NBER WORKING PAPER SERIES

OPTIMAL DEFAULT OPTIONS

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 February 2024

We are grateful to Eric Gao for able research assistance. 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
B. Douglas Bernheim and Jonas Mueller Gastell
NBER Working Paper No. 28254
December 2020, Revised February 2024
JEL No. D10,D11,D14

ABSTRACT

Previous studies of optimal default options demonstrate that either opt-out minimization or maximization is optimal under restrictive conditions. We obtain a general characterization of the solution by studying optimal defaults when one of the problem's parameters approaches a limiting value. We interpret these "asymptotic optima" as approximate optima for non-limiting cases and justify this interpretation through numerical simulations. When the designer and choosers agree about the activity's value, simple forms of weighted opt-out minimization are asymptotically optimal. Additional results encompass Pigouvian fees, normative ambiguity, and cases in which the designer and choosers disagree about the activity's value.

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

1 Introduction

Most decision problems implicitly or explicitly specify an option that serves as a default, in the following sense: if the individual 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). Because different people have different preferences, any given default may be appropriate for some and inappropriate for others. The literature has addressed the problem of choosing the optimal default primarily in the context of contribution rates for 401(k) pension plans, where a collection of empirical studies, starting with Madrian and Shea (2001), have revealed that changing the default option has a powerful effect on employees' contributions.[1] 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).

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. While their justification for the criterion is informal, the subsequent literature examines this issue more rigorously. Unfortunately, the general mathematical conditions that characterize optimal defaults do not have useful economic interpretations. Existing theoretical studies therefore focus on identifying restrictive sufficient conditions under which the Thaler-Sunstein rule is optimal; see especially Carroll et al. (2009) and Goldin and Reck (2022). Niether 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 the opposite rule, opt-out maximization, can also be optimal. And yet, analyses of empirically parametrized models suggest that opt-out minimization may be attractive more generally for reasons the theoretical literature has not identified. Specifically, based on numerical simulations, Bernheim, Fradkin, and Popov (2015) conclude that "the Thaler Sunstein opt-out-minimization criterion yields small welfare losses even when it is suboptimal; hence it is a reasonable rule of thumb" (see also Choukhmane 2023).

In this paper, we adopt an alternative methodological approach that allows us to characterize "appoximately" optimal defaults with much greater generality. The essence of our strategy is to consider a general version of the optimization problem, and to study the properties of its solution as one of the problem's parameters approaches a limiting value. The parameter in question governs the distribution of opt-out costs, and the limiting value yields a degenerate distribution (zero costs). Because the problem's mathematical structure simplifies as the parameter approaches this limit, we are able to provide a general characterization of "asymptotically optimal" defaults. We interpret this characterization as describing approximately optimal solutions for non-limiting opt-out cost distributions, and we justify this interpretation through numerical simulations. Conceptually, this

[1] See Beshears et al. (2018) for a summary of the subsequent literature. More recent evidence suggests that, for many workers, the effects may be transient (Choukhmane 2023).

approach is related to familiar strategies used in asymptotic econometrics,[2] as well as in computer science.[3]

We begin by analyzing settings with two distinguishing characteristics: first, the distribution of "ideal points" across heterogeneous individuals (i.e., the options they would actively select) is atomless; second, the designer's objectives concerning each individual's outcome coincide with the individual's objectives. Because the existing studies of optimal defaults impose these "baseline" assumptions, they provide a natural starting point for our analysis.

Certain applications fit the baseline assumptions well. For example, the manufacturer of a television screen establishes a default setting for brightness, but a customer can opt out by locating the control and altering the setting continuously. Likewise, when preparing a salad, a chef at a restaurant establishes a default for the quantity of salad dressing, but a customer can opt out by requesting dressing on the side and adding it to taste.

Other applications fit the baseline assumptions less well. In some cases, options are discrete and sparse. Examples include selecting from a short list of employer-sponsored health care plans or specifying whether one wishes to be a potential organ donor on a driver's license application. In other cases, atoms in the ideal-point distribution naturally appear at the boundaries of the opportunity set, or at special points on the interior of that set, such as "kink" points associated with caps on 401(k) contributions eligible for an employer match. The objectives of the designer and the chooser concerning the chooser's selection may also diverge. One possibility is that the designer's attitude toward the chooser is paternalistic. For example, an employer may worry that employees do not save enough for their own good. Another possibility is that the chooser's selection may create externalities. External cost and benefits may be direct, as when someone participates in an organ donation program. They may also arise indirectly by impacting market equilibria, as when improved individual choices exacerbate adverse selection problems (Handel 2013) or impact firms' pricing decisions (Ericson 2020).

The paper's first contribution is to provide a general characterization of asymptotically optimal defaults under the baseline assumptions. 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 with respect to their ideal points, as-if opt-out costs, 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-in maximization yields approximately optimal outcomes when opt-in frequencies are weighted according to the workers' characteristics,

[2] Absent restrictive assumptions, the small-sample properties of common estimators are often difficult to characterize, and the problem of determining the best (most efficient) estimator can be intractible. However, the mathematics often become tractable as the number of observations grows without bound. In that limit, one can determine, for example, that maximum is asymptotically efficient, in the sense that it achieves the Cramer-Rao lower bound (see, e.g., Greene 2003, Theorem 17.1, p. 473). To draw inferences about finite samples, economists typically rely on such asymptotic approximations, often supported by numerical simulations.

[3] In computer science, an algorithm is asymptotically optimal if (loosely), as the inputs become arbitrarily large, performance is improvable by no more than a fixed factor.

with weights given by a fixed formula.

Specifically, the weight on any worker's decision reflects two characteristics: the ratio of the worker's opt-out costs to the population mean (η), and the degree to which the worker over-remphasizes the importance of opt-out costs, relative to subsequent consequences, when making decisions ($1/\beta$).[4] The formula for the weight is simply $\omega(\eta, \beta) \equiv \eta \left(1 - \frac{1}{3\beta}\right)$. The roles of η and β are intuitive. Avoiding opt-out yields larger benefits for those with higher values of η. However, there is an offsetting cost: those who opt in end up with outcomes that are further from their ideal points. When workers' perceptions of opt-out costs are unbiased ($\beta = 1$), the benefit to each opt-in necessarily exceeds their cost, and a simple geometric construction shows that, asymptotically, the average cost is one-third the average benefit. When workers exaggerate opt-out costs ($\beta < 1$), the ideal points of the opt-ins are further on average from the default, so the offsetting cost is larger, as the formula implies. If the bias is large enough to inflate the average cost-offset by a factor of three or more, the cost exceeds the benefit, so the weight becomes negative.

Next, we use our general characterization to demonstrate that the Thaler-Sustein rule (*unweighted* opt-out minimization) is asymptotically optimal as long as ideal points and utility curvature parameters are distributed independently of opt-out costs and biases (η and β). When this independence condition is satisfied, the integral representing the weighted average of the asymptotic opt-in frequencies factors in a way that guarantees fixed proportionality between the weighted and unweighted averages. In effect, the assumption guarantees that those who opt in at any given default receive neither greater nor lesser weight than those who opt in at any other default, which means the weights simply rescale the entire problem. As a result, unweighted opt-out minimization inherits the asymptotic optimality property from weighted opt-out minimization. Moreover, under the alternative assumption that ideal points are distributed independently of the continuation value function's curvature properties as well as opt-out costs and biases, the asymptotically optimal default is simply the modal ideal point.

Notably, our general characterization 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. Specifically, an inspection of the formula for $\omega(\eta, \beta)$ reveals that, if pervasive biases are small ($\beta > \frac{1}{3}$), the weights on opt-in rates are positive, so opt-out minimization is asymptotically optimal. However, if these biases are large ($\beta < \frac{1}{3}$), the weights are negative, in which case opt-out maximization is asymptotically optimal.

After establishing these results, we relax the two baseline assumptions. First we consider the possibility that the distribution of ideal points may include atoms, either because ideal points "bunch" at certain options (e.g., boundary points and kink points), or because opportunity sets are discrete and sparse. We show that our general characterization extends to these settings, with a small adjustment: the opt-in weighting function that ensures the asymptotic optimality of weighted opt-out minimization 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

[4] For example, if the worker's preferences are quasi-hyperbolic, β is the "present bias" parameter.

strategy necessarily has the flavor of opt-out minimization rather than maximization, even when biases are large.

Next, we consider the possibility that the designer and the choosers have divergent objectives concerning the activity's level, and that the designer cannot resolve this conflict by imposing Pigouvian fees. To simplify the analytics, we make the assumptions that guarantee asymptotic optimality of the modal ideal point in the baseline model, and we add a positive linear externality. We demonstrate that the optimum is achievable asymptotically by maximizing the weighted sum of the ideal-point distribution's density and the negative of its first derivative. When nearly all weight falls on the density (as is the case when β exceeds $\frac{1}{3}$ and externalities are small), unweighted opt-out minimization is approximately optimal. But when nearly all weight falls on the density's derviative (as is the case when externalities are large and workers' loss functions are symmetric), setting the default at the distribution's upper inflection point is approximately optimal. For intermediate cases, whether the asymptotic optimum is closer to the mode or the upper inflection point depends on the magnitude of externalities, as well as the direction and degree of loss-function asymmetry.

We use numerical methods to explore the accuracy of our asymptotic approximations. Our simulations show that the limiting case provides a decent guide for a range of reasonable settings with substantial 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.

Our findings have substantive implications for the selection of default options in practical settings. The implications differ according to the designer's objectives (most notably, whether the choosers' selections generate externalities), whether the designer can deploy other instruments (such as Pigouvian fees), certain observable features of the setting (such as whether the choice set is sparse or approximately continuous), and the nature of the data the designer can access or generate. However, our analysis suggests that a few rules of thumb perform reasonably well in practice, and it offers guidance for data collection and directional tweaks that may improve that performance.

The remainder of the paper proceeds as follows. Section 2 details the baseline model. Section 3 demonstrates the asymptotic optimality of weighted opt-out minimization, identifies conditions under which unweighted opt-out minimization also coincides asymptotically with welfare maximization, and explains how our analysis applies to settings with normative ambiguity (as in the welfare framework of Bernheim and Rangel 2009). It also provides a more detailed account of the relationships between our findings and existing results. Section 4 examines extensions to settings with bunching (arising, for example, from boundary constraints or caps on matching provisions), and to decisions with sparse opportunity sets. Section 5 examines cases in which designers disagree with choosers' ideal points (for example, due to externalities). Section 6 describes our numerical simulations. Section 7 summarizes the implications of our analysis for the selection of defaults in practice. We close in Section 8 with some brief thoughts about directions for subsequent research.

Abbreviated proofs appear in the Appendix.

2 The baseline model

This section outlines our baseline model, in which the distribution of "ideal points" across heterogeneous individuals is atomless, and the designer's objectives concerning each individual's outcome coincide with the individual's objectives. 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-based 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 $X = [\underline{x}, \overline{x}]$. 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 opt-out ($\gamma I(x \neq D)$, where $I(x \neq D) = 1$ if $x \neq D$ and 0 otherwise):

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

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 pertaining the the evaluation of opt-out costs. We elaborate on the interpretation of this parameter below when discussing the employer's objectives.[5] 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; Goldin and Reck 2022), our analysis presupposes that opt-out costs reflect effort the worker must

[5] 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 5, but would otherwise leave our results unchanged.

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 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.[6]

For one of our results (Proposition 5), we assume that, in addition to specifying a default contribution rate D, the plan may also 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. 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 result on optimal fines, like other results in this literature, are applicable only in settings where implementation rather than deliberation is costly.

The worker hence chooses x to maximize $u(x, M - I(x = D)K; x^*, \rho, \beta, \lambda\eta, D)$, where M represents compensation received from the firm. 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 $C_\lambda(D, x^*, \theta) = 0$ 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

[6] 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.

$\underline{\beta}, \underline{\eta} > 0$. Let $G(\theta)$ denote the CDF governing the marginal distribution of θ across workers, and $F(x^* \mid \theta)$ denote the CDF governing the distribution of x^* conditional on θ.

Throughout, we impose minimal restrictions on V:

Assumption 1. *For all $(x, x^* \rho) \in X^2 \times [\underline{\rho}, \overline{\rho}]$, (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. It ensures that those who prefer higher contributions also benefit more (or suffer less) on the margin as contributions increase from any specified level. 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.

Without loss of generality, we normalize the total population size to unity ($\int_\Theta dG(\theta) = 1$). Except where stated otherwise, 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.*

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 a default $D \in [\underline{D}, \overline{D}]$ and, when permitted, the fine K that applies uniformly to everyone. To avoid some technical complications that arise for defaults at the boundaries of workers' opportunity sets, we assume that $\underline{D} > \underline{x}$ and $\overline{D} < \overline{x}$. We describe the reason for this restriction below, and separately explain how the ability to set defaults at the boundaries of X would affect our results.

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. We will assume that this disagreement is limited to the normatively appropriate value of β, which she takes to be unity. Thus, to 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.[7] For other interpretations of β, see Bernheim, Fradkin, and Popov (2015) and Goldin and Reck (2022).

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

A key feature of our baseline 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.[8] Whether this assumption is reasonable depends on the context; we relax it in Section 5. 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.

When imposing a fine for passive choice, the employer earns revenue B per worker. We assume the employer is indifferent, on the margin, between taking incremental revenue as profit and using it to enhance profits indirectly by increasing employee compensation.[9]Under this assumption, we can, without loss of generality, treat the employer as using the revenue to pay a lump sum bonus of B per worker. For simplicity, the employer levies fines and disburses these 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. Because utility is linear in income, the level of the worker's baseline compensation (before fines and bonuses) is immaterial, so we take it to be zero (in other words, $M = B$). When the employer chooses D and K, the resulting value of B is given by:

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

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 basic results

In this section, we characterize optimal defaults under the baseline assumptions. We start by establishing the asymptotic optimality of weighted opt-out minimization (Section 3.1). Then we derive conditions under which weighted opt-out minimization coincides with (i) the Thaler-Sunstein

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

[9]As long as the employer's profits are differentiable in the employee's compensation, this indifference is a consequence of profit maximization.

rule (unweighted opt-out minimization), and (ii) selecting the mode of the ideal-point distribution (Section 3.2). We also explain how our analysis accommodates any normative ambiguity stemming from possible controversy concerning the interpretation of $\beta < 1$ as a "bias" (Section 3.3). We close this section by clarifying the relationship between our findings and existing results (Section 3.4).

3.1 The asymptotic optimality of weighted opt-out minimization

In this section, we provide a general characterization of asymptotically optimal defaults that establishes a connection between welfare maximization and weighted opt-out minimization. Our main results show that, as $\lambda \to 0$, rescaled versions of aggregate welfare and of the weighted opt-out frequency (using specified weights) both converge uniformly to the same function, and consequently the defaults that maximize those functions also converge. To be clear, the same result does not necessarily hold for large λ,[10] which is why we verify the quality of the asymptotic approximation for plausible parameter values through simulations in Section 6. Throughout this section, we confine attention to settings in which the employer does not have the ability to impose fines for passive choice (i.e., we fix $B = 0$).

Because the opt-in 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,

[10] As $\lambda \to \infty$, only a trivial fraction of the population opts out and, critically, those individuals have small opt-out costs. Consequently, their total contribution to social welfare goes to zero. It follows that, in the limit, the employer sets the default to zero-out the average marginal utility of contributions for the entire population. That criterion does not generally coincide with weighted opt-out minimization.

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

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 $-\frac{1}{2}V_{11}(D,D,\rho)\,(D-x^*)^2 - \frac{\lambda\eta}{\beta}$ 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 $2\left(\frac{\lambda\eta}{\beta}\right)^{\frac{1}{2}}\left(\frac{1}{-\frac{1}{2}V_{11}(D,D,\rho)}\right)^{\frac{1}{2}}$.[11] 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. *Under Assumptions 1-3, the weighted opt-out-minimizing default option $D_\Omega(\lambda)$ converges to D^* as $\lambda \to 0$.*

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

[11] With a default of either $\underline{x}$ or $\overline{x}$, only half of this interval is relevant even when λ is arbitrarily small. Consequently, at these boundary points, the asymptotic opt-out frequency is half as a large as our approximation implies. Taking $\underline{D} > \underline{x}$ and $\overline{D} < \overline{x}$ avoids this technical complication.

Proposition 2. *Under Assumptions 1-3, 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:

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

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 $\lambda \to 0$.

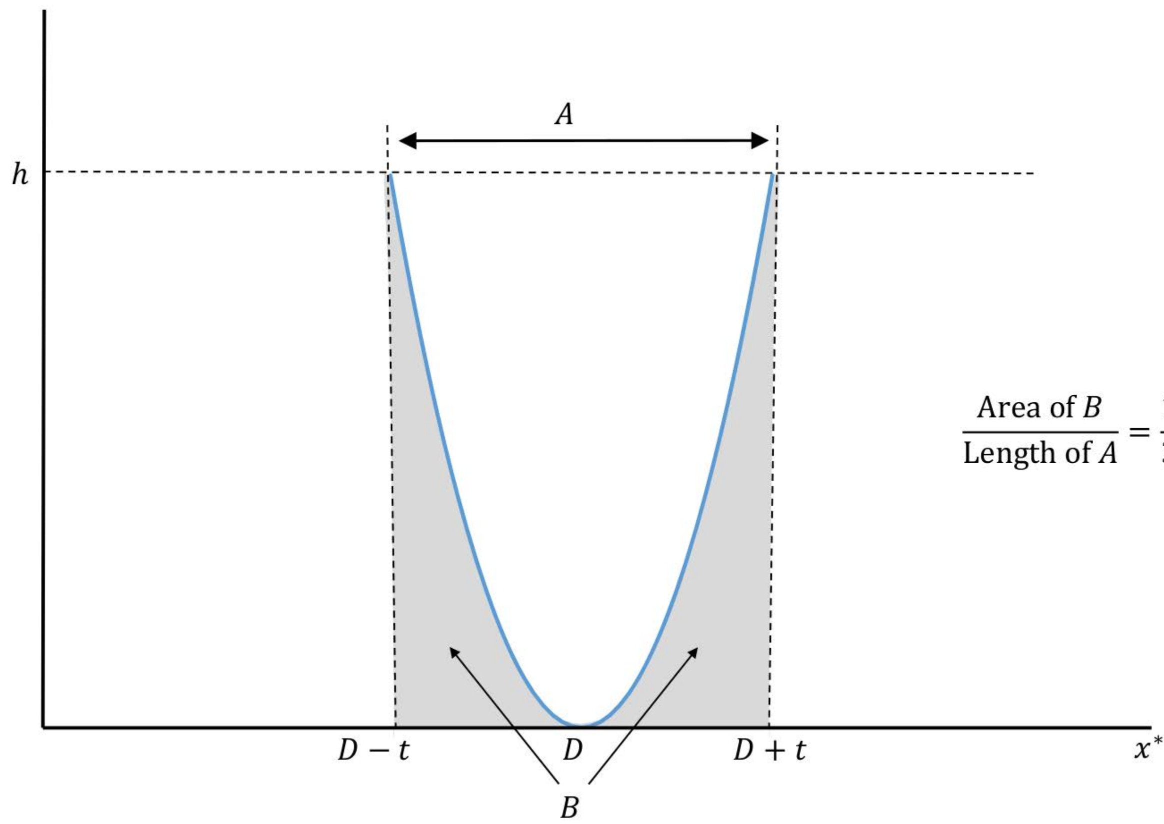
Here we encounter a small technical complication: as $\lambda \to 0$, the function L_λ converges to the function L_0, which maps all default rates to the same value. That property renders the limiting optimization problem unenlightening. To learn about optimization with small λ from the limiting case, 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

Figure 1: Second-order approximation to the conditional expectation



(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 $E(\cdot)$ term in equation (4) 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 quantity that is independent of D—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 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.

So far, we have assumed that $\underline{D} > \underline{x}$ and $\overline{D} < \overline{x}$. As noted in footnote 11, with an atomless ideal-point distribution, our approximation understates opt-out frequencies at boundary points. Setting the default at either boundary is therefore less attractive than our approximation would imply when the opt-out is detrimental in the aggregate (e.g., with $\underline{\beta} > \frac{1}{3}$), but more attractive and potentially optimal when opt-out is beneficial in the aggregate (e.g., with $\overline{\beta} < \frac{1}{3}$).

3.2 The asymptotic 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, $D_\Omega^U(\lambda) \equiv \min_{D \in X} \Omega_\lambda^U(D)$, and the opt-out maximizing default option, $d_\Omega^U(\lambda) \equiv \max_{D \in X} \Omega_\lambda^U(D)$. 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. The second route is to identify conditions under which the weighted opt-in frequency is separable into a component encompassing 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 small λ. Below, we consider two alternative conditions under which $\Pr_{x^*|\theta}\left[\Delta(D, x, \rho) \leq \frac{\lambda\eta}{\beta}\right]$ factors as desired in the limit.

Condition 1: (x^*, ρ) is distributed independently of (β, η). Recall that the (rescaled) weighted opt-in frequency, $\Omega_\lambda(D)$, converges uniformly to $\Omega(D)$, which we can write as follows:

$$\Omega(D) = \int_\theta \eta \left(1 - \frac{1}{3\beta}\right) Q(D, \theta) dG(\theta) \tag{5}$$

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

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

$$\Phi = \int_{\eta,\beta} \eta \left(1 - \frac{1}{3\beta}\right) \left(\frac{\eta}{\beta}\right)^{\frac{1}{2}} h(\eta, \beta) d\eta d\beta. \tag{6}$$

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

$$\pi \times \Omega^U(D) = \Omega(D), \tag{7}$$

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. *Suppose Assumptions 1-3 and Condition 1 hold. If $\Phi > 0$, then the unweighted opt-out-minimizing default option $D_{\Omega}^{U}(\lambda)$ 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 $\lambda \to 0$.*

Condition 2: x^* is distributed independently of θ, and x^* translates V. If we assume that x^* is distributed independent of θ (so that $f(D \mid \theta) = f(D)$), then we can rewrite equation (5) as follows:

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

Notice that the integral potentially depends on D only through the term $V_{11}(D, D, \rho)$. Even that dependence vanishes if we assume in addition that $V(x, x^*, \rho) = v(x - x^*, \rho)$ for some function v—in other words, that variations in x^* translate the function V. With this restriction, deviating from the worker's ideal point by a fixed amount inflicts a utility penalty that does not depend on the ideal point. In that case, we have $V_{11}(D, D, \rho) = v_{11}(0, \rho)$. It follows that

$$\Omega(D) = \Psi f(D)$$

where

$$\Psi_1 = \int_{\theta} \eta \left(1 - \frac{1}{3\beta}\right) \left(\frac{\eta}{\beta}\right)^{\frac{1}{2}} \left(\frac{1}{-\frac{1}{2}v_{11}(0, \rho)}\right)^{\frac{1}{2}} dG(\theta)$$

Trivially, D^* is then the ideal point that either maximizes or minimizes $f(D)$, depending on whether Ψ_1 is positive or negative. Consequently, by Proposition 2, we know that the welfare maximizing default option converges to the mode of the ideal point distribution if $\Psi_1 > 0$, and to a point of minimum density (the antimode) if $\Psi_1 < 0$.[12]

To characterize the default that minimizes the unweighted opt-out frequency, we again invoke the uniform convergence of $\Omega_{\lambda}^{U}(D)$ to $\Omega^{U}(D)$. Under Condition 2, we can rewrite the latter function as follows:

$$\Omega^{U}(D) = \psi f(D)$$

[12] For simplicity, we also take the antimode to be unique.

where

$$\psi = \int_\theta \left(\frac{\eta}{\beta}\right)^{\frac{1}{2}} \left(\frac{1}{-\frac{1}{2}v_{11}(0,\rho)}\right)^{\frac{1}{2}} dG(\theta)$$

Accordingly, the unweighted opt-out-minimizing default option also converges to the mode of the ideal point distribution, and the unweighted opt-out-maximizing default option converges to the antimode.

The following proposition summarizes these observations:

Proposition 4. *Suppose Assumptions 1-3 and Condition 2 hold. If $\Psi_1 > 0$, then the welfare-maximizing and unweighted opt-out-minimizing default options both converge to the mode of the ideal point distribution as $\lambda \to 0$. If $\Psi_1 < 0$, then the welfare-maximizing and unweighted opt-out-maximizing default options both converge to the antimode as $\lambda \to 0$.*

Fines for passive choice. Propositions 3 and 4 both point to opt-out minimization when the distribution of β is concentrated above $\frac{1}{3}$, and to opt-out maximization when the distribution is concentrated below $\frac{1}{3}$. However, to the extent the employer can impose non-disappative fines for passive choice, opt-out minimization becomes relatively more attractive.

We establish the preceding claim under Condition 1. (The analysis for Condition 2 is similar.) To allow for the possibility that the optimal fine shrinks along with the scale of opt-out costs, we write it as $K(\lambda)$. 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 (weakly) 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.

Proposition 5. *Suppose Assumptions 1-3 and Condition 1 hold, and assume Θ is degenerate. Fixing D, the optimal fine is $K^* = (1-\beta)\gamma$.*

The intuition for Proposition 5 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 the bias is large or small.

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 convenient as a practical matter, 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.

3.3 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 Bernheim (2009; 2016; 2021) 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 unambuouosly 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) provide 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

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

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

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.

3.4 Relation to existing results

The previous literature contains a few theoretical results related to our Propositions 3 and 4. Goldin and Reck (2020) study settings in which worker heterogeneity is limited to ideal points and as-if opt-out costs (our x^* and η). They focus on settings that satisfy our baseline assumptions and establish the optimal of unweighted opt-out minimization under the following sufficient conditions: (i-a) ideal points are distributed independently of opt-out costs; (i-b) the distribution of ideal points is single-peaked and symmetric; (ii-a) the utility derived from the action depends only on the difference between the action and the worker's ideal point; (ii-b) utility is a convex, single-peaked, symmetric function that difference; and (iii) as-if and normative opt-out costs are sufficiently similar. Furthermore, unweighted opt-out maximization becomes optimal when as-if and normative opt-out costs are sufficiently different. 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 and maximization are similar to those in Goldin and Reck (2020).

Our focus on asymptotic optimality rather than exact optimality allows us to treat the baseline problem with greater generality and to establish the desirability of unweighted opt-out minimization under substantially weaker sufficient conditions (Propositions 3 and 4). With respect to distributional assumptions, we allow for two additional dimensions of worker heterogeneity (bias and utility function curvature). Moreover, while Conditions 1 and 2 both impose (i-a), neither requires (i-b). With respect to assumptions concerning utility functions, Condition 2 retains (ii-a) while while Condition 1 dispenses with it; both replace (ii-b) with a modest single-crossing requirement. We also retain (iii), but identify a precise quantitative threshold (specifically, $\beta > \frac{1}{3}$, i.e., cognitive biases inflate as-if opt-out costs relative to continuation utility by a factor of no more than three). Additionally, our Proposition 5 extends the literature by showing, in effect, that one can dispense with (iii) in settings where the employer can impose financial penalties for passive choice.[13]

The existing literature also contains no counterpart to our Proposition 2, which provides a general characterization of asymptotically optimal defaults for settings that violate the Golden-Reck assumptions. The characterization involves a suprisingly simple form of weighted opt-in

[13] Bernheim, Fradkin, and Popov make a related point in arguing that the optimality of extreme and 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. Although they do not provide a theoretical result, their numerical calculations imply that 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, and by addressing the question theoretically rather than computationally.

maximization with weights $\omega(\eta, \beta) = \eta\left(1 - \frac{1}{3\beta}\right)$. Our analysis allows us to understand unweighted opt-out minimization as a special case of this general rule, one that emerges when the integral of the opt-in frequencies factors appropriately. The general characterization also helps to explain the previous computational finding that unweighted opt-out minimization yields small welfare losses even when it is suboptimal (Bernheim, Fradkin, and Popov 2015, and Choukhmane 2023): as long as correlations between x^* and $\omega(\eta, \beta)$ are not too large, one would not expect weighting to make much of a difference. Finally, because the sign of ω depends on the magnitude of β, the general characterization reveals that results concerning the optimality of the two opposing rules highlighted in the literature—unweighted opt-out minimization and maximization—are two sides of the same coin.

Goldin and Reck (2020) also address the topic of normative ambuity. While their analysis is related to ours, it is also distinct. The distinctions begin with different conceptualizations of normative ambiguity. Goldin and Reck proceed from the premise that people have "true" preferences (in the conventional sense), and consequently (to use our notation) that there is a single normatively correct value of β, call it β^T. In their view, ambiguity arises because the analyst may not know β^T. By focusing on the polar cases (one in which the perceived costs of opt-out equal their "true" costs, another in which the "true" costs are zero), they emphasize that one cannot tell whether employers should maximize or minimize opt-out until one nails down the value of β^T. In contrast, following Bernheim and Rangel (2009), we allow for the possibility that people construct their judgments contextually, in which case the notion of "true" preferences is an analytic fiction. Unlike Goldin and Reck, we therefore consider normative ambiguity potentially irreducible. From our perspective, the notable finding is that irreducible normative ambiguity need not impact the optimality of weighted opt-out minimization (or the formula for the weights) provided the values of β that govern naturalistically framed opt-out decisions exceed $\frac{1}{3}$. In contrast, if choices within the welfare-relevant domain are consistent with values of β that span $\frac{1}{3}$, then irreducible normative ambiguity may preclude discerning comparisons between minimization and maximization of the (weighted) opt-out frequency.

The previous theoretical literature concerning optimal defaults is also limited to settings that satisfy our two baseline assumptions: the distribution of x^* is atomless, and the designer's objectives concerning each individual's outcome coincide with the individual's objectives. In the next two sections, we extend the literature by relaxing those assumptions, using the same asymptotic approach as for the baseline model.

4 Optimal defaults with bunching and sparse menus

In this section, we relax the first of our two baseline assumptions by considering the possibility that the distribution of ideal points includes atoms. In some contexts, ideal points "bunch" at distinctive options. For example, with respect to 401(k) contributions, workers typically bunch at zero and at the option that exhausts their employer's matching contributions (due to the resulting kink point in the worker's opportunity set). In other contexts, the set of options is discrete and sparse. For example, employers typically present workers with a limited menu of health insurance plans. We show that our general characterization extends to these settings, with a small adjustment: the opt-in weighting function that ensures the asymptotic optimality of weighted opt-out minimization is simply $\omega(\eta) = \eta$. We explain this finding intuitively and reconcile it with our results for the baseline model.

4.1 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

$$\pi_n \equiv \int_\theta \Pr\left(x^*(y) = z_n \mid \theta\right) dG(\theta) = \int_\theta \left[F(\overline{y}_n \mid \theta) - F(\underline{y}_n \mid \theta)\right] dG(\theta). \tag{8}$$

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 (8) 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 $\lambda \to 0$.[14] 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 $\lambda \to 0$.

It is useful to rewrite the welfare function as follows:

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

$$
\begin{aligned}
\hat{W}_\lambda(D) &= \int_\theta \Pr\left(\Delta(D, x^*(y), \rho) \leq \tfrac{\lambda\eta}{\beta} \middle| \theta\right) \\
&\quad \times \left[\eta - \tfrac{1}{\lambda} E\left(\Delta(D, x^*(y), \rho) \mid \theta, \Delta(D, x^*(y), \rho) \leq \tfrac{\lambda\eta}{\beta}\right)\right] dG(\theta), \\
&= \hat{\Omega}_\lambda(D) - \int_\theta \Pr\left(\Delta(D, x^*(y), \rho) \leq \tfrac{\lambda\eta}{\beta} \middle| \theta\right) \\
&\quad \times \left[\tfrac{1}{\lambda} E\left(\Delta(D, x^*(y), \rho) \mid \theta, \Delta(D, x^*(y), \rho) \leq \tfrac{\lambda\eta}{\beta}\right)\right] dG(\theta)
\end{aligned} \tag{9}
$$

Now think about what happens to $\hat{W}_\lambda(D)$ as $\lambda \to 0$. For $D \notin Z$, $\Pr\left(\Delta(D, x^*(y), \rho) \leq \frac{\lambda\eta}{\beta} \middle| \theta\right) \to 0$, so the second term (specifically, everything after $\hat{\Omega}_\lambda(D)$) vanishes.[15] In contrast, for $D \in Z$, $\Pr\left(\Delta(D, x^*(y), \rho) \leq \frac{\lambda\eta}{\beta} \middle| \theta\right)$ need not vanish. The limiting behavior of the second term then depends on the bracketed expression in the last line. In Section 3.1, 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 z^*.

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

Proposition 6. *Suppose Assumptions 4 and 5 hold, and that z^* is unique. 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 $\lambda \to 0$.*

One important feature of this result is that the asymptotic optimum is always a contribution rate at which bunching occurs (formally, $z^* \in Z$). Another important feature is that the asymptotic optimum does not depend on the distribution of the bias parameter, β. In this context, because $\Pr\left(\Delta(D, x^*(y), \rho) \leq \frac{\lambda\eta}{\beta} \middle| \theta\right)$ converges to $\Pr\left(\Delta(D, x^*(y), \rho) = 0 \middle| \theta\right)$, the bias parameter can only enter in the limit through the bracketed term in the last line of equation (9). But as we have explained, that term disappears in the limit, which is why we arrive at the weight η rather than $\eta\left(1 - \frac{1}{3\beta}\right)$. Several implications follow.

First, because the weights η are strictly positive, the asymptotically optimal rule always has the flavor of opt-out minimization rather than maximization, even when bias is severe. In contrast to our baseline results, a fine for passive choice is not required for this property to hold.

Second, because the weights do not depend on β, weighted opt-out minimization is always normatively robust (in the sense that the asymptotic maximizers of EV_A and EV_B coincide).

Third, the asymptotic optimality of *unweighted* opt-out minimization only requires the independence of y (which stands in for x^*) and η. Relative to the corresponding results for the basline model (Propositions 3 and 4), we are able to dispense with the independence assumptions concerning β and ρ.

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

To understand this third point, notice that we can rewrite $\hat{\Omega}(z)$ as follows:

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

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

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.

Reconciling Propositions 2 and 6.

A comparison between Propositions 2 and 6 reveals an apparent tension: with no bunching, β appears in the weighting formula, but with even the tiniest amount of bunching, it vanishes. To underscore this tension, suppose Condition 1 is satisfied, and that $\beta < \frac{1}{3}$ for all workers. Assuming the distribution of x^* has a single atom, Proposition 6 implies that opt-out minimization is optimal. As we shrink the size of the atom toward zero, this implication is unchanged. But once we reach the limit and the atom disappears, Proposition 2 implies (through Proposition 3) that opt-out maximization is optimal.

The aforementioned tension reflects a general mathematical principle: in settings where multiple variables approach limiting values, the maximum of a well-behaved function can converge to different values depending on the order in which one takes the limits.[16] In the previous paragraph's example, the variables in question are λ and the size of the atom. When we invoke Proposition 6, we are taking the limit with respect to λ before shrinking the size of the atom. When we invoke Proposition 2, we are (in effect) taking limits in the opposite order. The term $\frac{1}{\lambda} E\left(\Delta(D, x^*(y), \rho) \mid \theta, \Delta(D, x^*(y), \rho) \leq \frac{\lambda\eta}{\beta}\right)$ converges to 0 in the first instance, but to $\frac{\eta}{3\beta}$ in the second.

Accounting for this tension mathematically leaves open the question of whether, in any given practical application, one ought to heed the implication of Proposition 2 or of Proposition 6. The answer to the question depends on which approximation is closer to the case of interest, the "small atom" limit or the "small opt-out cost" limit. That comparison boils down to whether, for $D \in Z$ (bunching points) and typical values of θ, the opt-in set $\left\{y \mid \Delta(D, x^*(y), \rho) \leq \frac{\lambda\eta}{\beta}\right\}$ is dominated by values for which $x^*(y) = D$ (in which case Proposition 6 provides the better approximation), or by values for which $x^*(y) \neq D$ (in which case Proposition 2 provides the better approximation).

Fortunately, this determination is relatively easy to make in practice. For each ideal point at

[16] For example, consider the function $f : \{0, 1\} \to \mathbb{R}$ such that $f(0) = \varsigma$ and $f(1) = \tau$. If we first take the limit of the function as $\varsigma \to 0$, and then take the limit as $\tau \to 0$, the sequence of maximizers converge to 1. If we take the limits in the opposite order, the sequence of maximizers converge to 0.

which bunching occurs ($z \in Z$), one can compute the ratio of the fraction of workers contributing z when D is far from z, to the fraction of workers contributing z when $D = z$. This ratio approximates the proportion of workers falling within the set $\left\{y \mid \Delta(z, x^*(y), \rho) \leq \frac{\lambda\eta}{\beta}\right\}$ for whom $x^*(y) = z$.[17]Let $\chi(z)$ denote that proportion. If $\chi(z)$ is close to unity for all $z \in Z$, bunching is relatively inconsequential for the analytics of optimal defaults, so Proposition 2 provides better guidance than Proposition 6. The opposite conclusion follows when $\chi(z)$ is close to zero for all $z \in Z$.

When $\chi(z)$ is close to neither zero nor unity for some $z \in Z$, it may be preferable to employ an asymptotic approximation for which λ and the size of each atom converge to zero simultaneously at rates that preserve $\chi(z)$.[18] With constant $\chi(z)$, we have[19]

$$\frac{1}{\lambda}E\left(\Delta(z, x^*(y), \rho) \mid \theta, \Delta(z, x^*(y), \rho) \leq \frac{\lambda\eta}{\beta}\right) = \chi\frac{1}{\lambda}E\left(\Delta(z, x^*(y), \rho) \mid \theta, z \neq \Delta(z, x^*(y), \rho) \leq \frac{\lambda\eta}{\beta}\right)$$

Furthermore, the right-hand-side expression converges in the aforementioned limit to $\chi\frac{\eta}{3\beta}$. Consequently, the asymptotically optimal rule is to maximize the weighted opt-in frequency using weights $\omega(\eta, \beta, D) = \eta\left(1 - \frac{\chi(D)}{3\beta}\right)$, where $\chi(D) \equiv 0$ for contribution rates at which ideal points do not bunch ($D \notin Z$). With a large degree of bunching, the asymptotic optimum tends to fall within Z because the associated opt-in frequency and weights are both larger than for options in the complement of Z. Propositions 2 and 6 thus emerge as boundary cases.

4.2 Sparse menus

To analyze environments with sparse menus, 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

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

[17]The ratio understates this proportion because, with any $D \neq z$, some of the workers with $x^*(y) = z$ will not make active choices. Using a default that is far from z when measuring the numerator minimizes the resulting bias.

[18]When taking this limit, we scale the objective function by $\lambda^{-\frac{3}{2}}$ as in Proposition 2, rather than by λ^{-1} as in Proposition 6.

[19]To obtain this expression, we use the fact that $\Delta(z, x^*(y), \rho) = 0$ iff $x^*(y) = z$.

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

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

As in the last subsection, we claim that the bracketed term converges to zero. Intuitively, 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 6. 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. The explanation is similar: the asymptotic optimum depends on the order in which one takes limits—in this context, whether one first lets increasingly fine discrete menus approach a continuum before or after taking the limit as λ goes to zero.

When considering any practical application, one must therefore determine whether the grid size is large or small relative to opt-out costs. If the grid size is small enough, the opt-in window for any default D will include workers with many other ideal points, and those with $x^* = D$ will consistute a small fraction of the total opt-ins. In that case, Proposition 2 provides a reasonable guide. Alternatively, if the the grid size is coarse enough, workers with $x^* = D$ will dominate the opt-ins. In that case, it is more appropriate to use the rule derived in the current section. Once again, one can evaluate these alternative possibilities by computing the ratio of the fraction of workers contributing x when D is far from x, to the fraction of workers contributing x when $D = x$.

5 Optimal defaults with divergent objectives

In this section, we relax the second of our two baseline assumptions by considering the possibility that the designer disagrees with the choosers concerning the activity's benefits or costs, and cannot resolve this disagreement by levying a Pigouvian fee. For the sake of analytic tractability, we focus primarily on settings where the discrepancy is linear in the activity's level—in other words, ones with linear externalities. Our formulation subsumes a variety of possibilities, including cases where the activity's overall level has desirable or undesirable general equilibrium effects, as in Handel (2013) and Ericson (2020). In keeping with some of the behavioral literature, one can alternatively interpret the discrepancy as reflecting an "internality" associated with cognitive biases regarding the level of the activity x (see, e.g., Bernheim and Taubinsky 2018).

Formally, we assume the employer evaluates the outcome associated with the choice of any given worker exactly as in Section 2, but also perceives an added benefit of σx, where σ parameterizes the importance of the externality. Accordingly, when the default is D, a worker with characteristics (x^*, θ) makes the following contribution to the employer's objective function:

$$\overline{U}_\lambda(D, x^*, \theta) = \tilde{U}_\lambda(D, x^*, \theta) + \sigma \left[x^* + (1 - C_\lambda(D, x^*, \theta))(D - x^*)\right],$$

where the functions $\tilde{U}_\lambda$ and C_λ are defined as before.

Throughout this section, we maintain Assumptions 1 and 2, and simplify by imposing Condition 2 (x^* is distributed independent of θ, and we can write $V(x, x^*, \rho)$ as $v(x - x^*, \rho)$). The employer's overall objective is then to maximize the following function:

$$\overline{L}_\lambda(D) = L_\lambda(D) + E(x^*) + \sigma \int_\theta \left[\int_{\Delta(D,x^*,\rho) \leq \frac{\lambda\eta}{\beta}} (D - x^*) dF(x^*) \right] dG(\theta)$$

We write the optimal default as $D_{\overline{L}}(\lambda)$.

As in Section 3, we encounter the difficulty that, as $\lambda \to 0$, the objective function becomes completely flat. To overcome this difficulty, we once again add a constant and rescale by $2\lambda^{-\frac{3}{2}}$:

$$\overline{W}_\lambda(D) \equiv W_\lambda(D) + J_\lambda(D).$$

where

$$J_\lambda(D) \equiv \frac{\sigma}{2\lambda^{\frac{3}{2}}} \int_\theta \left[\int_{\Delta(D,x^*,\rho) \leq \frac{\lambda\eta}{\beta}} (D - x^*) dF(x^*) \right] dG(\theta)$$

In establishing Proposition 4, we showed that under Condition 2, as $\lambda \to 0$, $W_\lambda(D)$ converges uniformly to $\Psi_1 f(D)$ for some fixed constant Ψ_1. The constant may be positive or negative depending on the distribution of β. The next step is to provide a similar asymptotic characterization of $J_\lambda(D)$. It turns out that $J_\lambda(D)$ converges uniformly to $-\sigma \left[v_{111}(0, \rho)\Psi_2 f(D) + \Psi_3 f'(D) \right]$, where Ψ_2 and Ψ_3 are strictly positive constants. Thus, the transformed objective function converges uniformly to $(\Psi_1 - \sigma v_{111}(0, \rho)\Psi_2) f(D) - \sigma\Psi_3 f'(D)$.

We define the optimum for the limiting case as follows:

$$\overline{D}^* \equiv \arg\max_{D \in X} \left[(\Psi_1 - \sigma v_{111}(0, \rho)\Psi_2) f(D) - \sigma\Psi_3 f'(D) \right],$$

As before, cases with multiple maxima are non-generic and therefore of little interest, so we rule them out by assuming that $\overline{D}^*$ is unique. If we strengthen part (i) of Assumption 1 to encompass fourth derivatives, we then have:

Proposition 7. *Suppose Assumptions 1-2 and Condition 2 hold, $v_{1111}(x - x^*, \rho)$ is continuous, and $\overline{D}^*$ is unique. Then the welfare-maximizing default option $\overline{D}_L(\lambda)$ converges to $\overline{D}^*$ as $\lambda \to 0$.*

Proposition 7 allows us to understand Proposition 4 as one end of a spectrum: when externalities are unimportant (σ is close to zero), the asymptotically optimal default roughly coincides with the option that maximizes the ideal-point density—in other words, the distribution's mode. At the other end of the spectrum, with sufficiently large σ, externalities are so important that $J_\lambda(D)$ dwarfs $W_\lambda(D)$. If, in addition, $v_{111}(0, \rho) = 0$ (so that, roughly speaking, opt-in windows are symmetric), the asymptotically optimal default roughly coincides with the option at which the ideal-point density has the greatest downward slope—in other words, an inflection point of the ideal-point distribution. For intermediate values of σ, the solution maximizes a weighted average of the density and the (negative) slope of the density, where the weight attached to the density is greater when opt-in windows are left-skewed ($v_{111}(0,\rho) < 0$) and smaller when they are right-skewed ($v_{111}(0,\rho) > 0$).

For distributions with common shapes, we can say more about the optimal default. A distribution is *quasi-bell-shaped* if the density function is three-times differentiable, single-peaked at some x_m, and has two inflection points, x_h and x_l, one above and one below the peak, with $f'''(x) < 0$ for $x \in (x_l, x_m)$ and $f'''(x) > 0$ for $x \in (x_m, x_h)$. For cases in which biases in assessing opt-out costs are not too severe (i.e., the distribution of β is mostly above $\frac{1}{3}$, so that $\Psi_1 > 0$) and opt-in windows are weakly left-skewed ($v_{111}(0,\rho) \leq 0$) we reach the following conclusions:[20]

Proposition 8. *Suppose Assumptions 1-2 and Condition 2 hold, $v_{1111}(x - x^*, \rho)$ is continuous, f is quasi-bell-shaped, $\Psi_1 > 0$, and $v_{111}(0,\rho) \leq 0$. For all $\sigma \geq 0$, the asymptotically optimal default option $\overline{D}^*$ is unique and lies (weakly) between the mode and the upper inflection point of the ideal-point distribution. It equals the mode when $\sigma = 0$ and rises monotonically with σ. When $v_{111}(0,\rho) = 0$, it converges to the upper inflection point as $\sigma \to \infty$.*

When $\Psi_1 \leq 0$ and $v_{111}(0,\rho) < 0$, the same result holds for sufficiently large σ. When $v_{111}(0,\rho) > 0$, the weight on $f(D)$ is negative for large σ (as well as for small σ if $\Psi_1 \leq 0$). Consequently, the asymptotic optimum lies either in the upper tail of the opt-out distribution beyond the upper inflection point, or in the lower tail.

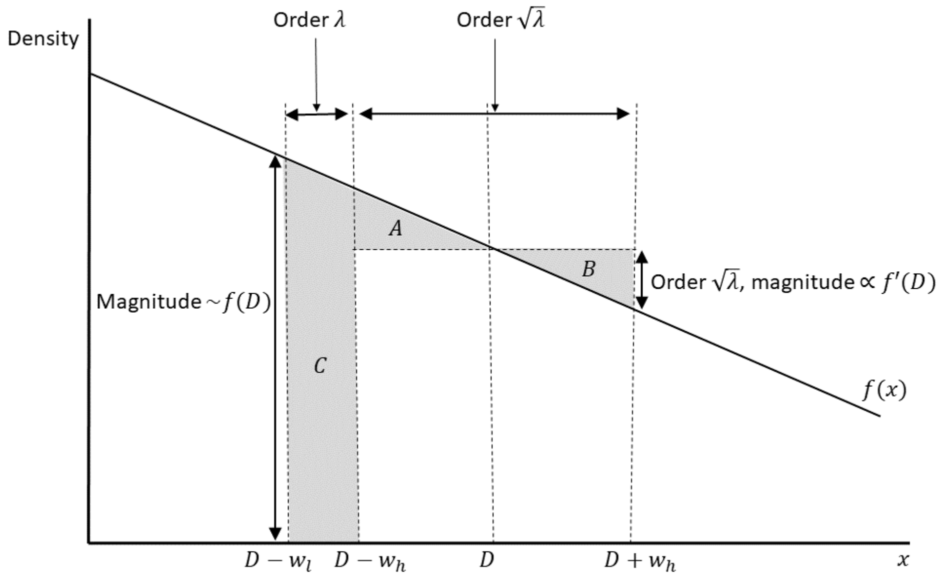
For the special case of a normal ideal-point distribution, under the conditions stated in Proposition 8, the asymptotically optimal default lies above, but not more than one standard deviation from, the mean. Equivalently, it falls between the 50th and 84th percentiles of the ideal-point distribution. When externalities are incidental, it falls close to the 50th percentile, but when they loom large, it approaches the 84th percentile if opt-in windows are symmetric.

Figure 2 provides intuition for the asymptotic characterization of $J_\lambda(D)$. The interval $[D - w_l, D + w_h]$ represents the opt-out window associated with the default D. The translation assumption ensures that w_l and w_h do not depend on D. For the purpose of this discussion, we assume the opt-in window is left-skewed ($w_l > w_h$). The figure separates the window into a symmetric part and an asymmetric part. The symmetric part, $[D - w_h, D + w_h]$, is centered on D. The asymmetric part, $[D - w_l, D - w_h]$, is entirely below D.

First, consider the symmetric part. When the designer sets a default D, any chooser with ideal points in the opt-in window end up with D rather than x^*. For choosers with $x^* \in [D - w_h, D]$, this effect yields an external benefit of $\sigma(D - x^*)$; for choosers with $x^* \in [D, D + w_h]$, it yields an external cost of $\sigma(x^* - D)$. The mass of people for whom there is an external benefit corresponds to the area beneath $f(x)$ over $[D - w_h, D]$, and the mass of people for whom there is an external cost corresponds to the area beneath $f(x)$ over $[D, D + w_h]$. Moreover, if we consider an individual from the first subinterval and one from the second with ideal points equidistant from D, the external benefits and costs just offset. Consequently, the net gain depends on the difference between the

[20] The third-derivative condition ensures uniqueness of the maximizer; the other properties hold without it.

Figure 2: The asymptotic characterization of $J_\lambda(D)$



population mass associated with the two subintervals. That difference is given by the sum of the two shaded areas, A and B, shown in the figure. The area of $A \cup B$ depends on the width of $[D - w_h, D + w_h]$ and the magnitude of $f(D - w_h) - f(D + w_h)$. The width is of order $\sqrt{\lambda}$. The term $f(D - w_h) - f(D + w_h)$ is therefore of order $\sqrt{\lambda}$ and proportional to $-f'(D)$. It follows that the area of $A \cup B$ is of order $\left(\sqrt{\lambda}\right)^2 = \lambda$ and proportional to $-f'(D)$. Furthermore, because the external benefit for someone with $x^* \in [D - w_h, 0]$ is $\sigma (D - x^*)$, the average external benefit is of order $\sqrt{\lambda}$. The total net benefit is therefore of order $\lambda^{3/2}$ and proportional to $f'(D)$. Thus, the symmetric part of the interval accounts for the $-\Psi_3 f'(D)$ term.

Next, consider the asymmetric part of the opt-in window, $[D - w_l, D - w_h]$. For those in this interval, selecting D rather than x^* generates an external benefit of $\sigma(D - x^*)$. The mass of people in this interval is given by the shaded area C. Here, there are no choosers for whom setting the default D creates offsetting costs. The height of C is approximately $f(D)$. The width depends on $v_{111}(0, \rho)$. If $v_{111}(0, \rho) = 0$, the width is too small to matter. Even when $v_{111}(0, \rho) \neq 0$, as $\lambda \to 0$, the the asymmetric part of the opt-in window becomes vanishingly small relative to the symmetric part. Specifically, the proof of the proposition shows that $w_l - w_h$ is of order λ. It then follows that the area C is of order λ and proportional to $f(D)$. Furthermore, because the external benefit for someone with $x^* \in [D - w_l, D - w_h]$ is $\sigma (D - x^*)$, the average external benefit is of order $\sqrt{\lambda}$. The total net benefit is therefore of order $\lambda^{3/2}$ and proportional to $f(D)$. Thus, the asymmetric part of the opt-in interval accounts for the $-v_{111}(0, \rho)\Psi_2 f(D)$ term.

We close this section with a discussion of some generalizations. First, recall that our analysis excludes the possibility of setting a default at the boundaries of X (formally, $\underline{D} > \underline{x}$ and $\overline{D} < \overline{x}$).

Assuming the externality is positive, our approximation overstates the net external benefit of setting $D = \underline{x}$ (because there are no ideal points below $\underline{x}$), and understates the net external benefit of setting $D = \overline{x}$ (because there are no ideal points above $\overline{x}$). In either case, if the ideal-point density falls to zero at a boundary point, one can easily adjust our formulas to accomodate a default at that boundary. The most interesting possibility arises when the ideal-point density is bounded away from zero near $\overline{x}$. In that case, the welfare effect in Figure 2 would include the total area under the density function between $D - w_h$ and D, rather than the area of $A \cup B$. Critically, the former area is of order $\sqrt{\lambda}$ rather than λ, which means the aggregate external benefits are of order λ rather than $\lambda^{\frac{3}{2}}$. It follows that, in the limit as $\lambda \to 0$, the positive externalities generated by setting $D = \underline{x}$ dominate all other welfare effects at all possible defaults. Hence, $D = \underline{x}$ is optimal.[21]

Next, consider a non-linear aggregate externality of the form $\sigma N(E(x))$ rather than $\sigma E(x)$. Asymptotically, the externality is approximately $\sigma \left[N(E(x^*)) + (E(x) - E(x^*))N'(E(x^*))\right]$, so the same characterization holds for $J_\lambda(D)$, with Ψ_2 and Ψ_3 subsuming $N'(E(x^*))$. Alternatively, with a non-linear individual-level internality of the form $\sigma N(x)$, the total internality is approximately $\sigma \left[E(N(x^*)) + E(N'(x^*)(x - x^*))\right]$. Notice that the second term integrates the product of $x - x^*$ and $N'(x^*)f(x^*)$. The expression for the total internality would therefore be the same in a setting with linear individual-level internalities and an ideal-point density function $\widetilde{f}(x^*) = N'(x^*)f(x^*)$. Accordingly, the same characterization holds for $J_\lambda(D)$ with a reweighted distribution.

6 Numerical simulations

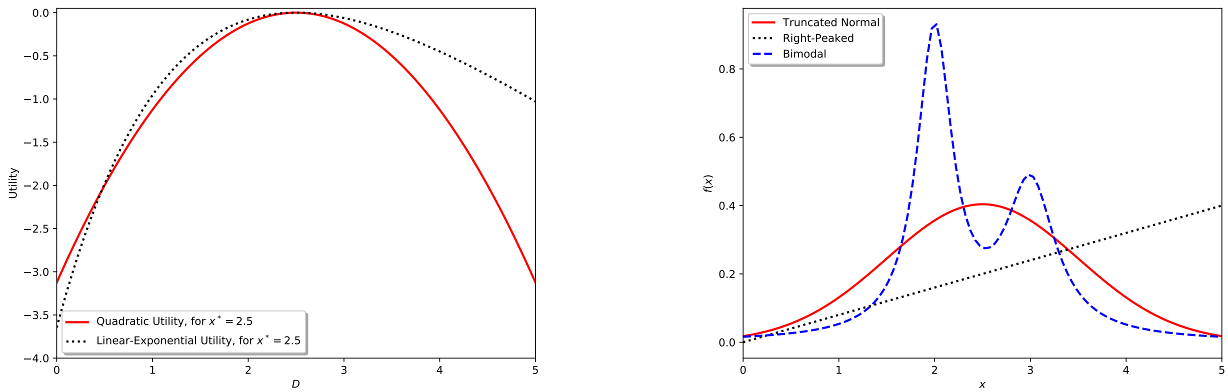
Next, we illustrate our main convergence result by simulating welfare-maximizing, weighted opt-out minimizing, and unweighted opt-out minimizing defaults under various parametric assumptions. These simulations also show that our asymptotic results provides decent approximations for settings with substantial 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.

6.1 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,

[21]When $\lim_{x \to \overline{x}} f(x) > 0$, it also may be natural to assume that there is an atom in the ideal-point distribution at $\overline{x}$. However, that consideration has no bearing on the size of the total externality created by setting $D = \overline{x}$. The reason is that those with $x^* = \overline{x}$ end up with their ideal points. That said, as discussed in Section 4.1, the existence of atoms can also change the analysis in other ways.

Figure 3: Utility Functions and Distribution Functions for Numerical Simulations



Notes: Left panel shows utility functions: quadratic utility (solid red) and linear-exponential asymmetric utility (dotted black) for defaults $D \in [0, 5]$ given ideal point $x^* = 2.5$. Right panel shows density of ideal point x^* over support $x \in [0, 5]$: truncated Normal distribution (solid red), right-peaked distribution (dotted black), and bimodal distribution (dashed blue).

the support of the ideal-point distribution is the interval $[0, 5]$, and the designer agrees with the choosers' assessments of the activity's value ($\sigma = 0$). Figure 3 depicts the alternatives we consider.

When introducing heterogeneity in β and η, we assume $\beta_i \in \{0.5, 0.8, 1\}$ and $\eta_i \in \{0.5, 1, 2\}$. Because the possibilities for correlations are virtually limitless, we either assume independence or 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.

6.2 Simulation results

Table 2 summarizes our main simulation results.[22] Each row represents a separate simulation. Columns (1) through (5) provide details concerning the parametrization. Columns (6) through (13) present results for different values of the cost-scaling parameter λ. For each simulation, we choose the value of λ to achieve the opt-out frequencies listed at the top of the columns: 95%, 90%, 75%, and 40%.[23] Converting values of λ into their implied opt-out frequencies renders the size of the parameter more easily interpretable.[24]

[22]The table omits the combination of a truncated Normal preference distribution and quadratic utility, as the opt-out minimizing and welfare maximizing defaults always coincide. 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 $1e - 12$.

[23]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.

[24]By way of comparison, in the sample studied by Choukhmane (2023), 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.

	(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: Simulation Results for *Weighted* Opt-Out Minimization. Values of λ chosen to achieve the opt-out frequencies indicated at the tops of columns (6)-(11). Checkmarks in columns (1)-(3) indicate whether the simulation includes forms of heterogeneity (in η, β, and interdependence, respectively). Columns (4) and (5) identify the specifications of the utility function and ideal-point distribution. Columns (6), (8), (10), and (12) display the absolute distance between the welfare maximizing default D_L and the weighted opt-out minimizing default D_Ω. Columns (7), (9), (11), and (13) display the percentage of potential welfare gains achived by weighted opt-out minimization; values that round to 100.00 appear as $\geq$ 99.99.

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 λ. 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 $D = 0$: $L_\lambda(D_L(\lambda)) - L_\lambda(0)$. Next we calculate the welfare gain achieved by the weighted opt-out minimizing policy relative to the same baseline: $L_\lambda(D_\Omega(\lambda)) - L_\lambda(0)$. We then define $\Delta_L(\lambda)$ as the ratio of the second welfare gain to the first, expressed as a percentage: $\Delta_L(\lambda) = 100\% \frac{L_\lambda(D_\Omega(\lambda)) - L_\lambda(0)}{L_\lambda(D_L(\lambda)) - L_\lambda(0)}$.

Convergence and the quality of the approximation.

Part A of Table 2 focuses on simulations for which heterogeneity is limited to ideal points. 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.

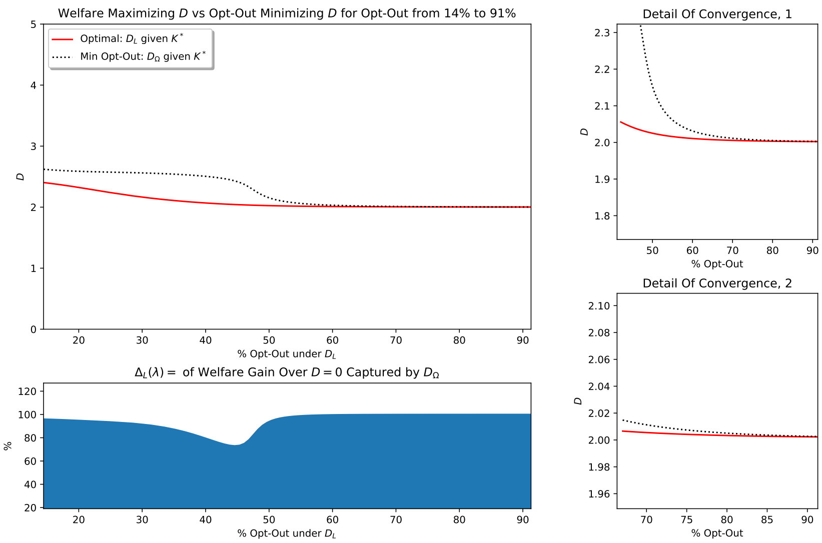
Second, for higher opt-out costs (lower opt-out rates), the correspondence between the two defaults remains close. With 75% opt-out, the maximal distance between $D_\Omega(\lambda)$ and $D_L(\lambda)$ (which again occurs for right-peaked preference distribution and quadratic utility), 0.0521, is only 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 4, 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

Figure 4: Illustration of welfare-maximizing and weighted opt-out-minimizing defaults



Notes: Simulations for linear-exponential utility function, $\beta = 0.8$, homogeneous opt-out costs, bimodal ideal-point density. Main panel: $D_L(\lambda)$ and $D_\Omega(\lambda)$ versus opt-out frequencies. Detail Panels 1 and 2 show same results at higher resolution for λ close to zero. Bottom panel: percentage of potential welfare gain achieved through the weighted opt-out-minimization.

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.

Limitations of the asymptotic approximation method

While the results in Table 2 generally support the asymptotic approximation method, they also help us identify the types of circumstances that render the method less reliable. Focusing on the table's final column, we see a handful of cases for which weighted opt-out minimization achieves only 75 to 80 percent of the potential welfare gain. These cases share two features: first, the distribution of ideal points is bimodal; second, opt-out costs are relatively large (in the sense that 60% of workers stick with the default).

Figure 4 illustrates why these features may erode the accuracy of the asymptotic approximation. With opt-out rates below 50%, weighted opt-out minimization favors defaults in the trough of the ideal-point distribution between the two peaks (at 2% and 3%). In contrast, welfare maximization favors defaults near the highest peak (at 2%) even when the opt-out rate is as low as 30%. This difference is intuitive: within the opt-out window, the average distance between a worker's ideal point and the default matters for the welfare criterion, but does not matter for the opt-out minimization criterion. While a default in the trough can achieve opt-out minimization when the opt-out window is large enough to encompass both peaks, a default near the highest peak can achieve greater welfare because it is closer, on average, to the ideal points of those who opt in. The lower-left panel of the figure shows that opt-out minimization achieves the smallest fraction of the potential welfare gains when the opt-out frequency is around 45%. Reducing opt-out costs from that level shrinks the opt-in window to the point where a default in the trough of the ideal-point distribution no longer captures enough of the mass around the peaks to yield the lowest possible opt-out rate. As a result, the opt-out-minimizing default shifts rapidly toward the higher peak, which brings it into closer alignment with welfare maximization.

Unweighted opt-out minimization

We have conducted analogous simulations for unweighted opt-out minimization. In other words, instead of finding the weighted opt-out minimizing default $D_\Omega(\lambda)$, we find the unweighted opt-out minimizing default $D^U_\Omega(\lambda)$ 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. We also make analogous welfare calculations.

Results appear in the Appendix. 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.

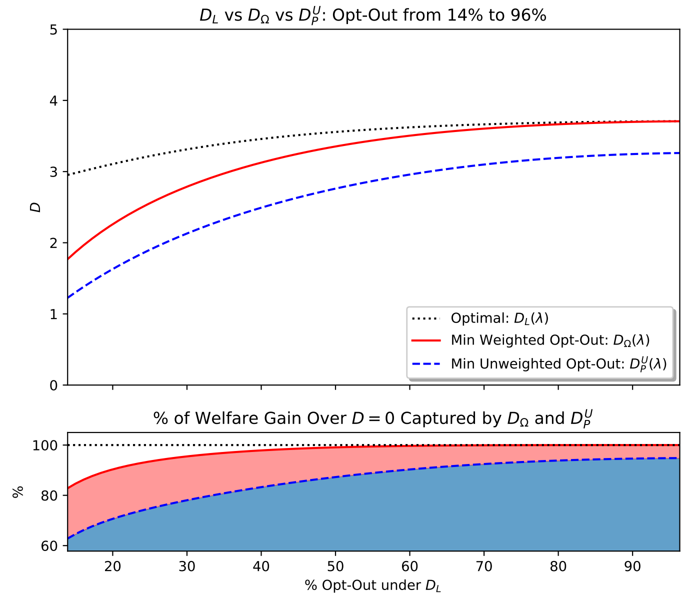
A case with strong correlation between ideal points and opt-out costs

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 asymptotic 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 5, we introduce a strong correlation between η and x^* (correlation coefficient of 0.8).[25] In particular, we assume the population falls into ten groups indexed $i \in$

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

Figure 5: A case where weighted and unweighted opt-out minimization converge to different limits.



Notes: Simulations for setting with ten groups of equal mass. For group i, x^* distributed normally with mean $\mu_i = \frac{5i}{11}$ and $\eta_i = i$; correlation = 0.7939. Top panel: $D_L(\lambda)$, $D_\Omega(\lambda)$, and $D_\Omega^U(\lambda)$ versus opt-out frequencies. Bottom panel: percent of potential welfare gain achieved through the weighted and unweighted opt-out-minimization, $D_\Omega(\lambda)$ and $D_P^U(\lambda)$.

$1, \ldots, 10$, 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.

The bottom panel of the figure shows how weighted and unweighted opt-out minimization perform relative to true welfare optimization. As in Figure 4 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.

7 Practical implications

In this section, we collect and summarize the practical implications of our analysis for the selection of default options. These implications differ according to the designer's objectives (most notably, whether the choosers' selections generate externalities), whether the designer can deploy other instruments (such as Pigouvian fees), certain observable features of the setting (such as whether the menu is sparse or approximately continuous), and the nature of the data the designer can access or generate.

First, suppose the designer agrees with the choosers' ideal points or resolves any disagreement through Pigouvian fees. It is relatively easy to discover the distribution of ideal points, for example through surveys, or more reliably by piloting an active-choice requirement. If there is no reason to suspect strong correlations between ideal points and other characteristics, our results suggest the designer ought to select either the modal ideal point or the antimode, depending on whether biases impacting the assessment of opt-out costs are thought to be mild or severe (Proposition 4).

To refine this selection, the designer must collect additional data. In principle, the relationship between the default and the opt-out frequency is discoverable through experimentation. When opt-out biases are mild (respectively, severe), selecting the opt-out-minimizing (respectively, maximizing) default rather than the modal ideal point (respectively, antimode) can yield an improvement (Proposition 3), because the ideal point may be systematically related to the width of the opt-in window (through the local curvature of the continuation value function).[26]However, the designer should be aware that, when opt-out rates are high and the distribution of ideal points is bimodal, the asymptotic approximation that justifies this rule of thumb may be less accurate.

Further refinements require information concerning the distributions of other chooser characteristics—specifically, as-if opt-out costs and biases pertaining to the evaluation of those costs. Simple experiments or surveys can provide reasonable proxies for those parameters. For example, it is possible to assess the degree of time inconsistency (a potential driver of β) in general population samples through unincentivized survey questions; see Goda et al. (2019), who adapt the approach used in Falk et al. (2023) . Furthermore, because it is only necessary to measure *relative* opt-out costs (in other words, η rather than $\eta\lambda$), measures of willingness-to-pay to avoid opt-out-like tasks, rather than the actual opt-out protocol, likely suffice and are easy to elicit. By conducting such elicitations while piloting an active-choice requirement, one can also determine how η and β covary with x^*.

Knowing the distribution of β, by itself, is useful for two reasons. First, this knowledge allows the designer to assess with greater precision whether the opt-out bias is mild (β generally exceeds $\frac{1}{3}$) or severe (β is generally less that $\frac{1}{3}$). Second, it allows the designer to determine the magnitude of an appropriate Pigouvian fee for passive choice, which (if feasible) can render opt-out minimization

[26]In some contexts, opt-out rates may be insensitive to defaults; see, for example, Brot-Goldberg et al. (2023). If this insensitivity reflects a relatively uniform ideal-point distribution, then our results imply that all alternatives are equally good. An alternative explanation is that people fall into two categories, some who always opt out, some who never opt out. The designer's problem then involves selecting an optimal default for the sugroup who never opt out. Because that group acts as if opt-out costs are infinite, our results do not apply.

superior to opt-out maximization under a much broader range of circumstances (Proposition 5). In principle, simultaneous optimization of the fee and the default could prove challenging because each might affect the other. However, we have shown that the designer can determine the optimal default based on opt-out patterns in a setting where no fine is imposed.

Information concerning the joint distribution of x^*, β, and η facilitates further improvements in the selection of the default D. In settings where the set of options is either continuous or approximates a continuum, and where there is no evidence that ideal points bunch to a significant degree, one can use this information to calculate the average weighting factor $\eta\left(1-\frac{1}{3\beta}\right)$ for each x^*. The best choice is then the default D that maximizes the product of this weight (for $x^* = D$) and the corresponding opt-in frequency (Proposition 2).

For settings where bunching dominates the distribution of selections in an active-choice pilot, the designer can identify the best default through a similar calculation, restricting attention to defaults that coincide with ideal points at which bunching occurs, and replacing the weight $\eta\left(1-\frac{1}{3\beta}\right)$ with the average of η for each such x^* (Proposition 6). A similar prescription applies to settings with sparse menus (see Section 4.2).

For settings with intermediate degrees of bunching, the designer can proceed as in the continuous case, replacing the weights $\eta\left(1-\frac{1}{3\beta}\right)$ with $\eta\left(1-\frac{\chi(x^*)}{3\beta}\right)$ for each ideal point at which meaningful bunching occurs, where $\chi(x^*)$ denotes the ratio of the number of choosers with an ideal point of x^* to the number who opt in when the default is x^* (see Section 4.1). The denominator of this ratio is directly observable. To approximate the numerator, the designer can examine the popularity of x^* when the default is far from x^*.

Next, suppose the designer disagrees with the choosers' ideal points, and cannot resolve this disagreement through Pigouvian fees. For the case of a continuous (or roughly continuous) ideal-point distribution, if there is no reason to suspect strong correlations between ideal points and other characteristics, and if the disagreement takes the form of a roughly linear externality (or internality), our results suggest the designer should set the default to maximize the weighted sum of the ideal-point density and its derivate (Proposition 7). The required information is recoverable by piloting an active-choice requirement. For cases in which (i) the externality is positive, and (ii) the biases impacting the assessment of opt-out costs are thought to be relatively mild, the weight on the ideal-point density is positive, and the weight on its derivative is negative. When the distribution of ideal points is quasi-bell-shaped, it then follows that the optimal default lies between the mode and the upper inflection point; whether it is closer to the latter or the former depends on whether the externality is large or small relative to opt-out costs (Proposition 8). It follows that, for a normal ideal-point distribution, the optimal default lies between the mean and the mean plus one standard deviation–in other words, between the 50th and 84th percentiles.

8 Conclusion

We have provided a general characterization of the solution to the problem of selecting a default option for a heterogeneous group of individuals, which we obtained by studying the asymptotic properties of welfare-maximizing defaults as one of the problem's parameters approaches a limiting value. We interpret these "asymptotic optima" as approximate optima for non-limiting cases and justify this interpretation through numerical simulations. Our main results show that, when the designer and choosers agree about the activity's value, simple forms of weighted opt-out minimization are asymptotically optimal. Existing results concerning the optimality of (unweighted) opt-out minimization and maximization emerge as special cases. Our extensions encompass setting with bunching and sparse menus, for which the weights simplify further. Additional results clarify the role Pigouvian fees and establish robustness with respect to normative ambiguity. A final extension explores settings in which the designer and choosers disagree about the activity's value, for example due to the existence of an externality. Under specified conditions, a designer who cares only about a linear externality should set the default equal to the upper inflection point of the ideal-point distribution; with a more modest externality, the optimum lies between the upper inflection point and the mode. We have also highlighted the practical implications of our analysis for the selection of default options.

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 2022), 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 useful to examine the robustness of our conclusions to these types of possibilities.

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.

— . 2021. In defense of behavioral welfare economics. *Journal of Economic Methodology* 28 (4): 385–400.

— . 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.

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.

Brot-Goldberg, Zarek, Timothy Layton, Boris Vabson, and Adelina Yanyue Wang. 2023. The behavioral foundations of default effecxts: theory and evidence from medicare part d. *American Economic Review* 113 (10): 2718–58.

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. 2023. Default options and retirement saving dynamics.

Ericson, Keith M. Marzilli. 2020. When consumers do not make an active decision: dynamic default rules and their equilibrium effects. *Games and Economic Behavior* 124:369–385.

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2023. The preference survey module: a validated instrument for measuring risk, time, an social preference. *Management Science* 69 (4): 1935–1950.

Goda, Gopi Shah, Matthew Levy, Colleen Flaherty Manchester, Aaron Sojourner, and Joshua Tasoff. 2019. Predicting retirement savings using survey measures of exponential-growth bias and present bias. *Economic Inquiry* 57 (3): 1636–1658.

Goldin, Jacob, and Daniel Reck. 2022. Optimal defaults with normative ambiguity. *Review of Economics and Statistics* 104 (1): 17–33.

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

Greene, William H. 2003. *Econometric analysis.* Prentice Hall.

Handel, Benjamin R. 2013. Adverse seleciton and inertia in health insurance markets: when nudging hurts. *Journal of Public Economics* 103 (7): 2643–2682.

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.*

Appendix

Proofs of Propositions

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. $\square$

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 $\Psi(\varepsilon) \equiv \left\{(D, x, \rho) \in X^2 \times \left[\underline{\rho}, \overline{\rho}\right] \big| \, |D - x| \geqq \varepsilon \right\}$ 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 $\Delta(D, x, \rho) \leq \frac{\lambda \eta}{\beta} < \frac{\lambda^*(\varepsilon) \overline{\eta}}{\underline{\beta}} = \sigma(\varepsilon)$. 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

$$|f(D \mid \theta) - f(x \mid \theta)| < \delta(\lambda) \tag{10}$$

and

$$|d(D, D, \rho) - d(D, x, \rho)| < \delta(\lambda). \tag{11}$$

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

$$|V_{11}(y, x, \rho) - V_{11}(x, x, \rho)| < M_1 |y - x| \, . \tag{12}$$

as well as $M_2 > 0$ for which

$$|V_{11}(x, y, \rho) - V_{11}(x, x, \rho)| < M_2 |y - x| \, . \tag{13}$$

Consequently,

$$\begin{array}{rcl} |d(D,D,\rho) - d(D,x,\rho)| & = & \frac{1}{2}\left|V_{11}(\tilde{x}(D,x,\rho),x,\rho) - V_{11}(D,D,\rho)\right| \\ & = & \frac{1}{2}\left|V_{11}(\tilde{x}(D,x,\rho),x,\rho) - V_{11}(D,x,\rho) + V_{11}(D,x,\rho) - V_{11}(D,D,\rho)\right| \\ & \leq \frac{1}{2} & \left|V_{11}(\tilde{x}(D,x,\rho),x,\rho) - V_{11}(D,x,\rho)\right| + \frac{1}{2}\left|V_{11}(D,x,\rho) - V_{11}(D,D,\rho)\right| \\ & < & M_1\left|\tilde{x}(D,x,\rho) - x\right| + M_2\left|D - x\right| \\ & \leq & (M_1 + M_2)\left|D - x\right|, \end{array} \tag{14}$$

where the second inequality follows from (12) and (13), 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 (14), 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,

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

$$S(D,\theta,\lambda) \supset \left(D - \left(\left(\frac{\lambda\eta}{\beta}\right)\frac{1}{d(D,D,\rho) + \delta(\lambda)}\right)^{\frac{1}{2}}, D + \left(\left(\frac{\lambda\eta}{\beta}\right)\frac{1}{d(D,D,\rho) + \delta(\lambda)}\right)^{\frac{1}{2}}\right) \tag{16}$$

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

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

It thus follows that

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

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

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

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,

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

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,[27] 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:

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

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

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:

$$Z_\lambda(D,\theta) = \frac{\mathbb{E}\left[\Delta(D,x,\rho)|\,\theta, \Delta(D,x,\rho) \leq \frac{\lambda\eta}{\beta}\right]}{\lambda} = \frac{\mathbb{E}\left[\Delta(D,x,\rho)\mathbf{1}_{\Delta(D,x,\rho)\leq\frac{\lambda\eta}{\beta}} \mid \theta\right]}{\lambda \Pr\left[\Delta(D,x,\rho) \leq \frac{\lambda\eta}{\beta} \mid \theta\right]} \tag{17}$$

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 $D \in D$, $\rho \in [\underline{\rho}, \overline{\rho}]$), the numerator of (17) is bounded above by:

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

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

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

where the inequality in the first line follows from (10), (11), and (15) (given that the integrand is strictly positive). It then follows from (17) that

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

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 $X \times [\underline{\rho}, \overline{\rho}]$. Thus, $0 < f^{min}\left(\frac{\underline{\underline{\eta}}}{\beta}\right)^{\frac{1}{2}}\left(\frac{1}{-\frac{1}{2}\underline{v}_{11}}\right)^{\frac{1}{2}} \leq Q(D,\theta) \leq f^{max}\left(\frac{\overline{\eta}}{\underline{\beta}}\right)^{\frac{1}{2}}\left(\frac{1}{-\frac{1}{2}\overline{v}_{11}}\right)^{\frac{1}{2}}$. 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 (17) is bounded below by:

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

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 3.1, 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

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

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 5 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 6 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 $\Delta(D, y, \rho) \leq \frac{\lambda\eta}{\beta} < \frac{\lambda^*(\varepsilon)\overline{\eta}}{\underline{\beta}} = \sigma(\varepsilon)$. 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 (9) tells us that

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

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 $D_{\hat{W}}(\lambda) \to z^*$. □

Proof of Proposition 7 Under Condition 2 (specifically, the translation portion), Assumption 1 (specifically, single crossing) implies $v_{11}(w, \rho) < 0$ for all feasible z. Because v_{11} is continuous (also Assumption 1), we have

$$v_{11}^{max} \equiv \max_{(x,x^*\rho)\in X^2\times[\underline{\rho},\overline{\rho}]} v_{11}(x-x^*,\rho) < 0$$

By definition,

$$v(w,\rho) = v(0,\rho) - \frac{\lambda\eta}{\beta}$$

Because $v(w,\rho)$ is strictly concave in w and maximized at $w=0$, for sufficiently small λ there are two solutions, $w_l(\lambda,\theta) < 0$ and $w_h(\lambda,\theta) > 0$. Furthermore, Taylor's Theorem implies that

$$v(w,\rho) \leq v(0,\rho) + w^2 v_{11}^{max}$$

It follows that

$$\max\{w_h(\lambda,\theta), |w_l(\lambda,\theta)|\} \leq \left(-\frac{\lambda\eta}{\beta v_{11}^{max}}\right)^{\frac{1}{2}}$$

Because $\underline{\beta}, \underline{\eta} > 0$ and $\overline{\eta}$ is finite, we know that $w_h(\lambda,\theta)$ and $w_l(\lambda,\theta)$ converge to zero uniformly over θ as $\lambda \to 0$.

Under Condition 2 (in particular, the translation assumption), we can write the opt-in interval for any given θ as $[\max\{\underline{x}, D + w_l(\lambda,\theta)\}, \min\{\overline{x}, D + w_h(\lambda,\theta)\}]$, where $w_l(\lambda,\theta) < 0$ and $w_h(\lambda,\theta) > 0$ are independent of D. Due to the uniform convergence property notied above, there is some λ' such that, for $\lambda < \lambda'$, we have $\underline{D} + w_l(\lambda,\theta) > \underline{x}$ and $\overline{D} + w_l(\lambda,\theta) < \overline{x}$ for all θ, which means we can write the opt-in interval more simply as $[D + w_l(\lambda,\theta), D + w_h(\lambda,\theta)]$ for all D.

Lemma 4. *Let*

$$A(\theta) = \left(\frac{-2\eta}{\beta v_{11}(0,\rho)}\right)^{1/2} > 0$$

and

$$B(\theta) = \frac{\eta v_{111}(0,\rho)}{3\beta[v_{11}(0,\rho)]^2}.$$

For $\lambda > 0$, *we have*

$$w_h(\lambda,\theta) = A(\theta)\lambda^{1/2} + B(\theta)\lambda + C_h(\lambda,\theta)\lambda^{3/2}$$

and

$$w_l(\lambda,\theta) = -A(\theta)\lambda^{1/2} + B(\theta)\lambda + C_l(\lambda,\theta)\lambda^{3/2},$$

where $C_h(\theta,\lambda)$ *and* $C_l(\theta,\lambda)$ *have finite upper and lower bounds.*

Proof. For any $\ell > 0$, let $w_\theta(\ell)$ be either the positive or the negative solutions to

$$v(w,\rho) = v(0,\rho) - \frac{\ell^2\eta}{\beta}$$

Note that $w_\theta(\ell)$ equals either $w_l(\ell^2,\theta)$ or $w_h(\ell^2,\theta)$, depending on whether we consider the positive or negative solutions. With this change of variables, we can obtain exact second-order Taylor series expansions of $w_l(\ell^2, \theta)$ and $w_h(\ell^2, \theta)$ in terms of ℓ around $\ell = 0$.[28]

From the preceding expression, it follows from implicit differentiation that

$$w'_\theta(\ell) = -\frac{2\ell\eta}{\beta V_1(w_\theta(\ell), \rho)}$$

When $\ell = 0$, the numerator and denominator are both zero. Applying L'Hopital's Rule, we have:

$$\begin{aligned} \lim_{\ell\to 0} w'_\theta(\ell) &= \lim_{\ell\to 0} \tfrac{-2\ell\eta}{\beta v_1(w_\theta(\ell),\rho)} \\ &= \lim_{\ell\to 0} \tfrac{-2\eta}{\beta v_{11}(w_\theta(\ell),\rho) w'_\theta(\ell)} \\ &= \tfrac{-2\eta}{\beta v_{11}(0,\rho) \lim_{\ell\to 0} w'_\theta(\ell)} \end{aligned}$$

Solving for $\lim_{\ell\to 0} w'_\theta(\ell)$, we then have

$$\lim_{\ell\to 0} w'_\theta(\ell) = \pm A(\theta)$$

Whether the limit is positive or negative depends on whether one focuses on the positive or negative solutions (in other words, on $w_l(\lambda, \theta)$ or $w_h(\lambda, \theta)$). Under our assumptions, for any finite $\bar{\ell}$, $w'_\theta(\ell)$ is continuous in ℓ and θ on the compact set $[0,\bar{\ell}] \times \Theta$ (where we interpret $w'_\theta(0)$ as $\lim_{\ell\to 0} w'_\theta(\ell)$). Consequently, it has finite upper and lower bounds.

We obtain the second-order term for the Taylor's expansion around $\ell = 0$ by computing the second derivative and then taking the limit as $\ell \to 0$. For the second derivative, we have

$$\begin{aligned} w''_\theta(\ell) &= -\tfrac{d}{d\ell}\left[\tfrac{2\ell\eta}{\beta v_1(w_\theta(\ell),\rho)}\right] \\ &= -\tfrac{2\eta}{\beta v_1(w_\theta(\ell),\rho)} + \tfrac{2\ell\eta}{\beta [v_1(w_\theta(\ell),\rho)]^2} v_{11}(w_\theta(\ell), \rho) w'_\theta(\ell) \\ &= -\tfrac{2\eta}{\beta v_1(w_\theta(\ell),\rho)} - [w'_\theta(\ell)]^2 \tfrac{v_{11}(w_\theta(\ell),\rho)}{v_1(w_\theta(\ell),\rho)} \\ &= -\tfrac{1}{v_1(w_\theta(\ell),\rho)}\left[\tfrac{2\eta}{\beta} + [w'_\theta(\ell)]^2 v_{11}(w_\theta(\ell), \rho)\right] \end{aligned}$$

When $\ell = 0$, the denominator, $v_1(w_\theta(\ell), \rho)$, is zero, but so is the numerator. (To see this point, substitute our expression for $\lim_{\ell\to 0} w'_\theta(\ell)$.) So as with the first derivative, we use L'Hopital's rule to evaluate the second derivative:

$$\begin{aligned} \lim_{\ell\to 0} w''_\theta(\ell) &= -\lim_{\ell\to 0} \tfrac{2w'_\theta(\ell) w''_\theta(\ell) v_{11}(w_\theta(\ell),\rho) + \left[w'_\theta(\ell)\right]^3 v_{111}(w_\theta(\ell),\rho)}{v_{11}(w_\theta(\ell),\rho) w\theta'(\ell)} \\ &= -2\lim_{\ell\to 0} w''_\theta(\ell) - [\lim_{\ell\to 0} w'_\theta(\ell)]^2 \tfrac{v_{111}(0,\rho)}{v_{11}(0,\rho)} \end{aligned}$$

[28]We use an expansion in ℓ rather than λ because the derivatives of w with respect to λ at $\lambda = 0$ are infinite.

Solving for $\lim_{\ell\to 0} w''_\theta(\ell)$, we then have

$$\lim_{\ell\to 0} w''(\ell) = 2B(\theta)$$

Under our assumptions, for any finite $\bar{\ell}$, $w''_\theta(\ell)$ is continuous in ℓ and θ on the compact set $[0,\bar{\ell}]\times\Theta$ (where we interpret $w''_\theta(0)$ as $\lim_{\ell\to 0} w''_\theta(\ell)$). Consequently, it has finite upper and lower bounds.

We likewise obtain the third-order term for the Taylor's expansion around $\ell = 0$ by computing the third derivative and then taking the limit as $\ell \to 0$. For the third derivative, we have

$$\begin{aligned} w'''_\theta(\ell) &= -\tfrac{d}{d\ell}\left[\tfrac{1}{v_1(w_\theta(\ell),\rho)}\left[\tfrac{2\eta}{\beta} + [w'_\theta(\ell)]^2 v_{11}(w_\theta(\ell),\rho)\right]\right] \\ &= -\tfrac{w'_\theta(\ell)\left[3w''_\theta(\ell)v_{11}(w_\theta(\ell),\rho)+\left[w'_\theta(\ell)\right]^2 v_{111}(w_\theta(\ell),\rho)\right]}{v_1(w_\theta(\ell),\rho)} \end{aligned}$$

When $\ell = 0$, the denominator, $v_1(w_\theta(\ell),\rho)$, is zero, but so is the numerator. (To see this point, substitute our expressions for $\lim_{\ell\to 0} w'_\theta(\ell)$ and $\lim_{\ell\to 0} w''_\theta(\ell)$.) So as with the first and second derivatives, we use L'Hopital's rule to evaluate the third derivative:

$$\begin{aligned} \lim_{\ell\to 0} w'''_\theta(\ell) &= -\lim_{\ell\to 0}\Big[3w'_\theta(\ell)w'''_\theta(\ell)v_{11}(w_\theta(\ell),\rho) + 3\left[w''_\theta(\ell)\right]^2 v_{11}(w_\theta(\ell),\rho) \\ &\qquad +6\left[w'_\theta(\ell)\right]^2 w''_\theta(\ell)v_{111}(w_\theta(\ell),\rho) + \left[w'_\theta(\ell)\right]^3 v_{1111}(w_\theta(\ell),\rho)\Big] / v_{11}(w_\theta(\ell),\rho)w'_\theta(\ell) \\ &= -3\lim_{\ell\to 0} w'''_\theta(\ell) - \lim_{\ell\to 0}\Big[3\left[w''_\theta(\ell)\right]^2 v_{11}(w_\theta(\ell),\rho) + 6\left[w'_\theta(\ell)\right]^2 w''_\theta(\ell)v_{111}(w_\theta(\ell),\rho) \\ &\qquad + \left[w'_\theta(\ell)\right]^3 v_{1111}(w_\theta(\ell),\rho)\Big] / v_{11}(w''_\theta(\ell),\rho)w'_\theta(\ell) \end{aligned}$$

Solving for $\lim_{\ell\to 0} w'''_\theta(\ell)$, we then have

$$\lim_{\ell\to 0} w'''_\theta(\ell) = -\frac{1}{4}\frac{3\left[w''_\theta(0)\right]^2 v_{11}(0,\rho) + 6\left[w'_\theta(0)\right]^2 w''_\theta(0)v_{111}(0,\rho) + \left[w'_\theta(0)\right]^3 v_{1111}(0,\rho)}{v_{11}(0,\rho)w'_\theta(0)}$$

Under our assumptions, for any finite $\bar{\ell}$, $w'''_\theta(\ell)$ is continuous in ℓ and θ on the compact set $[0,\bar{\ell}]\times\Theta$ (where we interpret $w'''_\theta(0)$ as $\lim_{\ell\to 0} w'''_\theta(\ell)$). Consequently, it has finite upper and lower bounds.

According to Taylor's Theorem, the positive solution for $w(\ell)$ satisfies

$$w^+_\theta(\ell) = A(\theta)\ell + B(\theta)\ell^2 + \frac{1}{6}w'''_\theta(\varsigma_h(\ell,\theta))\ell^3$$

some $\varsigma_h(\ell) \in [0,\ell]$ (where $w'''_\theta(\varsigma_h(\ell))$ is evaluated using the positive solutions), while the negative solution satisfies

$$w^-_\theta(\ell) = -A(\theta)\ell + B(\theta)\ell^2 + \frac{1}{6}w'''_\theta(\varsigma_l(\ell,\theta))\ell^3$$

for some $\varsigma_l(\ell,\theta), \varsigma_h(\ell,\theta) \in [0,\ell]$ (where $w'''_\theta(\varsigma_h(\ell))$ is evaluated using the negative solutions). Substituting $\sqrt{\lambda}$ for ℓ in these equations, defining $C_k(\lambda,\theta) = \frac{1}{6}w'''_\theta(\varsigma_k(\ell))$ for $k = h, \ell$, and defining $\xi_k(\lambda,\theta) = \sqrt{\varsigma_k(\ell,\theta)}$ for $k = h, \ell$ establishes the lemma. $\square$

Now observe that we can write $J_\lambda(D)$ as

$$J_\lambda(D) = \int_\theta K_\lambda(D) dG(\theta)$$

where

$$K_\lambda(D,\theta) = \frac{\sigma}{2\lambda^{\frac{3}{2}}}\left[\int_{\Delta(D,x^*,\rho)\leq\frac{\lambda\eta}{\beta}} (D - x^*)f(x^*)dx^*\right]$$

Lemma 5. *As $\lambda \to 0$, $K_\lambda(D,\theta)$ converges uniformly to $-\sigma v'''(0)\Lambda_2(\theta)f(D) - \sigma\Lambda_3(\theta)f'(D)$, where*

$$\Lambda_2(\theta) = \frac{2^{\frac{1}{2}}\eta^{\frac{3}{2}}}{3\beta^{\frac{3}{2}}[-v_{11}(0,\rho)]^{\frac{5}{2}}} > 0$$

and

$$\Lambda_3(\theta) = \frac{1}{3}\left[\frac{-2\eta}{\beta v_{11}(0,\rho)}\right]^3 > 0$$

Proof. Define

$$w_h^0(\lambda,\theta) = A(\theta)\lambda^{1/2} + B(\theta)\lambda$$

and

$$w_l^0(\lambda,\theta) = -A(\theta)\lambda^{1/2} + B(\theta)\lambda$$

In light of the assumed bounds on η, β, v_{11}, and v_{111}, there is some $\lambda'' > 0$ with $\lambda'' < \lambda'$ such that, for $\lambda < \lambda''$, we have $w_h^0(\lambda,\theta) > 0$ and $w_l^0(\lambda,\theta) < 0$ for all θ (because the $\lambda^{1/2}$ terms dominate). Henceforth, we will confine attention to such λ.

To approximate $K_\lambda(D,\theta)$, we will consider the following function:

$$K_\lambda^0(D,\theta) = \frac{\sigma}{2\lambda^{\frac{3}{2}}}\left[\int_{D+w_l^0(\lambda,\theta)}^{D+w_h^0(\lambda,\theta)} (D - x^*)f(x^*)dx^*\right]$$

We will focus on values of ρ for which $v_{111}(0,\rho) \leq 0$, so that $\left|w_l^0(\lambda,\theta)\right| \geq w_h(\lambda,\theta)$ (a left-skewed approximate opt-in window); for the case of $v_{111}(0,\rho) > 0$ (i.e., $\left|w_l^0(\lambda,\theta)\right| < w_h^0(\lambda,\theta)$), the proof requires small modifications. With this focus, we can write

$$K_\lambda^0(D,\theta) = K_\lambda^S(D,\theta) + K_\lambda^A(D,\theta),$$

where $K_\lambda^S(D,\theta)$ is the symmetric part of the approximate opt-in window,

$$K_\lambda^S(D,\theta) \equiv \frac{\sigma}{2\lambda^{\frac{3}{2}}}\left[\int_{D-w_h^0(\lambda,\theta)}^{D+w_h^0(\lambda,\theta)} (D - x^*)f(x^*)dx^*\right],$$

and $K^A_\lambda(D,\theta)$ is the asymmetric part,

$$K^A_\lambda(D,\theta) \equiv \frac{\sigma}{2\lambda^{\frac{3}{2}}}\left[\int_{D+w^0_l(\lambda,\theta)}^{D-w^0_h(\lambda,\theta)}(D-x^*)f(x^*)dx^*\right].$$

STEP 1: $K^S_\lambda(D,\theta)$ converges uniformly to $-\sigma\Lambda_3(\theta)f'(D)$.

By Taylor's Theorem, there exists $\xi_1(x^*)$ between x^* and D s.t.

$$\begin{aligned} K^S_\lambda(D,\theta) &= \tfrac{\sigma}{2\lambda^{\frac{3}{2}}}\left[\int_{D+w^0_h(\lambda,\theta)}^{D+w^0_h(\lambda,\theta)}(D-x^*)\left[f(D)+f'(D)(x^*-D)+\tfrac{1}{2}f''(\xi_1(x^*))(x^*-D)^2\right]dx^*\right] \\ &= \tfrac{\sigma}{2\lambda^{\frac{3}{2}}}\left[-\tfrac{f(D)}{2}(x^*-D)^2-\tfrac{f'(D)}{3}(D-x^*)^3\right]_{D-w^0_h(\lambda,\theta)}^{D+w^0_h(\lambda,\theta)} \\ &\qquad -\tfrac{\sigma}{4\lambda^{\frac{3}{2}}}\int_{D-w^0_h(\lambda,\theta)}^{D+w^0_h(\lambda,\theta)} f''(\xi_1(x^*))(x^*-D)^3dx^* \end{aligned}$$

For the first term, we have

$$\frac{\sigma}{2\lambda^{\frac{3}{2}}}\left[-\frac{f(D)}{2}(x^*-D)^2-\frac{f'(D)}{3}(x^*-D)^3\right]_{D-w^0_h(\lambda,\theta)}^{D+w^0_h(\lambda,\theta)} = -\frac{\sigma}{3}f'(D)\left[A(\theta)+B(\theta)\lambda^{\frac{1}{2}}\right]^3$$

In light of the assumed bounds on η, β, v_{11}, v_{111}, and f', this term converges uniformly as $\lambda \to 0$ to $\frac{\sigma}{3}f'(D)\left[A(\theta)\right]^3$.

Let $f''_{abs} = \max_{x\in X}|f''(x)|$. Note that, under our assumptions, f''_{abs} is finite. Then, for the second term, we have

$$\begin{aligned} \left|\tfrac{\sigma}{4\lambda^{\frac{3}{2}}}\int_{D-w^0_h(\lambda,\theta)}^{D+w^0_h(\lambda,\theta)} f''(\xi_1(x^*))(x^*-D)^3dx^*\right| &\le \tfrac{\sigma}{4\lambda^{\frac{3}{2}}}\int_{D-w^0_h(\lambda,\theta)}^{D+w^0_h(\lambda,\theta)} |f''(\xi_1(x^*))|\left|x^*-D\right|^3dx^* \\ &< \tfrac{\sigma}{4\lambda^{\frac{3}{2}}}\int_{D-w^0_h(\lambda,\theta)}^{D+w^0_h(\lambda,\theta)} f''_{abs}\left[w^0_h(\lambda,\theta)\right]^3dx^* \\ &= \tfrac{\sigma}{2}f''_{abs}\left[A(\theta)\lambda^{\frac{1}{8}}+B(\theta)\lambda^{\frac{5}{8}}\right]^4 \end{aligned}$$

In light of the assumed bounds on η, β, v_{11}, and v_{111},, this term converges uniformly to 0 as $\lambda \to 0$.

Combining these conclusions, we see that $K^S_\lambda(D,\theta)$ converges uniformly to

$$-\frac{\sigma}{3}f'(D)\left[\frac{-2\eta}{\beta v_{11}(0,\rho)}\right]^3$$

The desired conclusions for Step 1 follows immediately.

STEP 2: $K^A_\lambda(D,\theta)$ converges uniformly to $-\sigma v_{111}(0,\rho)\Lambda_2(\theta)f(D)$.

By Taylor's Theorem, there exists $\xi_2(x^*)$ between x^* and D s.t.

$$K_\lambda^A(D,\theta) = \frac{\sigma}{2\lambda^{\frac{3}{2}}}\left[\int_{D+w_l^0(\lambda,\theta)}^{D-w_h^0(\lambda,\theta)}(D-x^*)\left[f(D)+f'(\xi_2(x^*))(x^*-D)\right]dx^*\right]$$

$$= \frac{\sigma}{2\lambda^{\frac{3}{2}}}\left[-\frac{f(D)}{2}(x^*-D)^2\right]_{D+w_l^0(\lambda,\theta)}^{D-w_h^0(\lambda,\theta)} - \frac{\sigma}{2\lambda^{\frac{3}{2}}}\int_{D+w_l^0(\lambda,\theta)}^{D-w_h^0(\lambda,\theta)} f'(\xi_2(x^*))(x^*-D)^2dx^*$$

For the first term, we have

$$\frac{\sigma}{2\lambda^{\frac{3}{2}}}\left[-\frac{f(D)}{2}(x^*-D)^2\right]_{D+w_l^0(\lambda,\theta)}^{D-w_h^0(\lambda,\theta)} = -\sigma f(D)A(\theta)B(\theta)$$

Let $f_{abs}^{'} = \max_{x\in X}|f'(x)|$. Note that, under our assumptions, $f_{abs}^{'}$ is finite. Then, for the second term, we have

$$\left|\frac{\sigma}{2\lambda^{\frac{3}{2}}}\int_{D+w_l^0(\lambda,\theta)}^{D-w_h^0(\lambda,\theta)} f'(\xi_2(x^*))(x^*-D)^2dx^*\right| \le \frac{\sigma}{2\lambda^{\frac{3}{2}}}\int_{D+w_l^0(\lambda,\theta)}^{D-w_h^0(\lambda,\theta)} |f'(\xi_2(x^*))|\,|x^*-D|^2\,dx^*$$

$$< \frac{\sigma}{4\lambda^{\frac{3}{2}}}\int_{D+w_l^0(\lambda,\theta)}^{D-w_h^0(\lambda,\theta)} f'_{abs}\left|w_l^0(\lambda,\theta)\right|^2 dx^*$$

$$= -\frac{\sigma}{4\lambda^{\frac{3}{2}}} f'_{abs}\left|w_l^0(\lambda,\theta)\right|^2\left(w_h^0(\lambda,\theta)+w_l^0(\lambda,\theta)\right)$$

$$= -\frac{\sigma}{4\lambda^{\frac{3}{2}}} f'_{abs}\left|-A(\theta)+B(\theta)\lambda^{\frac{1}{2}}\right|^2 2B(\theta)\lambda^{\frac{1}{2}}$$

In light of the assumed bounds on η, β, v_{11}, and v_{111}, this term converges uniformly to 0 as $\lambda \to 0$.

Combining these conclusions, we see that $K_\lambda^A(D,\theta)$ converges uniformly to

$$-\sigma f(D)\left[\frac{2^{\frac{1}{2}}\eta^{\frac{3}{2}}}{3\beta^{\frac{3}{2}}[-v_{11}(0,\rho)]^{\frac{5}{2}}}\right]v_{111}(0,\rho)$$

The desired conclusion for Step 2 follows immediately.

Step 3: $K_\lambda(D,\theta) - K_\lambda^0(D,\theta)$ converges uniformly to 0.

The formulas for $K_\lambda(D,\theta)$ and $K_\lambda^0(D,\theta)$ differ only with respect to the limits of integration: for the lower limit, $w_l^0(\lambda,\theta)$ may differ from $w_l(\lambda,\theta)$, and for the upper limit, $w_h^0(\lambda,\theta)$ may differ from $w_h(\lambda,\theta)$. We will bound the difference between $K_\lambda(D,\theta)$ and $K_\lambda^0(D,\theta)$ resulting from a discrepancy between $w_h^0(\lambda,\theta)$ and $w_h(\lambda,\theta)$ when $w_h^0(\lambda,\theta) < w_h(\lambda,\theta)$. Parallel arguments apply to the case of $w_h^0(\lambda,\theta) > w_h(\lambda,\theta)$, as well as to the lower limits of integration.

Using the same Taylor's expansion as in Step 2, we know there exists $\xi_2(x^*)$ between x^* and D s.t.

$$\frac{\sigma}{2\lambda^{\frac{3}{2}}}\left[\int_{D+w_h(\lambda,\theta)}^{D+w_h^0(\lambda,\theta)}(D-x^*)f(x^*)dx^*\right] = \frac{\sigma}{2\lambda^{\frac{3}{2}}}\left[\int_{D+w_h(\lambda,\theta)}^{D+w_h^0(\lambda,\theta)}(D-x^*)\left[f(D)+f'(\xi_2(x^*))(x^*-D)\right]dx^*\right]$$

$$= \frac{\sigma}{2\lambda^{\frac{3}{2}}}\left[-\frac{f(D)}{2}(x^*-D)^2\right]_{D+w_h(\lambda,\theta)}^{D+w_h^0(\lambda,\theta)} - \frac{\sigma}{2\lambda^{\frac{3}{2}}}\int_{D+w_h(\lambda,\theta)}^{D+w_h^0(\lambda,\theta))} f'(\xi_2(x^*))(x^*-D)^2dx^*$$

Using Lemma 4, we can rewrite the first term as

$$\frac{\sigma}{2\lambda^{\frac{3}{2}}}\left[-\frac{f(D)}{2}(x^*-D)^2\right]_{D+w_h(\lambda,\theta)}^{D+w_h^0(\lambda,\theta)} = -\frac{1}{4}\sigma f(D)\left[(C_h(\lambda,\theta)-C_l(\lambda,\theta))\left(A(\theta)\lambda^{\frac{1}{2}}-B(\theta)\lambda\right) + \left([C_h(\lambda,\theta)]^2-[C_l(\lambda,\theta)]^2\right)\lambda^{\frac{3}{2}}\right.$$

Because all of the terms in this expression are bounded, it converges to 0 uniformly over D and θ as $\lambda \to 0$.

For the second term, we have

$$\left|\frac{\sigma}{2\lambda^{\frac{3}{2}}}\int_{D+w_h(\lambda,\theta)}^{D+w_h^0(\lambda,\theta)} f'(\xi_2(x^*))(x^*-D)^2dx^*\right| \leq \frac{\sigma}{2\lambda^{\frac{3}{2}}}\int_{D+w_h(\lambda,\theta)}^{D+w_h^0(\lambda,\theta)}\left|f'(\xi_2(x^*))\right|\left|x^*-D\right|^2dx^*$$

$$< \frac{\sigma}{2\lambda^{\frac{3}{2}}}\int_{D+w_h(\lambda,\theta)}^{D+w_h^0(\lambda,\theta)} f'_{abs}\left|w_h^0(\lambda,\theta)\right|^2dx^*$$

$$= \frac{\sigma}{4\lambda^{\frac{3}{2}}}f'_{abs}\left|w_h^0(\lambda,\theta)\right|^2\left(w_h^0(\lambda,\theta)-w_h(\lambda,\theta)\right)$$

$$= \frac{\sigma}{4}f'_{abs}\left|A(\theta)\lambda^{\frac{1}{2}}+B(\theta)\lambda\right|^2 C_h(\lambda,\theta)$$

(where the final step follows from Lemma 4). Because all of the terms in this expression are bounded, it converges to 0 uniformly over D and θ as $\lambda \to 0$.

The desired conclusion for Step 3 follows immediately.

The lemma follows from the combination of Steps 1 through 3. □

From Lemma 2, it follows that $J_\lambda(D)$ converges uniformly to the following function:

$$\int_\theta K(D,\theta)dG(\theta) = -\int_\theta\left[\sigma v_{111}(0,\rho)\Lambda_2(\theta)f(D)+\sigma\Lambda_3(\theta)f'(D)\right]dG(\theta)$$

$$= -\sigma v_{111}(0,\rho)\Psi_2 f(D) - \sigma\Psi_3 f'(D)$$

where $\Psi_k = \int_\theta \Lambda_k(\theta)dG(\theta)$ for $k = 2, 3$.

In the proof of Proposition 2, we showed that $\overline{W}_\lambda(D)$ converges uniformly to $\Omega(D)$. Furthermore, in demonstrating Proposition 4, we showed that $\Omega(D) = \Psi_1 f(D)$ under Condition 2, where Ψ_1 is defined in Section 3.2. It follows that $\overline{W}_\lambda(D)$ converges uniformly to $\Psi_1 f(D)$.

Combining these conclusions proves the proposition. □

Proof of Proposition 8. Under the stated assumptions, $\Psi_1 > 0$, $\Psi_2 \leq 0$, and $\Psi_3 > 0$, so in the asymptotic objective function, the total weight on $f(D)$ is strictly positive while the total weight on $f'(D)$ is negative.

For any $D < x_m$, we have $f(D) < f(x_m)$ and $f'(D) > 0 = f'(x_m)$. Therefore, the value of the objective function is strictly greater at x_m than at D. For any $D > x_h$, we have $f(D) > f(x_h)$ and $f'(D) > f'(x_h)$. Therefore, the value of the objective function is strictly greater at x_h than at D. It follows that any maximizer lies within $[x_m, x_h]$.

Because $[x_m, x_h]$ is interior to X, a maximizer must satisfy the first-order condition,

$$[\Psi_1 - \sigma v_{111}(0,\rho)\Psi_2]\, f'(D) = \sigma \Psi_3 f''(D)$$

Within $[x_m, x_h]$, the left-hand side of the first-order condition is decreasing in D, while the right-hand side is increasing (given that $f'''(x) > 0$ for $x \in (x_m, x_h)$), so the maximizer, $\overline{D}^*$, is necessarily unique. The asymptotic optimality of $\overline{D}^*$ then follows from Proposition 7.

Because the objective function exhibits increasing differences in D and σ (the cross-partial derivative being $-\Psi_3 f''(D) > 0$ on $[x_m, x_h)$), $\overline{D}^*$ is increasing in σ by Topkis's Theorem. Suppose $v_{111}(0,\rho) = 0$. Let $R(x) = \max_{x \in [x_m, x']} \frac{\Psi_1 f'(x)}{\Psi_3 f''(x)}$. Because (i) f' and f'' are continuous, (ii) $f'(x) < 0$ and $f''(x) < 0$ for $x \in [x_m, x']$, and (iii) $[x_m, x']$ is compact, it follows that $R(x)$ is well-defined and strictly positive. Consider any $x' < x_h$. For $\sigma > R(x')$, no $x \in [x_m, x_h]$ can satisfy the first-order condition. We therefore have $\overline{D}^* \in (x', x_h)$. It follows immediately that $\lim_{\sigma \to \infty} \overline{D}^* = x_h$. □

Additional simulations

The table below replicates the simulation results of Table 2 for *unweighted* (rather than weighted) opt-out minimization. In other words, instead of finding the weighted opt-out minimizing default $D_\Omega(\lambda)$, we find the unweighted opt-out minimizing default $D^U_\Omega(\lambda)$ 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. The table reports the absolute value of the difference between the welfare-maximizing and (unweighted) opt-out-minimizing defaults, i.e., $|D_L(\lambda) - D^U_\Omega(\lambda)|$. We also calculate the welfare gain achieved by the unweighted opt-out minimizing policy relative to the same baseline as before: $L_\lambda(D_\Omega(\lambda)) - L_\lambda(0)$. We then define $\Delta^U_L(\lambda)$ as the ratio of the second welfare gain to the first, expressed as a percentage: $\Delta^U_L(\lambda) = 100\% \frac{L_\lambda\left(D^U_\Omega(\lambda)\right) - L_\lambda(0)}{L_\lambda(D_L(\lambda)) - L_\lambda(0)}$.

	(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	$\|\|D_L - D_\Omega^U\|\|$	Δ_L^U %	$\|\|D_L - D_\Omega^U\|\|$	Δ_L^U %	$\|\|D_L - D_\Omega^U\|\|$	Δ_L^U %	$\|\|D_L - D_\Omega^U\|\|$	Δ_L^U %
				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. Values of λ chosen to achieve the opt-out frequencies indicated at the tops of columns (6)-(11). Checkmarks in columns (1)-(3) indicate whether the simulation includes forms of heterogeneity (in η, β, and interdependence, respectively). Columns (4) and (5) identify the specifications of the utility function and ideal-point distribution. Columns (6), (8), (10), and (12) display the absolute distance between the welfare maximizing default D_L and the unweighted opt-out minimizing default D_Ω. Columns (7), (9), (11), and (13) display the percentage of potential welfare gains achived by unweighted opt-out minimization; values that round to 100.00 appear as ≥ 99.99.