eta,\beta)=\eta\left(1-\frac{1}{3\beta}\right)$.

From these observations, it follows that maximization of EV_A coincides (asymptotically) with maximization of EV_B, and consequently that the optimal default is robust with respect to strategic ambiguity, as long as the same default achieves weighted opt-out minimization with weights $\eta\left(1-\frac{1}{3\beta}\right)$ and $\eta\left(1-\frac{1}{3}\right)$. This condition is obviously satisfied when there is no heterogeneity in β (the case considered in Bernheim, Fradkin, and Popov 2015), provided $\beta>\frac{1}{3}$. More generally, the result survives the introduction of heterogeneity with respect to present focus as long as the distribution of β is independent of x^*, ρ, and η. Notice that the latter condition does not require that η is independent of x^* and ρ, which means that weighted opt-out minimization may be normatively robust even when it diverges from

unweighted opt-out minimization.

6 Extensions

Even though Assumptions 1 and 2 are relatively mild in a technical sense, they do not fit all applications. One issue is that, in some practical settings, the distribution of ideal options includes atoms. For example, in the context of 401(k)s, the possibility of exhausting an employer's matching contributions creates a kink point in the worker's opportunity set, which can lead to bunching at the kink point. One also typically observes bunching at zero, the lower boundary. A second practical consideration is that the choice set is sometimes finite. In this section, we consider these cases as extensions.

6.1 Bunching

We take the view that bunching usually results from a characteristic of the opportunity set, such as a kink or a boundary, rather than from atoms in the underlying distribution of workers' characteristics. Accordingly, we model the ideal point, $x^*(y)$, as depending on some latent characteristic, $y \in Y$, where Y is a compact set. We assume there is a finite set of disjoint non-degenerate intervals, $Y_1, ..., Y_N$, where $Y_n = \left[\underline{y}_n, \overline{y}_n\right]$, along with a set of associated contribution levels, $Z = \{z_1, ..., z_N\} \subset X$, such that all values of $y \in Y_n$ map to the same value, $x^*(y) = z_n$. For example, y might represent the worker's long-run discount factor, and x^* might be choices from a non-linear budget set, in which case the z_n correspond to kink points in an opportunity set, induced for example by a cap on employer matching contributions, or alternatively z_n might be a boundary point of X. We will assume that, outside $Y_0 \equiv \cup_{n=1}^{N} Y_n$, $x^*(y)$ is strictly increasing and differentiable with a derivative that is uniformly bounded away from 0.

Consistent with these modifications, we now model continuation utility, $V(x, y, \rho)$, as depending on the latent characteristic y rather than the ideal point $x^*(y)$, which is presumably specific to the opportunity set. Here it is important to avoid the assumption of differentiability, precisely because an underlying kink in an opportunity set generally translates into a point of non-differentiability, which in turn produces the bunching assumed above. Accordingly, we make the following weak assumption concerning V:

Assumption 4. *For all $(x, y, \rho) \in X \times Y \times \left[\underline{\rho}, \overline{\rho}\right]$, $V(x, y, \rho)$ is real-valued, continuous, and uniquely maximized at $x = x^*(y)$.*

Notice that Assumption 4 dispenses not only with our differentiability assumptions, but also with the single-crossing property.

To conserve on new notation, we will use F to represent the distribution of y rather than x^*. We can also make due with weaker assumptions concerning F:

Assumption 5. *F and G are atomless distributions with well-defined densities. There exists $f^{max} > 0$ such that for f, the density function of F, $f(y \mid \theta) < f^{max}$ holds for all $y \in Y$, $\theta \in \Theta$.*

While Assumption 2 did not explicitly call out the existence of an upper bound on $f(x \mid \theta)$, that property followed from the assumed continuity of the density as well as the compactness of the sets X and Θ. Thus, Assumption 5 is unambiguously weaker than Assumption 2.

Under these assumptions, the fraction of the population with an ideal point of $z_n \in Z$ is

$$\pi_n \equiv \int_\theta \Pr\left(x^*(y) = z_n \mid \theta\right) dG(\theta) = \int_\theta \Big[F(\overline{y}_n \mid \theta) - F(\underline{y}_n \mid \theta)\Big]\, dG(\theta). \tag{7}$$

Were we to assume full support (as in Section 2), we would have $\pi_n > 0$. Here we will assume only that there is some $z_n \in Z$ with $\pi_n > 0$, which implies the existence of bunching. Notice that the analog of expression (7) is 0 for any $D \notin Z$.

Our analysis will focus on weighted opt-out minimization with weights $\omega(\eta) = \eta$. The weighted opt-in frequency is then

$$\hat{\Omega}_\lambda(D) \equiv \int_\theta \eta \Pr\left(\Delta(D, x^*(y), \rho) \leq \frac{\lambda\eta}{\beta} \middle| \theta\right) dG(\theta)$$

Let $D_{\hat{\Omega}}(\lambda)$ denote any default that maximizes this objective function.

For any $D \notin Z$, it is easily verified that $\hat{\Omega}_\lambda(D) \to 0$ as $\lambda \to 0$.[7] Consequently, it is natural to conjecture that, as $\lambda \to 0$, the weighted opt-out minimizing default $D_{\hat{\Omega}}(\lambda)$ converges to $z^* \in Z$, defined as $\arg\max_{z \in Z} \hat{\Omega}(z)$, where (for $z \in Z$),

$$\hat{\Omega}(z) \equiv \int_\theta \eta \Pr\left(x^*(y) = z \mid \theta\right) dG(\theta).$$

We will assume that z^* is unique within Z, a property that holds generically.

Now we turn to welfare maximization. Equation (4), which defines the aggregate welfare function $L_\lambda(D)$, is unchanged, except that we replace x^* with $x^*(y)$. Following the structure of the arguments in Section 3.1, we define

$$\hat{W}_\lambda(D) \equiv \frac{L_\lambda(D) + \int_\theta \lambda\eta dG(\theta)}{\lambda}.$$

Notice that we use a different scaling factor here, λ^{-1} rather than $= (2\lambda)^{-\frac{3}{2}}$, to ensure that the objective function neither explodes to infinity nor collapses everywhere to zero. The reason for the change in scaling is that, here, some probabilities do not converge to zero. Let $D_{\hat{W}}(\lambda)$ be any welfare-maximizing default, given λ. Our objective is to characterize the limiting behavior of $D_{\hat{W}}(\lambda)$ as $\lambda \to 0$.

[7] The convergence is not necessarily uniform, however, since there are values of $D \notin Z$ that are arbitrarily close to points in Z.

It is useful to rewrite the welfare function as follows:

$$\begin{aligned}\hat{W}_\lambda(D) &= \int_\theta \Pr\left(\Delta(D,x^*(y),\rho) \leq \tfrac{\lambda\eta}{\beta}\middle|\,\theta\right) \\ &\qquad \times\left[\eta - \tfrac{1}{\lambda}E\left(\Delta(D,x^*(y),\rho) \mid \Delta(D,x^*(y),\rho) \leq \tfrac{\lambda\eta}{\beta}\middle|\,\theta\right)\right] dG(\theta), \\ &= \hat{\Omega}_\lambda(D) - \int_\theta \Pr\left(\Delta(D,x^*(y),\rho) \leq \tfrac{\lambda\eta}{\beta}\middle|\,\theta\right) \\ &\qquad \times\left[\tfrac{1}{\lambda}E\left(\Delta(D,x^*(y),\rho) \mid \Delta(D,x^*(y),\rho) \leq \tfrac{\lambda\eta}{\beta}\middle|\,\theta\right)\right] dG(\theta)\end{aligned} \tag{8}$$

Now think about what happens to $\hat{W}_\lambda(D)$ as $\lambda \to 0$. For $D \notin Z$, $\Pr\left(\Delta(D,x^*(y),\rho) \leq \tfrac{\lambda\eta}{\beta}\middle|\,\theta\right) \to 0$, so the second term (specifically, everything after $\hat{\Omega}_\lambda(D)$) vanishes.[8] In contrast, for $D \in Z$, $\Pr\left(\Delta(D,x^*(y),\rho) \leq \tfrac{\lambda\eta}{\beta}\middle|\,\theta\right)$ need not vanish. The limiting behavior of the second term then depends on the bracketed expression in the last line. In Section 3.2, we showed that, with no bunching, that expression converges to $\frac{\lambda\eta}{3\beta}$. In the current context, it converges to zero. Intuitively, for such $D = z_n \in Z$, as $\lambda \to 0$, the fraction of workers choosing z_n for whom $x^*(y) \neq z_n$ converges to zero. The conditional expectation is therefore governed entirely by workers for whom $x^*(y) = z_n$. But for those workers, $\Delta(z_n, x^*(y), \rho) = 0$. It is therefore intuitive that $\hat{W}_\lambda(D) - \hat{\Omega}_\lambda(D)$ converges to 0, and consequently that $D_{\hat{W}}(\lambda)$ also converges to D^*.

By articulating this intuition while attending to a number of technical issues, we prove the following result:

Proposition 5. *The weighted opt-out-minimizing default option $D_{\hat{\Omega}}(\lambda)$ and the welfare-maximizing default option $D_{\hat{W}}(\lambda)$ both converge to z^* as $\lambda \to 0$.*

Because the applicable weight is simply η, the asymptotically welfare-maximizing default option does not depend on the distribution of the bias parameter, β. In this context, because $\Pr\left(\Delta(D,x^*(y),\rho) \leq \tfrac{\lambda\eta}{\beta}\middle|\,\theta\right)$ converges to $\Pr\left(\Delta(D,x^*(y),\rho) = 0|\,\theta\right)$, bias can only enter in the limit through the bracketed term in the last line of equation (8). But as we have explained, that term disappears in the limit. Several implications follow.

First, because the weight is simply the relative opt-out cost η, which is always positive, we see that the welfare-maximizing strategy involves minimizing opt-out rather maximizing it, even when the bias is severe. Consequently, in settings with bunching, we can dispense entirely with the Goldin-Reck assumption that as-if opt-out costs are sufficiently normative, at least asymptotically.

Second, in this context, the asymptotic optimality of *unweighted* opt-out minimization only requires the independence of y (which stands in for x^*) and η. To understand this assertion, notice that we can rewrite $\hat{\Omega}(z)$ as follows:

[8] A technicality here is that it does not vanish uniformly.

$$\hat{\Omega}(z) = \int_{\eta} \eta \left[\int_{\rho} \int_{\beta} \Pr\left(x^*(y) = z \mid \beta, \rho\right) h(\beta, \rho \mid \eta) d\beta d\rho \right] k(\eta) d\eta$$

$$= \int_{\eta} \eta \Pr\left(x^*(y) = z \mid \eta\right) k(\eta) d\eta,$$

where $k(\eta)$ is the density for the marginal distribution of η, while $h(\beta, \rho \mid \eta)$ is the density for the joint distribution of β and ρ, conditional on η. If we assume y and η are independent, then the probability term factors out, which means we are left with $\hat{\Omega}(z) = \Pr\left(x^*(y) = z\right)$, the (limiting) unweighted opt-out frequency. Relative to the corresponding result in Section 4, we are able to dispense with all of the independence assumptions concerning β and ρ.

A comparison between Propositions 2 and 5 reveals an apparent tension: with no bunching, β appears in the weighting formula, but with even the tiniest amount of bunching, it vanishes. This tension is resolved by the observation that the convergence of the bracketed term to zero becomes slower and slower as the amount of bunching shrinks. When bunching is barely detectable, this term resembles $\eta\left(1 - \frac{1}{3\beta}\right)$ rather than η until λ is small enough to cause the probability atom at the kink point to dominate the cumulative density within any opt-in window containing the kink point.

An additional implication follows from the fact that β does not appear in the opt-out frequency weights for settings with bunching: weighted opt-out minimization is normatively robust (in the sense that the asymptotic maximizers of EV_A and EV_B coincide) even when β is heterogeneous and correlated with x^*, ρ, and η.

6.2 Finite menus

To analyze environments with finite sets of alternatives, we modify the model of Section 2. For simplicity, we assume the action x takes on one of two values, 0 or 1. Our analysis extends to settings with more than two discrete options in an obvious but tedious way, and this simplification allows us to illustrate the applicable principles while avoiding uninstructive notational complexity. The problem of setting default options for choices with binary alternatives is also of independent practical interest because it regularly arises in practice, for example with respect to organ donation elections.

As before, we assume we can write continuation utility, $V(x, x^*, \rho)$, as a function of the action x, a characteristic x^* governing the individual's preferred option, and a characteristic ρ governing the intensity of that preference. Here, however, x and x^* belong to different sets ($\{0, 1\}$ and X, respectively), so we reinterpret x^* as a latent characteristic rather than an ideal point. The incremental continuation utility the individual derives from action 1 relative to action 2 is then

$$C(x^*, \rho) = V(1, x^*, \rho) - V(0, x^*, \rho)$$

To the assumptions listed in the previous subsection (V real-valued and continuous), we add that $C(x^*, \rho)$ is strictly increasing in x^*. This assumption is simply a matter of arranging latent types in order of increasing preference for option 1. We also assume that $C(\underline{x}, \rho) < 0$ and $C(\overline{x}, \rho) > 0$, so that some people strictly prefer each option. These assumptions plainly imply the existence of some threshold value x_T such that $C(x_T, \rho) = 0$.

Next we define

$$\Delta(D, x^*, \rho) = \max\left\{0, (-1)^{1-D} C(x^*, \rho)\right\}$$

In other words, when $D = 0$, $\Delta(D, x^*, \rho)$ equals $C(x^*, \rho)$ truncated below at zero, while if $D = 1$, it equals $-C(x^*, \rho)$ truncated below at zero. This function has precisely the same interpretation as in previous sections: it measures the difference between the utility the individual derives from receiving his most preferred option, and the utility he derives from receiving another specified alternative (which may or may not be his most preferred option). It follows that the individual opts out of the default when $\Delta(D, x^*, \rho) > \frac{\lambda\eta}{\beta}$, exactly as before.

As in the last subsection, our analysis will focus on weighted opt-out minimization with weights $\omega(\eta) = \eta$. The weighted opt-out frequency is then

$$\tilde{\Omega}_\lambda(D) \equiv \int_\theta \eta \Pr\left(\Delta(D, x^*, \rho) \leq \frac{\lambda\eta}{\beta} \middle| \theta\right) dG(\theta).$$

Let $D_{\tilde{\Omega}}(\lambda)$ denote any default that maximizes this objective function. It is easy to see that, as $\lambda \to 0$, $\tilde{\Omega}_\lambda(D)$ converges to

$$\tilde{\Omega}(D) \equiv \int_\theta \eta \Pr\left((-1)^D (x^* - x_T) \leq 0 \mid \theta\right) dG(\theta)$$

Let D^* be the default that maximizes $\tilde{\Omega}(D)$. It is straightforward to establish the existence of some $\lambda_{\tilde{\Omega}} > 0$ such that, for $\lambda < \lambda_{\tilde{\Omega}}$, we have $D_{\tilde{\Omega}}(\lambda) = D^*$.

Even though we have altered our original model, equation (4) for $L_\lambda(D)$ continues to describe aggregate welfare. In parallel with the preceding subsection, we define

$$\tilde{W}_\lambda(D) \equiv \frac{L_\lambda(D) + \int_\theta \lambda\eta dG(\theta)}{\lambda},$$

which we can rewrite as

$$\begin{aligned} \tilde{W}_\lambda(D) &= \tilde{\Omega}_\lambda(D) - \int_\theta \Pr\left(\Delta(D, x^*, \rho) \leq \tfrac{\lambda\eta}{\beta} \middle| \theta\right) \\ &\quad \times \left[\tfrac{1}{\lambda} E\left(\Delta(D, x^*, \rho) \mid \Delta(D, x^*, \rho) \leq \tfrac{\lambda\eta}{\beta} \middle| \theta\right)\right] dG(\theta). \end{aligned}$$

As in the last subsection, we claim that the bracketed term converges to zero. Intu-

itively, for either value of D, as $\lambda \to 0$, the fraction of workers choosing z_n for whom $(-1)^D (x^* - x_T) > 0$, and hence for whom $\Delta(D, x^*, \rho) > 0$, converges to zero. The conditional expectation is therefore governed entirely by workers for whom $(-1)^D (x^* - x_T) \leq 0$. Because $\Delta(D, x^*, \rho) = 0$ for those workers, the bracketed term converges to zero. It follows that $\tilde{W}_\lambda(D) - \tilde{\Omega}_\lambda(D)$ converges to zero, which in turn means that $\tilde{W}_\lambda(D) - \tilde{\Omega}(D)$ converges to zero. An immediate implication is that there is some $\lambda_{\tilde{W}} > 0$ such that, for $\lambda < \lambda_{\tilde{W}}$, we have $D_{\tilde{W}}(\lambda) = D^*$. Thus, the weighted opt-out minimizing default with weights $\omega(\eta) = \eta$ is welfare-optimal for sufficiently small λ. While the preceding discussion omits some details, they are easy to fill in, and indeed they involve simpler versions of the arguments used in the proof of Proposition 5. Because the weights are the same as for settings with bunching, the same conclusions follow.

As in the previous section, there appears some tension between Proposition 2 and the conclusions we have just reached: with a continuous menu, β appears in the weighting formula, but with any finite menu, no matter how fine, it vanishes. This tension is resolved by the observation that the convergence of the bracketed term to zero becomes slower and slower as the cardinality of the menu grows. With an astronomical but finite number of alternatives, this term resembles $\eta \left(1 - \frac{1}{3\beta}\right)$ rather than η until λ is small enough to exclude all but a few alternatives from the opt-in window.

7 Numerical simulations

In this section, we illustrate our main convergence result by simulating welfare-maximizing, weighted opt-out minimizing, and unweighted opt-out minimizing defaults in settings that violate specific assumptions imposed in Carroll et al. (2009) and Goldin and Reck (2019). These simulations also show that our limiting result provides a decent approximation for settings with larger opt-out costs and, consequently, meaningful social stakes. We also investigate the magnitude of the inefficiencies resulting from minimizing the unweighted opt-out frequency, rather than the weighted opt-out frequency, in settings with correlations between x^*, η, and β.

7.1 Parametrizations

Table 1 summarizes the various parametric specifications used in our main simulations. For V, we employ a quadratic utility function, which exhibits the symmetry property imposed in the prior literature, and an asymmetric linear-exponential utility function (Martinez-Mora and Puy (2012)). For F, we examine a truncated Normal distribution that exhibits the symmetry and single peakedness properties imposed in the prior literature, a highly asymmetric distribution with a unique mode at a boundary value, and an asymmetric bimodal distribution. In all simulations, the support of the ideal-point distribution is the interval $[0, 5]$. Figure 2 depicts these alternatives.

When introducing heterogeneity with respect to the bias and opt-out cost parameters, for the sake of analytic tractability we assume $\beta_i \in \{0.5, 0.8, 1\}$ and $\eta_i \in \{0.5, 1, 2\}$. In some settings we assume that β_i, η_i, and x_i^* are distributed independently, and in others we allow for mild 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. Our specific distributional assumptions appear in Table 1.

7.2 Simulation results

Table 2 summarizes our main simulation results by comparing welfare-maximization to weighted opt-out minimization.[9] 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 λ. 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%.[10] Converting values of λ into their implied opt-out frequencies renders the size of the parameter more easily interpretable.[11]

For each specification and opt-out frequency, the table reports the distance between the welfare-maximizing default option $D_L(\lambda)$ and the weighted opt-out-minimizing default option $D_\Omega(\lambda)$, as well as the fraction of the potential welfare gain, $\Delta_L(\lambda)$, achieved by the opt-out-minimizing default option relative to a zero-default policy. Both of these metrics require explanation. For each simulation, we first find the default $D_L(\lambda)$ that maximizes welfare; to obtain $D_\Omega(\lambda)$, we then minimize weighted opt-out for the same γ. The table reports the absolute value of the difference between these two defaults, i.e., $|D_L(\lambda) - D_\Omega(\lambda)|$. To compute $\Delta_L(\lambda)$, we first evaluate the welfare gain achieved by the welfare-optimal policy relative to a baseline scenario in which the default is $D = 0$: $L_\lambda\left(D_L(\lambda)\right) - L_\lambda(0)$. Next we calculate the welfare gain achieved by the weighted opt-out minimizing policy relative to the same baseline: $L_\lambda\left(D_\Omega(\lambda)\right) - L_\lambda(0)$. We then define $\Delta_L(\lambda)$ as the ratio of the second welfare gain to the first, expressed as a percentage: $\Delta_L(\lambda) = 100\% \frac{L_\lambda(D_\Omega(\lambda)) - L_\lambda(0)}{L_\lambda(D_L(\lambda)) - L_\lambda(0)}$.

[9] 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 employed a maximal function value tolerance of $1e-11$ and maximal absolute quadrature error of $1e-12$.

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

[11] 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**	**Var.**	**Max.**	**Corr**(x, η)
Truncated Normal $f(x) = H * \phi\,(x - 2.5)$	2.5	2.5	≈ 0.911	2.5	-0.1531
Right-peaked $f(x) = H * x$	$3.\overline{3}$	≈ 3.538	≈ 1.389	5	-0.1608
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	$\{2, 3\}$	-0.0942

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. Var. displays the variance and Max. lists the (local) maximand(s) of the distribution. "Corr(x, η)" refers to the correlation between x, the ideal point, and η, the cost parameter, in the case of interdependence, as detailed in Table c).

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

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 2: Utility Functions and Distribution Functions for Numerical Simulations: Illustrations

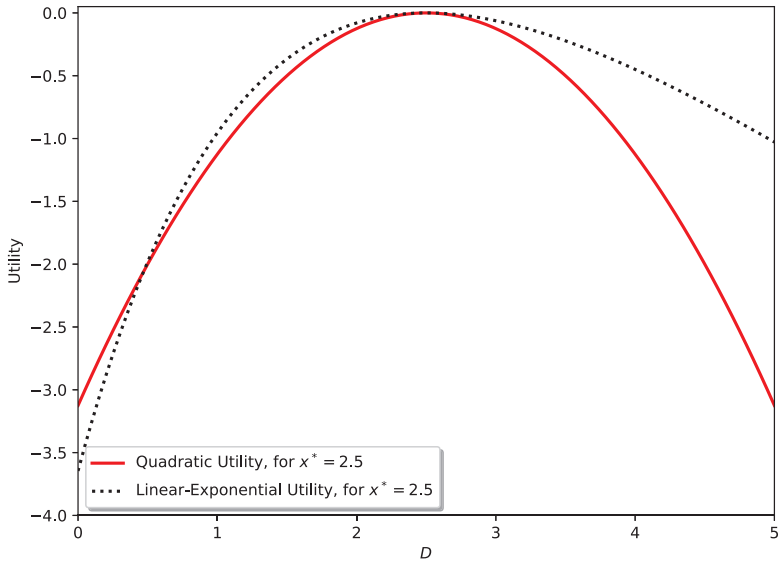

Figure 2a): Quadratic utility (in solid red) and linear-exponential asymmetric utility (in dotted black) for defaults $D \in [0, 5]$ given ideal point $x^* = 2.5$.

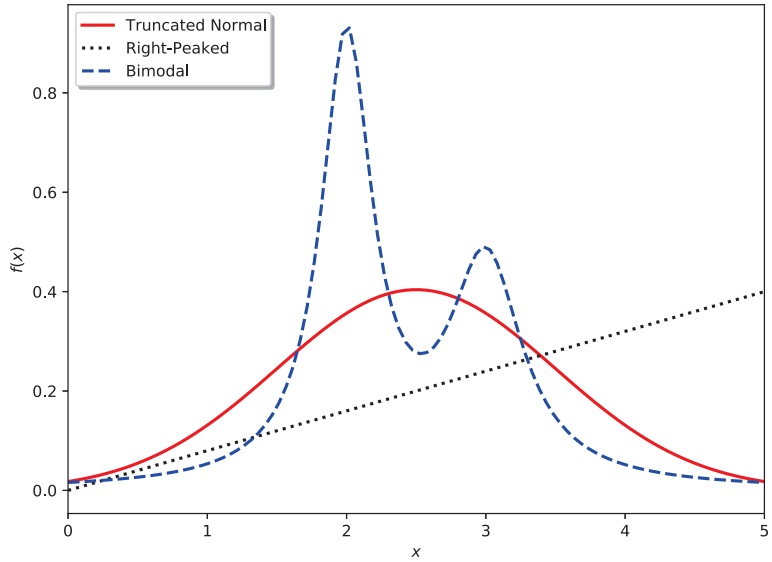

Figure 2b): Density of ideal point x^* over support $x \in [0, 5]$ for the three distributions studied: in solid red, the truncated Normal distribution, in dotted black, the right-peaked distribution, and in dashed blue the bimodal distribution.

Convergence and the quality of the approximation.

Focusing first on simulations in which heterogeneity is limited to ideal points, Part A of Table 2 explores settings that violate specific assumptions imposed in Carroll et al. (2009) and Goldin and Reck (2019). Several notable patterns emerge. First, we see numerical corroboration of Proposition 2: in each case, when λ is low enough to produce an opt-out frequency of 95%, $D_\Omega(\lambda)$ and $D_L(\lambda)$ are nearly identical. The maximal difference between the two, 0.0075, for the case of a right-peaked distribution and quadratic utility, is only 0.63% of the standard deviation of the ideal points x^* under this distribution, and the fraction of the potential welfare gain achieved through opt-out minimization, $\Delta_L(\lambda)$, is larger than 98% in all cases we consider.

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 $D_\Omega(\lambda)$ and $D_L(\lambda)$ (which again occurs for right-peaked preference distribution and quadratic utility), 0.0521, is only 4.4% of the standard deviation of x^*, and the corresponding weighted opt-out minimizing default achieves 97.84% of the total attainable welfare improvement. Even for the smallest opt-out percentage considered in the table, 40%, the approximations remain surprisingly good, with between 80% and 99.9% of welfare gain achieved across the parameterizations.

Figure 3, which focuses on the specification with an asymmetric linear-exponential utility function along with a bimodal ideal-point distribution, shows the relationship between $D_\Omega(\lambda)$ and $D_L(\lambda)$, as well as welfare losses, for λ yielding opt-out frequencies between roughly 14% and 91%. The limiting approximation is extremely good for parameters that produce opt-out rates above 50%, and remains reasonably good even with higher opt-out costs (lower opt-out rates).

Additional dimensions of heterogeneity.

The rest of Table 2 introduces various forms of heterogeneity. We allow η and x^* to vary independently across workers in Part B, and introduce correlation between them in Part E. Parts C and F are analogous, with β varying rather than η. We allow all three parameters to vary independently across workers in Part D, and introduce correlations among them in Part G. None of these changes produce meaningful divergences between the limiting values of the welfare-maximizing and weighted opt-out minimizing defaults. Moreover, we see only small divergences and modest inefficiencies from weighted opt-out minimization even when opt-out costs are high enough to produce opt-out frequencies as low as 40%: despite allowing for full interdependent heterogeneity, the weighted opt-out minimizing default captures at least 84% of the achievable welfare gains and above 95% in three of the five simulation cases

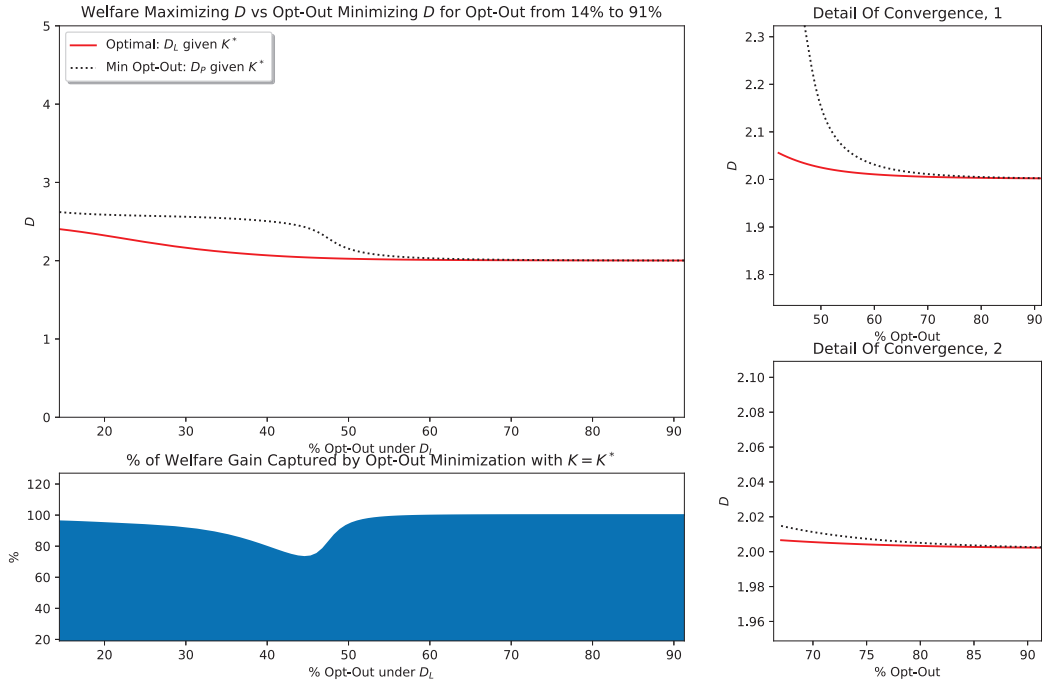

Figure 3: Illustration of welfare-maximizing and weighted opt-out-minimizing defaults. Simulations for a case in which the utility function is linear-exponential with asymmetry factor $\alpha = 0.75$, the present bias parameter is $\beta = 0.8$, opt-out costs are homogeneous, and the ideal-point density is bimodal with peaks at 2 and 3. The main panel shows the welfare-maximizing default $D_L(\lambda)$ and the weighted opt-out minimizing default $D_\Omega(\lambda)$, plotted for λ yielding opt-out frequencies between 14% and 91% for the welfare-maximizing default. Detail Panels 1 and 2 reproduce the main panel at higher resolution for opt-out frequencies closer to unity (λ close to zero), zooming in on the y-axis. The bottom panel displays the percentage of the potential welfare gain achieved through the weighted opt-out-minimization policy $D_\Omega(\lambda)$.

Unweighted opt-out minimization

In Table 3 we present the analogous simulation results for unweighted opt-out minimization. In other words, instead of finding the weighted opt-out minimizing default D_Ω, we find the unweighted opt-out minimizing default D_P^U for a given combination of the cost scaling factor λ and the specified utility function, distribution of x, and individual cost and bias parameters η and β. For cases with no heterogeneity, and with independence between η, β,and x^*, the simulation results match those in Table 2 for weighted opt-out minimization: weighted and unweighted opt-out minimization perform equally well relative to the welfare maximizing default option, just as our analytical results imply.

For cases involving non-independence between the distributions of x and η and/or β, unweighted opt-out minimization still performs comparably to weighted opt-out minimization. The percentage of the maximal welfare gain captured is slightly smaller in some instances and slightly higher in others, depending on parameterization and cost scaling factor λ.

A case with strong correlation between ideal points and opt-out costs

Proposition 3 tells us that unweighted opt-out minimization is asymptotically optimal as long as x^* and ρ are distributed independently of η and β. In the preceding simulations, it also performs well when η and β are mildly correlated with x^* (correlation coefficients ranging from -0.09 to -0.16). We now show through an additional simulation that a sufficiently strong correlation between x^* and η can significantly erode the limiting performance of the unweighted procedure. However, weighted opt-out minimization continues to maximize welfare in the limit (as Proposition 2 guarantees), and the asymptotic approximation remains accurate even with relatively low opt-out rates.

For the case depicted in Figure 4, we introduce a strong correlation between η and x^* (correlation coefficient of 0.8).[12] In particular, we assume the population falls into ten groups indexed $i \in 1, \ldots, 10$, each with equal mass. The cost scaling factor for each group, η_i, simply equals i. Ideal points are Normally distributed with means $\mu_i = 5\frac{i}{11}$. Thus, the ideal point for group 10 is ten times as large as for group 1, and group 10 faces ten times the opt-out cost of group 1 for any given λ. For all workers, we assume $\beta = 0.8$ and take the continuation utility function to be quadratic (with the same curvature) around the ideal point. The top panel of the figure shows the welfare-maximizing, unweighted opt-out minimizing, and weighted-opt-out-minimizing default options at various opt-out frequencies. Because unweighted opt-out minimization attaches too much importance to the workers with low as-if opt-out costs, it prescribes default contribution rates that are too low compared to the welfare maximizing defaults. In contrast, weighted opt-out minimization coincides with welfare maximization in the limit, and approximates the welfare-optimal solution to a high degree of accuracy at much lower opt-out rates.

[12] Note that the relevant consideration is the absolute magnitude of the correlation coefficient, not the sign.

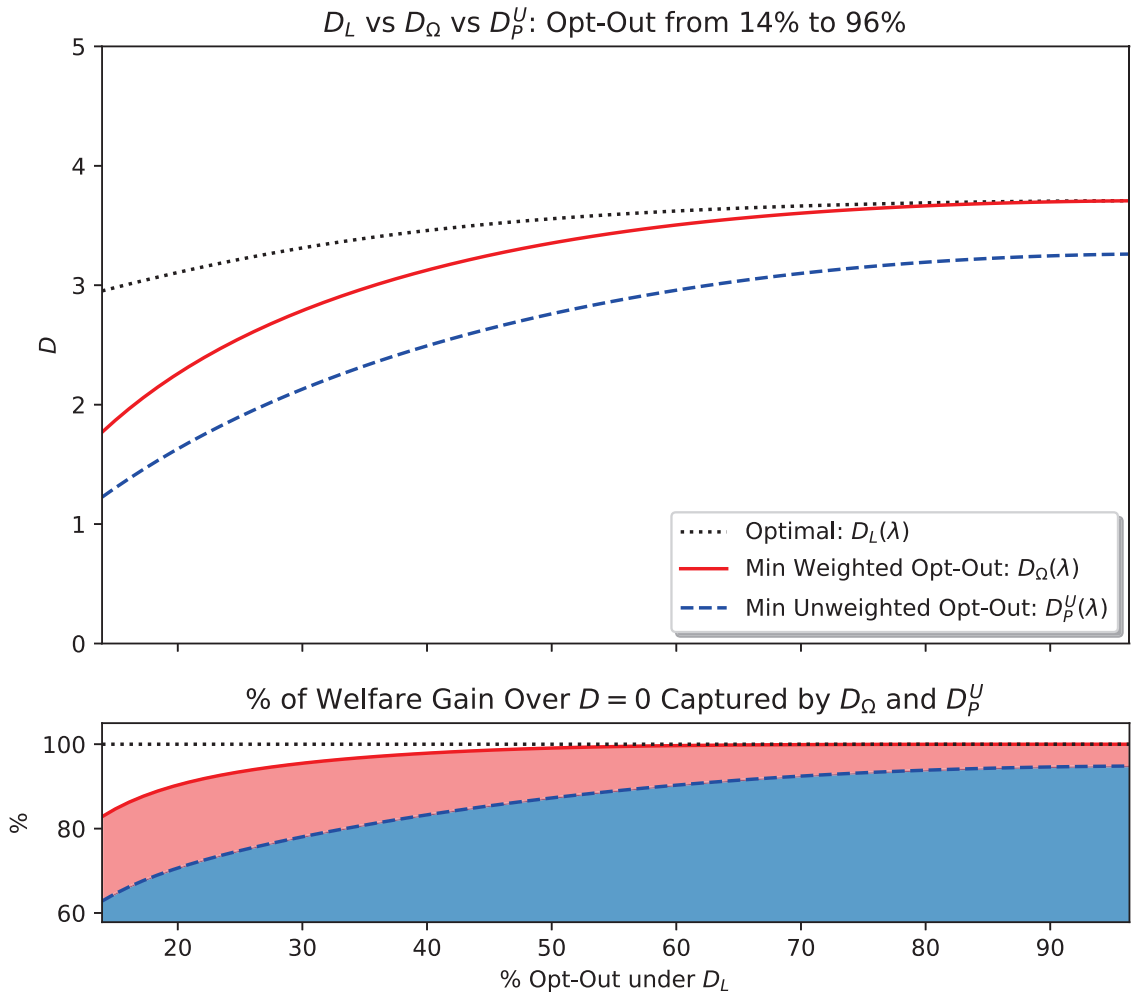

Figure 4: Weighted and unweighted opt-out minimization converge to different limits. Simulations for a case in which the population consists of ten groups, indexed by $i \in 1, \ldots, 10$, each with equal mass. For group i, ideal points are Normally distributed with mean $\mu_i = \frac{5i}{11}$, and the cost scaling factor is $\eta_i = i$, so that ideal points and opt-out costs are strongly correlated. The top panel displays the welfare-maximizing default $D_L(\lambda)$, the weighted opt-out minimizing default $D_\Omega(\lambda)$, and the unweighted opt-out minimizing default $D_P^U(\lambda)$, plotted for λ that yield opt-out frequencies between 14% and 96% under the welfare-maximizing default. The bottom panel displays the percentage of the potential welfare gain achieved through the weighted opt-out-minimization policy $D_\Omega(\lambda)$, and through the unweighed opt-out minimization policy $D_P^U(\lambda)$. The correlation between the ideal point and the cost parameter in this case is 0.7939.

The bottom panel of the figure shows how weighted and unweighted opt-out minimization perform relative to true welfare optimization. As in Figure 3 above, we express the welfare gain achieved through weighted and unweighted opt-out minimization as a fraction of the greatest possible gain. (In each case, we measure the gain relative to setting a default rate of zero.) Weighted opt-out minimization performs extremely well: it acheives more than 99% of the potential welfare gain as long as the opt-out rate exceeds 51%. In contrast, unweighted opt-out minimization does not achieve 95% of the potential welfare gain at for any opt-out rate.

8 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 option 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	$\|\|D_L - D_\Omega\|\|$	Δ_L %	$\|\|D_L - D_\Omega\|\|$	Δ_L %	$\|\|D_L - D_\Omega\|\|$	Δ_L %	$\|\|D_L - D_\Omega\|\|$	Δ_L %
				Quadratic	Right-peaked	0.00746	98.74	0.01637	98.55	0.05209	97.84	0.1523	95.91
				Quadratic	Bimodal	4e-05	>99.99	0.00016	>99.99	0.00157	>99.99	0.38688	80.24
A	X	X	X	Lin.-Exp.	Trunc. Normal	0.00028	>99.99	0.00111	>99.99	0.00706	>99.99	0.04871	99.89
				Lin.-Exp.	Right-peaked	0.00738	98.76	0.01595	98.60	0.04755	98.03	0.17391	95.79
				Lin.-Exp.	Bimodal	9e-05	>99.99	0.00037	>99.99	0.00322	99.99	0.43694	79.56
				Quadratic	Right-peaked	0.0105	99.04	0.02429	98.78	0.08873	97.64	0.36898	93.31
				Quadratic	Bimodal	5e-05	>99.99	0.00023	>99.99	0.00257	99.99	0.31874	89.22
B	✓	X	X	Lin.-Exp.	Trunc. Normal	0.0004	>99.99	0.00166	>99.99	0.0106	99.99	0.07152	99.77
				Lin.-Exp.	Right-peaked	0.01035	99.06	0.02344	98.83	0.07895	97.91	0.37275	93.14
				Lin.-Exp.	Bimodal	0.00014	>99.99	0.00055	>99.99	0.00483	99.97	0.38133	88.73
				Quadratic	Right-peaked	0.02498	93.90	0.05073	93.32	0.1421	91.23	0.14928	96.41
				Quadratic	Bimodal	6e-05	>99.99	0.00023	>99.99	0.00186	99.99	0.38453	77.51
C	X	✓	X	Lin.-Exp.	Trunc. Normal	0.00047	>99.99	0.00171	>99.99	0.01082	99.99	0.06328	99.80
				Lin.-Exp.	Right-peaked	0.02391	94.02	0.04923	93.55	0.12872	92.03	0.16041	96.40
				Lin.-Exp.	Bimodal	0.00013	>99.99	0.00054	>99.99	0.00372	99.98	0.43259	76.51
				Quadratic	Right-peaked	0.03737	95.11	0.07883	94.21	0.10346	97.16	0.2363	96.12
				Quadratic	Bimodal	8e-05	>99.99	0.00034	>99.99	0.00286	99.99	0.31678	87.70
D	✓	✓	X	Lin.-Exp.	Trunc. Normal	0.00066	>99.99	0.00261	>99.99	0.01594	99.99	0.08273	99.67
				Lin.-Exp.	Right-peaked	0.03674	95.21	0.07574	94.47	0.09278	97.46	0.23705	96.17
				Lin.-Exp.	Bimodal	0.0002	>99.99	0.00079	>99.99	0.00524	99.96	0.38829	86.60
				Quadratic	Right-peaked	0.01179	99.07	0.02735	98.78	0.10458	97.43	0.40534	95.23
				Quadratic	Bimodal	5e-05	>99.99	0.00022	>99.99	0.00232	99.99	0.31358	88.45
E	✓	X	✓	Lin.-Exp.	Trunc. Normal	0.0004	>99.99	0.00158	>99.99	0.01	99.99	0.02217	99.98
				Lin.-Exp.	Right-peaked	0.0116	99.09	0.02606	98.84	0.09193	97.75	0.35565	95.52
				Lin.-Exp.	Bimodal	0.00013	>99.99	0.00052	>99.99	0.00441	99.97	0.37815	86.61
				Quadratic	Right-peaked	0.02499	95.36	0.05243	94.76	0.14598	92.68	0.15553	96.24
				Quadratic	Bimodal	5e-05	>99.99	0.00021	>99.99	0.00178	>99.99	0.38785	78.11
F	X	✓	✓	Lin.-Exp.	Trunc. Normal	0.00038	>99.99	0.00157	>99.99	0.00985	>99.99	0.13201	98.86
				Lin.-Exp.	Right-peaked	0.02468	95.44	0.05016	94.96	0.13198	93.37	0.1609	96.64
				Lin.-Exp.	Bimodal	0.00013	>99.99	0.00048	>99.99	0.00361	99.98	0.43612	77.22
				Quadratic	Right-peaked	0.03933	95.84	0.08327	94.90	0.10962	97.25	0.24618	96.89
				Quadratic	Bimodal	8e-05	>99.99	0.00031	>99.99	0.00271	99.99	0.32685	86.23
G	✓	✓	✓	Lin.-Exp.	Trunc. Normal	0.0006	>99.99	0.00236	>99.99	0.01481	99.99	0.05568	99.88
				Lin.-Exp.	Right-peaked	0.03861	95.93	0.07873	95.16	0.09748	97.56	0.22373	97.02
				Lin.-Exp.	Bimodal	0.00018	>99.99	0.00072	>99.99	0.005	99.96	0.39577	84.90

Table 2: Overview of Simulation Results for *Weighted* Opt-Out Minimization. Each row is a separate simulation. The cost-scaling parameter λ is chosen to achieve the opt-out frequency detailed in columns (6) through (11). 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 D_L and the weighted opt-out minimizing default D_Ω 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 weighted opt-out minimizing default (Δ_Ω). A value of ≥ 99.99 indicates that the percentage would round to 100.00. We omit the combination of truncated Normal preference distribution and quadratic utility, as the two defaults coincide for any λ.

	(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	$\|\|D_L - D_P^U\|\|$	Δ_L %	$\|\|D_L - D_P^U\|\|$	Δ_L %	$\|\|D_L - D_P^U\|\|$	Δ_L %	$\|\|D_L - D_P^U\|\|$	Δ_L %
				Quadratic	Right-peaked	0.00746	98.74	0.01637	98.55	0.05209	97.84	0.1523	95.91
				Quadratic	Bimodal	4e-05	>99.99	0.00016	>99.99	0.00157	>99.99	0.38688	80.24
A	X	X	X	Lin.-Exp.	Trunc. Normal	0.00028	>99.99	0.00111	>99.99	0.00706	>99.99	0.04871	99.89
				Lin.-Exp.	Right-peaked	0.00738	98.76	0.01595	98.60	0.04755	98.03	0.17391	95.79
				Lin.-Exp.	Bimodal	9e-05	>99.99	0.00037	>99.99	0.00322	99.99	0.43694	79.56
				Quadratic	Right-peaked	0.0105	99.04	0.02429	98.78	0.08873	97.64	0.36898	93.31
				Quadratic	Bimodal	4e-05	>99.99	0.00016	>99.99	0.00151	>99.99	0.24805	92.96
B	✓	X	X	Lin.-Exp.	Trunc. Normal	0.00027	>99.99	0.00113	>99.99	0.00717	>99.99	0.04471	99.91
				Lin.-Exp.	Right-peaked	0.01035	99.06	0.02344	98.83	0.07895	97.91	0.06884	99.60
				Lin.-Exp.	Bimodal	9e-05	>99.99	0.00036	>99.99	0.00286	99.99	0.30148	92.14
				Quadratic	Right-peaked	0.02498	93.90	0.05073	93.32	0.1421	91.23	0.35871	86.08
				Quadratic	Bimodal	6e-05	>99.99	0.00025	>99.99	0.00212	99.99	0.3922	76.84
C	X	✓	X	Lin.-Exp.	Trunc. Normal	0.00051	>99.99	0.00188	>99.99	0.01193	99.99	0.07063	99.75
				Lin.-Exp.	Right-peaked	0.02391	94.02	0.04923	93.55	0.12872	92.03	0.16041	96.40
				Lin.-Exp.	Bimodal	0.00014	>99.99	0.0006	>99.99	0.00421	99.97	0.44213	75.80
				Quadratic	Right-peaked	0.03737	95.11	0.07883	94.21	0.10346	97.16	0.2363	96.12
				Quadratic	Bimodal	7e-05	>99.99	0.00029	>99.99	0.00218	99.99	0.25037	91.76
D	✓	✓	X	Lin.-Exp.	Trunc. Normal	0.00059	>99.99	0.00231	>99.99	0.01395	99.99	0.06538	99.79
				Lin.-Exp.	Right-peaked	0.03674	95.21	0.07574	94.47	0.09278	97.46	0.23705	96.17
				Lin.-Exp.	Bimodal	0.00018	>99.99	0.00068	>99.99	0.00395	99.98	0.31661	90.20
				Quadratic	Right-peaked	0.01179	99.07	0.02735	98.78	0.10458	97.43	0.23072	98.62
				Quadratic	Bimodal	3e-05	>99.99	0.00015	>99.99	0.00149	>99.99	0.28482	90.10
E	✓	X	✓	Lin.-Exp.	Trunc. Normal	0.00028	>99.99	0.00111	>99.99	0.00701	>99.99	0.01952	99.99
				Lin.-Exp.	Right-peaked	0.0116	99.09	0.02606	98.84	0.09193	97.75	0.04471	99.94
				Lin.-Exp.	Bimodal	9e-05	>99.99	0.00036	>99.99	0.00289	99.99	0.3417	88.42
				Quadratic	Right-peaked	0.02499	95.36	0.05243	94.76	0.14598	92.68	0.15553	96.24
				Quadratic	Bimodal	5e-05	>99.99	0.00023	>99.99	0.00197	99.99	0.39318	77.65
F	X	✓	✓	Lin.-Exp.	Trunc. Normal	0.00041	>99.99	0.00171	>99.99	0.01073	99.99	0.16545	98.50
				Lin.-Exp.	Right-peaked	0.02468	95.44	0.05016	94.96	0.13198	93.37	0.16337	96.55
				Lin.-Exp.	Bimodal	0.00014	>99.99	0.00053	>99.99	0.00399	99.98	0.44291	76.72
				Quadratic	Right-peaked	0.03933	95.84	0.08327	94.90	0.10962	97.25	0.24544	96.90
				Quadratic	Bimodal	7e-05	>99.99	0.00026	>99.99	0.00208	99.99	0.27154	89.83
G	✓	✓	✓	Lin.-Exp.	Trunc. Normal	0.00053	>99.99	0.00207	>99.99	0.01284	99.99	0.05791	99.87
				Lin.-Exp.	Right-peaked	0.03861	95.93	0.07873	95.16	0.09748	97.56	0.05647	99.84
				Lin.-Exp.	Bimodal	0.00016	>99.99	0.00062	>99.99	0.00382	99.98	0.33803	88.02

Table 3: Overview of Simulation Results for *Unweighted* Opt-Out Minimization. Each row is a separate simulation. The cost-scaling parameter λ is chosen to achieve the opt-out frequency detailed in columns (6) through (11). 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 D_L and the unweighted opt-out minimizing default D_P^U 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 unweighted opt-out minimizing default (Δ_P^U). A value of $\geq$ 99.99 indicates that the percentage would round to 100.00. We omit the combination of truncated Normal preference distribution and quadratic utility, as the two defaults coincide for any λ.

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. 2009. Behavioral welfare economics. *Journal of the European Economics Association* 7 (2-3): 267–319.

— . Forthcoming. In defense of behavioral welfare economics. *Journal of Economic Methodology*.

— . 2016. The good, the bad, and the ugly: a unified approach to behavioral welfare economics. *Journal of Benefit-Cost Analysis* 7 (1): 12–68.

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 Antonio Rangel. 2009. Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics. *Quarterly Journal of Economics* 124 (1): 51–104.

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.

— . 2020. Revealed-preference analysis with framing effects. *Journal of Political Economy* 128 (7): 2759–2794.

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).

O'Donoghue, Ted, and Matthew Rabin. 1999. Doing it now or later. *American Economic Review* 89 (1): 103–124.

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*.

Mathematical 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.

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

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

$$\Delta(D, x_1, \rho) = V(x_1, x_1, \rho) - V(D, x_1, \rho)$$

$$= [V(x_1, x_1, \rho) - V(x_2, x_1, \rho)] + [V(x_2, x_1, \rho) - V(D, x_1, \rho)]$$

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

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

where the first inequality follows from the optimality of x_1 for a worker with ideal point x_1, and the second follows from single crossing ($V_{12} > 0$) (Assumption 1, (iii)). It follows that opt-out from D by x_2 implies opt-out from D byt x_1, and opt-in to D by x_1 implies opt-in to D by x_2. An analogous argument establishes that a symmetric property holds for $x_1 > D$ and $x_2 \in (D, x_1)$. Furthermore, $\triangle(D, x, \rho)$ 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. □

Lemma 2. *$Q_\lambda(D,\theta)$ is continuous in D and converges uniformly to $Q(D,\theta)$ as $\lambda \to 0$.*

Proof. Continuity of Q_λ follows from the continuity of V and G. For subsequent reference, define $\overline{v}_{11} = \max_{x \in X, \rho \in [\underline{\rho}, \overline{\rho}]} V_{11}(x, x, \rho)$. Because $X \times [\underline{\rho}, \overline{\rho}]$ is compact and V_{11} is continuous, the maximum is well-defined. Adding the fact that $V_{11}(x, x, \rho) < 0$ for all $(x, \rho) \in X \times [\underline{\rho}, \overline{\rho}]$, we see that $\overline{v}_{11} < 0$. Further, using Taylor's theorem, we know there is some $\tilde{x}(D, x, \rho) \in [\min\{D, x\}, \max\{D, x\}]$ such that

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

It will then be convenient to define

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

We note for subsequent reference that, trivially, $\tilde{x}(D, D, \rho) = D$, which implies $d(D, D, \rho) \equiv -\frac{1}{2} V_{11}(D, D, \rho)$.

The proof of uniform convergence proceeds in a series of steps. The arguments reference the opt-in window, $S(D, \theta, \lambda) \equiv [x_l(D, \theta, \lambda), x_h(D, \theta, \lambda)]$, identified in Lemma 1. Throughout, we use the symbol $\rightrightarrows$ to denote uniform convergence.

Step 1: For each $\lambda > 0$, there exists $\nu(\lambda) > 0$ with $\lim_{\lambda \to 0} \nu(\lambda) = 0$ such that, for all $D \in X$, $\theta \in \Theta$, and $x \in S(D, \theta, \lambda)$, we have $|D - x| \leq \nu(\lambda)$.

We establish the claim by constructing the requisite function:

$$\nu(\lambda) \equiv \max_{(D,\theta) \in X \times \Theta,\ x \in S(D,\theta,\lambda)} |D - x|$$

Because the objective function is continuous and the contraint set is compact, the maximum exists.

To complete Step 1, we must prove that $\lim_{\lambda \to 0} \nu(\lambda) = 0$. Our strategy is to show that, for all $\varepsilon > 0$, there exists $\lambda^*(\varepsilon)$ such that $\lambda < \lambda^*(\varepsilon)$ implies $|D - x| < \varepsilon$ for all $(D, \theta) \in X \times \Theta$ and $x \in S(D, \theta, \lambda)$. For such λ, it must then be the case that $\nu(\lambda) < \varepsilon$.

For any $\varepsilon > 0$, we define $\Psi(\varepsilon) \equiv \left\{(D, x, \rho) \in X^2 \times \left[\underline{\rho}, \overline{\rho}\right] \middle| \, |D - x| \geqq \varepsilon\right\}$ and $\sigma(\varepsilon) \equiv \min_{(D,x,\rho) \in \Psi(\varepsilon)} \Delta(D, x, \rho)$. Existence of $\sigma(\varepsilon)$ follows from continuity of the objective function and compactness of the constraint set. Because we have assumed that $\Delta(D, x, \rho) > 0$ whenever $D \neq x$, we know that $\sigma(\varepsilon) > 0$. Let $\lambda^*(\varepsilon) \equiv \frac{\beta \sigma(\varepsilon)}{\overline{\eta}} > 0$. For $\lambda < \lambda^*(\varepsilon)$, any $x \in S(D, \theta, \lambda)$ satisfies $\Delta(D, x, \rho) \leq \frac{\lambda \eta}{\beta} < \frac{\lambda^*(\varepsilon) \overline{\eta}}{\beta} = \sigma(\varepsilon)$. But then we must have $(D, x, \rho) \notin \Psi(\varepsilon)$, which means $|D - x| < \varepsilon$, as desired.

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

$$|f(D \mid \theta) - f(x \mid \theta)| < \delta(\lambda) \tag{9}$$

and

$$|d(D, D, \rho) - d(D, x, \rho)| < \delta(\lambda). \tag{10}$$

First consider f. Because F is twice-continuously differentiable and X and Θ are compact, f is Lipschitz-continuous on $X \times \Theta$. Accordingly, there exists $M_f > 0$ such that $|f(D \mid \theta) - f(x \mid \theta)| < M_f |D - x|$. In Step 1, we showed that $|D - x| \leq \nu(\lambda)$ for all $D \in X$, $\theta \in \Theta$, and $x \in S(D, \theta, \lambda)$. Therefore $|f(D \mid \theta) - f(x \mid \theta)| < M_f \nu(\lambda)$ for all $D \in X$, $\theta \in \Theta$, and $x \in S(D, \theta, \lambda)$.

Now consider d. Because V has continuous third derivatives and X and Θ are compact, $V_{11}(D, x, \rho)$ is Lipschitz-continuous on $X^2 \times \Theta$. It follows that there exists $M_1 > 0$ for which

$$|V_{11}(y,x,\rho) - V_{11}(x,x,\rho)| < M_1 |y-x| . \tag{11}$$

as well as $M_2 > 0$ for which

$$|V_{11}(x,y,\rho) - V_{11}(x,x,\rho)| < M_2 |y-x| . \tag{12}$$

Consequently,

$$\begin{aligned} |d(D,D,\rho) - d(D,x,\rho)| &= \tfrac{1}{2} |V_{11}(\tilde{x}(D,x,\rho),x,\rho) - V_{11}(D,D,\rho)| \\ &= \tfrac{1}{2} |V_{11}(\tilde{x}(D,x,\rho),x,\rho) - V_{11}(D,x,\rho) + V_{11}(D,x,\rho) - V_{11}(D,D,\rho)| \\ &\leq \tfrac{1}{2} |V_{11}(\tilde{x}(D,x,\rho),x,\rho) - V_{11}(D,x,\rho)| + \tfrac{1}{2} |V_{11}(D,x,\rho) - V_{11}(D,D,\rho)| \\ &< M_1 |\tilde{x}(D,x,\rho) - x| + M_2 |D-x| \\ &\leq (M_1 + M_2) |D-x| , \end{aligned} \tag{13}$$

where the second inequality follows from (11) and (12), and the final inequality follows from the fact that $\tilde{x}(D,x,\rho) \in [\min\{D,x\}, \max\{D,x\}]$. In Step 1, we showed that $|D-x| \leq \nu(\lambda)$ for $x \in S(D,\theta,\lambda)$. Substituting into (13), we obtain $|d(D,D,\rho) - d(D,x,\rho)| < (M_1+M_2)\,\nu(\lambda)$ for $x \in S(D,\theta,\lambda)$.

To complete Step 2, we simply define $\delta(\lambda) \equiv \max\{M_f, (M_1+M_2)\} \cdot \nu(\lambda)$.

Step 3: Proof of the lemma.

From Step 2, we know that for all $x \in S(D,\theta,\lambda)$, we have

$$d(D,D,\rho) - \delta(\lambda) < d(D,x,\rho) < d(D,D,\rho) + \delta(\lambda)$$

Because $d(D,D,\rho) \geq \frac{-\overline{v}_{11}}{2} > 0$ and $\lim_{\lambda \to 0} \delta(\lambda) = 0$, there exists λ^c such that $\lambda < \lambda^c$ implies $d(D,D,\rho) - \delta(\lambda) > 0$ for all $D \in X$, $\rho \in \left[\underline{\rho}, \overline{\rho}\right]$. It follows that, for $\lambda < \lambda^c$ and all $x \in S(D,\theta,\lambda)$,

$$0 < (d(D,D,\rho) - \delta(\lambda))(D-x)^2 < \Delta(D,x,\rho) < (d(D,D,\rho) + \delta(\lambda))(D-x)^2$$

Accordingly, $\Delta(D,x,\rho) \leq \frac{\lambda\eta}{\beta}$ (i.e., $x \in S(D,\theta,\lambda)$) implies $(D-x)^2 < \left(\frac{\lambda\eta}{\beta}\right) \frac{1}{d(D,D,\rho) - \delta(\lambda)}$, and $\Delta(D,x,\rho) > \frac{\lambda\eta}{\beta}$ (i.e., $x \notin S(D,\theta,\lambda)$) implies $(D-x)^2 > \left(\frac{\lambda\eta}{\beta}\right) \frac{1}{d(D,D,\rho) + \delta(\lambda)}$. Thus,

$$S(D,\theta,\lambda) \subset \left(D - \left(\left(\frac{\lambda\eta}{\beta}\right) \frac{1}{d(D,D,\rho) - \delta(\lambda)} \right)^{\frac{1}{2}}, D + \left(\left(\frac{\lambda\eta}{\beta}\right) \frac{1}{d(D,D,\rho) - \delta(\lambda)} \right)^{\frac{1}{2}} \right) \tag{14}$$

$$S(D,\theta,\lambda) \supset \left(D - \left(\left(\frac{\lambda\eta}{\beta}\right) \frac{1}{d(D,D,\rho) + \delta(\lambda)} \right)^{\frac{1}{2}}, D + \left(\left(\frac{\lambda\eta}{\beta}\right) \frac{1}{d(D,D,\rho) + \delta(\lambda)} \right)^{\frac{1}{2}} \right) \tag{15}$$

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

$$
\begin{aligned}
2\left(f(D \mid \theta)+\delta(\lambda)\right)\left(\left(\tfrac{\lambda\eta}{\beta}\right)\tfrac{1}{d(D,D,\rho)-\delta(\lambda)}\right)^{\frac{1}{2}} &> \Pr\left[\Delta(D,x,\rho)\leq \tfrac{\lambda\eta}{\beta}\middle|\,\theta\right] \\
&> 2\left(f(D \mid \theta)-\delta(\lambda)\right)\left(\left(\tfrac{\lambda\eta}{\beta}\right)\tfrac{1}{d(D,D,\rho)+\delta(\lambda)}\right)^{\frac{1}{2}}
\end{aligned}
$$

It thus follows that

$$
\begin{aligned}
\left(f(D \mid \theta)+\delta(\lambda)\right)\left(\left(\tfrac{\eta}{\beta}\right)\tfrac{1}{-\frac{1}{2}V_{11}(D,D,\rho)-\delta(\lambda)}\right)^{\frac{1}{2}} &> Q_\lambda(D,\theta) \\
&> \left(f(D \mid \theta)-\delta(\gamma)\right)\left(\left(\tfrac{\eta}{\beta}\right)\tfrac{1}{-\frac{1}{2}V_{11}(D,D,\rho)+2\delta(\lambda)}\right)^{\frac{1}{2}}
\end{aligned}
$$

As $\lambda \to 0$, both sides converge to the same value: $f(D \mid \theta)\left(\left(\tfrac{\eta}{\beta}\right)\tfrac{1}{-\frac{1}{2}V_{11}(D,D,\rho)}\right)^{\frac{1}{2}} = Q(D,\theta)$. Therefore we know that $Q_\lambda(D,\theta)$ converges pointwise to $Q(D,\theta)$.

To show that convergence is uniform, notice first that, by construction, $Q(D,\theta)$ lies within the same bounds. We consider the difference between the upper and lower bounds on $Q_\lambda(D,\theta)$ and $Q(D,\theta)$:

$$
\begin{aligned}
\xi(D,\theta,\lambda) &= \left(f(D \mid \theta)+\delta(\lambda)\right)\left(\left(\tfrac{\eta}{\beta}\right)\tfrac{1}{-\frac{1}{2}V_{11}(D,D,\rho)-\delta(\lambda)}\right)^{\frac{1}{2}} \\
&\quad -\left(f(D \mid \theta)-\delta(\gamma)\right)\left(\left(\tfrac{\eta}{\beta}\right)\tfrac{1}{-\frac{1}{2}V_{11}(D,D,\rho)+\delta(\lambda)}\right)^{\frac{1}{2}} \\
&> 0
\end{aligned}
$$

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

$$
\xi(D,\theta,\lambda) < \left(\frac{\overline{\eta}}{\underline{\beta}}\right)^{\frac{1}{2}}\left[\left(f^{max}+\delta(\lambda)\right)\left(\frac{1}{-\frac{1}{2}\overline{v}_{11}-\delta(\lambda)}\right)^{\frac{1}{2}} - \left(f^{max}-\delta(\lambda)\right)\left(\frac{1}{-\frac{1}{2}\overline{v}_{11}+\delta(\lambda)}\right)^{\frac{1}{2}}\right] \equiv \bar{\xi}(\lambda)
$$

The right-hand side of this expression converges to 0 as $\lambda \to 0$, and does not depend upon D or θ. Therefore, we have $Q_\lambda(D,\theta) \rightrightarrows Q(D,\theta)$. □

Proof of Proposition 1 We claim that $\Omega_\lambda(D) \rightrightarrows \Omega(D)$. To prove the claim, we write:

$$
\begin{aligned}
|\Omega_\lambda(D)-\Omega(D)| &\leq \int_\theta \eta\left|1-\frac{1}{3\beta}\right| |Q_\lambda(D,\theta)-Q(D,\theta)|\, dG(\theta) \\
&\leq \bar{\eta}\phi\bar{\xi}(\lambda)
\end{aligned}
$$

where $\phi \equiv \max\left\{\left|1-\frac{1}{3\underline{\beta}}\right|, \left|1-\frac{1}{3\overline{\beta}}\right|\right\}$, and $\bar{\xi}(\lambda)$ is defined in the proof of Lemma 2. Uniform convergence follows from the fact that $\bar{\xi}(\lambda) \to 0$ as $\lambda \to 0$. Because $\Omega_\lambda(D) \rightrightarrows \Omega(D)$ and

$\Omega(D)$ is bounded on X,[13] we know that the maximizers of $\Omega_\lambda(D)$ converge to the maximizer of $\Omega(D)$. Proposition 1 follows. □

Our next result concerns the limiting behavior of the following function:

$$Z_\lambda(D,\theta) \equiv \frac{\mathbb{E}\left[\Delta(D,x,\rho) \mid \theta, \Delta(D,x,\rho) \leq \frac{\lambda\eta}{\beta}\right]}{\lambda}$$

Lemma 3. *$Z_\lambda(D,\theta)$ converges uniformly to $\frac{\eta}{3\beta}$ as $\lambda \to 0$.*

Proof. Because $0 < \mathbb{E}\left[\Delta(D,x,\rho)|\,\theta, \Delta(D,x,\rho) \leq \frac{\lambda\eta}{\beta}\right] < \frac{\lambda\eta}{\beta}$ for all λ, we know that $Z_\lambda(D,\theta)$ is bounded between 0 and $\frac{\eta}{\beta}$. Observe that:

$$Z_\lambda(D,\theta) = \frac{\mathbb{E}\left[\Delta(D,x,\rho)|\,\theta, \Delta(D,x,\rho) \leq \frac{\lambda\eta}{\beta}\right]}{\lambda} = \frac{\mathbb{E}\left[\Delta(D,x,\rho)\mathbf{1}_{\Delta(D,x,\rho)\leq\frac{\lambda\eta}{\beta}} \mid \theta\right]}{\lambda \Pr\left[\Delta(D,x,\rho) \leq \frac{\lambda\eta}{\beta} \mid \theta\right]} \tag{16}$$

The denominator equals $2Q_\lambda(D,\theta)\lambda^{\frac{3}{2}}$.

Defining $\delta(\lambda)$ and λ^c as in the proof of Lemma 2, Step 3, as long as $\lambda < \lambda^c$ (i.e., so that $d(D,D,\rho) - \delta(\lambda) > 0$ for all $D \in \mathbf{D}$, $\rho \in \left[\underline{\rho}, \overline{\rho}\right]$), the numerator of (16) is bounded above by:

$$\mathbb{E}\left[\Delta(D,x,\rho)\mathbf{1}_{\Delta(D,x,\rho)\leq\frac{\lambda\eta}{\beta}} \mid \theta\right] \leq \int_{D-\left(\left(\frac{\lambda\eta}{\beta}\right)\frac{1}{d(D,D,\rho)-\delta(\lambda)}\right)^{\frac{1}{2}}}^{D+\left(\left(\frac{\lambda\eta}{\beta}\right)\frac{1}{d(D,D,\rho)-\delta(\lambda)}\right)^{\frac{1}{2}}} (d(D,D,\rho)+\delta(\lambda))\,(D-x)^2\,(f(D \mid \theta)+\delta(\lambda))\,dx$$

$$= \frac{1}{3}(f(D \mid \theta)+\delta(\lambda))(d(D,D,\rho)+\delta(\lambda))\left(\frac{\lambda\eta}{\beta}\right)^{\frac{3}{2}}\left(\frac{2}{(d(D,D,\rho)-\delta(\lambda))^{\frac{3}{2}}}\right)$$

$$= \frac{2}{3}Q(D,\theta)\left(1+\frac{\delta(\lambda)}{f(D \mid \theta)}\right)\left(1+\frac{\delta(\lambda)}{d(D,D,\rho)}\right)\lambda^{\frac{3}{2}}\left(\frac{\eta}{\beta}\right)\left(1-\frac{\delta(\lambda)}{d(D,D,\rho)}\right)^{-\frac{3}{2}}$$

where the inequality in the first line follows from (9), (10), and (14) (given that the integrand is strictly positive). It then follows from (16) that

$$Z_\lambda(D,\theta) \leq \frac{1}{3}\left(\frac{Q(D,\theta)}{Q_\lambda(D,\theta)}\right)\left(1+\frac{\delta(\lambda)}{f(D \mid \theta)}\right)\left(1+\frac{\delta(\lambda)}{d(D,D,\rho)}\right)\left(\frac{\eta}{\beta}\right)\left(1-\frac{\delta(\lambda)}{d(D,D,\rho)}\right)^{-\frac{3}{2}}$$

$$\equiv \overline{Z}_\lambda(D,\theta)$$

[13] This claim follows from the fact that f and η are bounded above, while V_{11} and β are bounded away from zero.

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

We use a similar strategy to derive a lower bound on $Z_\lambda(D, \theta)$. Because $\lim_{\lambda\to 0} \delta(\lambda) = 0$, there exists λ^f such that $\lambda < \lambda^f$ implies $f^{min} > \delta(\lambda)$. As long as $\lambda < \lambda^f$ (which ensures $f(D \mid \theta) - \delta(\lambda) > 0$ for all $D \in X$), the numerator of (16) is bounded below by:

$$\mathbb{E}\left[\Delta(D, x, \rho)\mathbf{1}_{\Delta(D,x,\rho)\leq\frac{\lambda\eta}{\beta}} \mid \theta\right] \geq \int_{D-\left(\left(\frac{\lambda\eta}{\beta}\right)\frac{1}{d(D,D,\rho)+\delta(\lambda)}\right)^{\frac{1}{2}}}^{D+\left(\left(\frac{\lambda\eta}{\beta}\right)\frac{1}{d(D,D,\rho)+\delta(\lambda)}\right)^{\frac{1}{2}}} (d(D, D, \rho) - \delta(\lambda))(D - x)^2 (f(D \mid \theta) - \delta(\lambda))\, dx$$

A parallel argument then implies that

$$Z_\lambda(D, \theta) \geq \frac{1}{3}\left(\frac{Q(D, \theta)}{Q_\lambda(D, \theta)}\right)\left(1 - \frac{\delta(\lambda)}{f(D \mid \theta)}\right)\left(1 - \frac{\delta(\lambda)}{d(D, D, \rho)}\right)\left(\frac{\eta}{\beta}\right)\left(1 + \frac{\delta(\lambda)}{d(D, D, \rho)}\right)^{-\frac{3}{2}}$$
$$\equiv \underline{Z}_\lambda(D, \theta)$$

Reasoning as for the upper bound, we have $\underline{Z}_\lambda(D, \theta) \rightrightarrows \frac{\eta}{3\beta}$ as $\lambda \to 0$.

Because the upper and lower bounds both converge uniformly to $\frac{\eta}{3\beta}$, we can infer that $Z_\lambda(D, \theta) \rightrightarrows \frac{\eta}{3\beta}$ as $\lambda \to 0$. □

Proof of Proposition 2 Notice that we can rewrite the function $W_\lambda(D)$, which we defined in Section 3.2, as follows:

$$W_\lambda(D) \equiv \int_\theta Q_\lambda(D, \theta)\left[\eta - Z_\lambda(D, \theta)\right] dG(\theta)$$

It follows that

$$W_\lambda(D) - \Omega_\lambda(D) = \int_\theta Q_\lambda(D, \theta)\left(\frac{\eta}{3\beta} - Z_\lambda(D, \theta)\right) dG(\theta),$$

Choosing λ sufficiently small so as to insure $\delta(\lambda) < \min\left\{-\frac{1}{4}\overline{v}_{11}, f^{max}\right\}$, we have

$$0 < Q_\lambda(D,\theta) < (f(D \mid \theta) + \delta(\lambda)) \left(\frac{\eta}{\beta}\right)^{\frac{1}{2}} \left(\frac{1}{d(D,D,\rho) - \delta(\lambda)}\right)^{\frac{1}{2}} < 4f^{max} \left(\frac{\overline{\eta}}{\underline{\beta}}\right)^{\frac{1}{2}} \left(\frac{1}{-\overline{v}_{11}}\right)^{\frac{1}{2}} \equiv C$$

Consequently,

$$|W_\lambda(D) - \Omega_\lambda(D)| \leq \int_\theta C \left| Z_\lambda(D,\theta) - \frac{\eta}{3\beta} \right| dG(\theta).$$

According to Lemma 3, $Z_\lambda(D,\theta) \rightrightarrows \frac{\eta}{3\beta}$, which means that for any $\varepsilon > 0$, there exists $\lambda_\varepsilon > 0$ such that $\left| Z_\lambda(D,\theta) - \frac{\eta}{3\beta} \right| < \varepsilon$ for all $\lambda < \lambda_\varepsilon$. But then we have

$$|W_\lambda(D) - \Omega_\lambda(D)| \leq \int_\theta C\varepsilon dG(\theta) = C\varepsilon.$$

It follows that $W_\lambda(D) - \Omega_\lambda(D) \rightrightarrows 0$ as $\lambda \to 0$. Because $\Omega_\lambda(D) \rightrightarrows \Omega(D)$, we then have $W_\lambda(D) \rightrightarrows \Omega(D)$. Because $\Omega(D)$ is bounded on X (see the proof of Proposition 1), we know that the maximizers of $W_\lambda(D)$ converge to the maximizer of $\Omega(D)$. Proposition 1 follows. □

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

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

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

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

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. □

Proof of Proposition 5 Define the opt-in set for given D, θ, λ.

$$S(D,\theta,\lambda) \equiv \left\{ y \in Y \mid \Delta(D,y,\rho) \leq \frac{\lambda\eta}{\beta} \right\}$$

Because we have not assumed single-crossing, we cannot guarantee that $S(D,\theta,\lambda)$ is an interval. However, we can still begin with essentially the same step as in the proof of Lemma 2.

Step 1: For each $\lambda > 0$, there exists $\nu(\lambda) > 0$ with $\lim_{\lambda\to 0}\nu(\lambda) = 0$ such that, for all $D \in X$, $\theta \in \Theta$, and $y \in S(D,\theta,\lambda)$, we have $|D - x^*(y)| \leq \nu(\lambda)$.

We establish the claim by constructing the requisite function:

$$\nu(\lambda) \equiv \max_{(D,\theta)\in X\times\Theta,\ y\in S(D,\theta,\lambda)} |D - x^*(y)|$$

Because the objective function is continuous and the contraint set is easily shown to be compact, the maximum exists.

To complete Step 1, we must prove that $\lim_{\lambda\to 0}\nu(\lambda) = 0$. Our strategy is to show that, for all $\varepsilon > 0$, there exists $\lambda^*(\varepsilon)$ such that $\lambda < \lambda^*(\varepsilon)$ implies $|D - x^*(y)| < \varepsilon$ for all $(D,\theta) \in X \times \Theta$ and $y \in S(D,\theta,\lambda)$. For such λ, it must then be the case that $\nu(\lambda) < \varepsilon$.

For any $\varepsilon > 0$, we define $\Psi(\varepsilon) \equiv \left\{(D,y,\rho) \in X \times Y \times \left[\underline{\rho},\overline{\rho}\right] \middle|\, |D - x^*(y)| \geqq \varepsilon\right\}$ and $\sigma(\varepsilon) \equiv \min_{(D,y,\rho)\in\Psi(\varepsilon)} \Delta(D,y,\rho)$. Existence of $\sigma(\varepsilon)$ follows from continuity of the objective function and compactness of the constraint set. Because we have assumed that $\Delta(D,y,\rho) > 0$ whenever $D \neq x^*(y)$, we know that $\sigma(\varepsilon) > 0$. Let $\lambda^*(\varepsilon) \equiv \frac{\underline{\beta}\sigma(\varepsilon)}{\overline{\eta}} > 0$. For $\lambda < \lambda^*(\varepsilon)$, any $y \in S(D,\theta,\lambda)$ satisfies $\Delta(D,y,\rho) \leq \frac{\lambda\eta}{\underline{\beta}} < \frac{\lambda^*(\varepsilon)\overline{\eta}}{\underline{\beta}} = \sigma(\varepsilon)$. But then we must have $(D,x,\rho) \notin \Psi(\varepsilon)$, which means $|D - x^*(y)| < \varepsilon$, as desired.

Throughout the remaining steps of this proof, we will focus on λ sufficiently small so that, for all $z, z' \in Z$, $[z - \nu(\lambda), z + \nu(\lambda)] \cap [z' - \nu(\lambda), z' + \nu(\lambda)] = \emptyset$. (This is possible because Z is a finite set.) For any such λ, we will define $Z^\nu(\lambda) \equiv \cup_{z\in Z}[z - \nu(\lambda), z + \nu(\lambda)]$ and $X^\nu(\lambda) \equiv X \setminus Z^\nu(\lambda)$.

Step 2: $\lim_{\lambda\to 0} D_{\hat{\Omega}}(\lambda) \to z^*$.

Recalling our assumption that the derivative of $x^*(y)$ is uniformly bounded away from zero outside Y_0, we know there exists $\delta > 0$ such that $\frac{dx^*(y)}{dy} > \delta$ for $y \in Y \setminus Y_0$. Using Step 1, we then have:

$$\hat{\Omega}_\lambda(D) \leq \begin{cases} \overline{\eta} f^{max} \frac{2\nu(\lambda)}{\delta} & \text{if } D \in X^\nu(\lambda) \\ \hat{\Omega}(z) + \overline{\eta} f^{max} \frac{2\nu(\lambda)}{\delta} & \text{if } D \in [z - \nu(\lambda), z + \nu(\lambda)] \text{ for some } z \in Z \end{cases}$$

Let

$$\varepsilon^* = \frac{1}{2}\left[\hat{\Omega}(z^*) - \max_{z\in Z\setminus z^*} \hat{\Omega}(z)\right] > 0.$$

We know there exists λ^* such that, for all $\lambda < \lambda^*$,

$$\nu(\lambda) < \frac{\delta\varepsilon^*}{2\overline{\eta} f^{max}}.$$

For such λ, we have

$$\hat{\Omega}_\lambda(D) \leq \begin{cases} \varepsilon^* & \text{if } D \in X^\nu(\lambda) \\ \hat{\Omega}(z) + \varepsilon^* & \text{if } D \in [z - \nu(\lambda), z + \nu(\lambda)] \text{ for some } z \in Z \setminus z^* \end{cases}$$

Now we claim that if $D \in [z - \nu(\lambda), z + \nu(\lambda)]$ for some $z \in Z \setminus z^*$, then $\hat{\Omega}_\lambda(D) \leq \hat{\Omega}_\lambda(z^*) - \varepsilon^*$. By the definition of ε^*, we have $\hat{\Omega}(z^*) \geq \hat{\Omega}(z) + 2\varepsilon^*$ for $z \in Z \setminus z^*$. From the definitions of $\hat{\Omega}$ and $\hat{\Omega}_\lambda$, we know that $\hat{\Omega}_\lambda(z^*) \geq \hat{\Omega}(z^*)$. Combining these inequalities, we have $\hat{\Omega}_\lambda(z^*) - \varepsilon^* \geq \hat{\Omega}(z) + \varepsilon^*$. But we have just shown that, for such D, we have $\hat{\Omega}(z) + \varepsilon^* \geq \hat{\Omega}_\lambda(D)$. Combining the last two inequalities yields the desired conclusion.

Next we claim that if $D \in X^\nu(\lambda)$, then $\hat{\Omega}_\lambda(D) < \hat{\Omega}_\lambda(z^*)$. We know from the last claim that, for $z \in Z \setminus z^*$, we have $\hat{\Omega}_\lambda(z) \leq \hat{\Omega}_\lambda(z^*) - \varepsilon^*$. Furthermore, $\hat{\Omega}_\lambda(z) > 0$. It follows that $\hat{\Omega}_\lambda(z^*) \geq \hat{\Omega}_\lambda(z) + \varepsilon^* > \varepsilon^* \geq \hat{\Omega}_\lambda(D)$ for such D.

Putting these two claims together, we conclude that, for $\lambda < \lambda^*$, we must have

$$D_{\hat{\Omega}}(\lambda) \in [z^* - \nu(\lambda), z^* + \nu(\lambda)]$$

(because we have shown that any other D yields a lower value of the objective function than z^*). Letting $\lambda \to 0$, we see that $D_{\hat{\Omega}}(\lambda) \to z^*$.

Step 3: $\lim_{\lambda \to 0} W_{\hat{\Omega}}(\lambda) \to z^*$.

Equation (8) tells us that

$$\begin{aligned} \hat{W}_\lambda(D) &= \hat{\Omega}_\lambda(D) - \textstyle\int_\theta \Pr\left(\Delta(D, x^*(y), \rho) \leq \frac{\lambda\eta}{\beta} \middle| \theta\right) \\ &\quad \times \left[\tfrac{1}{\lambda} E\left(\Delta(D, x^*(y), \rho) \mid \Delta(D, x^*(y), \rho) \leq \tfrac{\lambda\eta}{\beta} \middle| \theta\right)\right] dG(\theta), \end{aligned}$$

from which it follows immediately that $\hat{W}_\lambda(D) \leq \hat{\Omega}_\lambda(D)$ for all $D \in X$.

We now claim that $\lim_{\lambda \to \infty} \hat{W}_\lambda(z^*) = \hat{\Omega}(z^*)$. Because the probability term in the integrand is bounded between 0 and 1, we can demonstrate this claim by showing that the bracketed term in the integrand converges uniformly to zero. Using the fact that

$$E\left(\Delta(z^*, x^*(y), \rho) \mid \Delta(z^*, x^*(y), \rho) \leq \frac{\lambda\eta}{\beta} \middle| \theta\right) \leq \frac{\lambda\eta}{\beta} \leq \frac{\lambda\overline{\eta}}{\underline{\beta}},$$

we have

$$\begin{aligned} \tfrac{1}{\lambda} E\left(\Delta(z^*, x^*(y), \rho) \mid \Delta(z^*, x^*(y), \rho) \leq \tfrac{\lambda\eta}{\beta} \middle| \theta\right) &\leq \tfrac{1}{\lambda} \frac{0 \times \Pr(\Delta(z^*, x^*(y), \rho) = 0 | \theta) + \left(\frac{\lambda\overline{\eta}}{\underline{\beta}}\right) \Pr\left(0 < \Delta(z^*, x^*(y), \rho) \leq \frac{\lambda\eta}{\beta} \middle| \theta\right)}{\Pr\left(\Delta(z^*, x^*(y), \rho) \leq \frac{\lambda\eta}{\beta} \middle| \theta\right)} \\ &< \frac{2\overline{\eta} f^{max} \nu(\lambda)}{\delta \underline{\beta} \Pr(x^*(y) = z^* | \theta)}, \end{aligned}$$

which implies the desired convergence property.

The preceding argument implies that there is some $\lambda^0 \in (0, \lambda^*)$ such that, for $\lambda > \lambda^0$,

we have $\hat{W}_\lambda(z^*) \geq \hat{\Omega}_\lambda(z^*) - \varepsilon^*$. It follows that, for all $D \neq [z^* - \nu(\lambda), z^* + \nu(\lambda)]$, we have

$$\hat{W}_\lambda(D) \leq \hat{\Omega}_\lambda(D) < \hat{\Omega}_\lambda(z^*) - \varepsilon^* \leq \hat{W}_\lambda(z^*).$$

We then have, for $\lambda < \lambda^0$,

$$D_{\hat{W}}(\lambda) \in [z^* - \nu(\lambda), z^* + \nu(\lambda)]\,.$$

Letting $\lambda \to 0$, we see that $D_{\hat{W}}(\lambda) \to z^*$. $\square$