Appendix (For Online Publication Only)

A Appendix: Proofs

Table 1: Composition of Groups for a Change in the Default

Group	Behavior when default is:		Characterization
	d_0	d_1	
Always Active (AA)	$a_i(d_0) > 0$	$a_i(d_1) > 0$	$u_i(x_i^*) - \max\{u_i(d_0),\ u_i(d_1)\} > \gamma_i$
Always Passive (PP)	$a_i(d_0) \leq 0$	$a_i(d_1) \leq 0$	$u_i(x_i^*) - \min\{u_i(d_0),\ u_i(d_1)\} \leq \gamma_i$
Active-to-Passive (AP)	$a_i(d_0) > 0$	$a_i(d_1) \leq 0$	$u_i(x_i^*) - u_i(d_0) > \gamma_i \geq u_i(x_i^*) - u_i(d_1)$
Passive-to-Active (PA)	$a_i(d_0) \leq 0$	$a_i(d_1) > 0$	$u_i(x_i^*) - u_i(d_1) > \gamma_i \geq u_i(x_i^*) - u_i(d_0)$

Note: This table describes how the composition of the four groups described in Section 2 is determined in terms of the behavioral parameters from Equation (2). The characterization of these groups in terms of primitives is used in several of the proofs.

Lemma 1:

$$W(d) = E[u_i(x_i^*) - \pi_i\gamma_i | a_i(d) > 0]\ (1 - F_{a;d}(0)) + E[u_i(d)|a_i(d) \leq 0]\, F_{a;d}(0)$$

where $F_{a;d}(\cdot)$ denotes the cumulative density function of $a_i(d)$.

Proof: From the definition of the social welfare function we know that $W(d) = E[v_i(d)]$. By the law of iterated expectations,

$$W(d) = E[v_i(d)|a_i(d) > 0]P(a_i(d) > 0) + E[v_i(d)|a_i(d) \leq 0]P(a_i(d) \leq 0)$$

We know from the consumer's problem and the definition of $a_i(d)$ that 1) $a_i(d) \leq 0 \implies x_i(d) = d$ and 2) $a_i(d) > 0 \implies x_i(d) = x_i^* = \arg\max u_i(x)$. Substituting these into $v_i(d) = w_i(x_i(d), d) = u_i(x_i(d)) - \pi_i\gamma_i 1\{x_i(d) \neq d\}$ gives the result. ■

Proposition 1: *For any two defaults $d_0, d_1 \in X$:*

$$W(d_1) - W(d_0) = E[u_i(x^*) - u_i(d_0) - \pi_i\gamma_i|PA]\, p(PA) - E[u_i(x^*) - u_i(d_1) - \pi_i\gamma_i|AP]\, p(AP) + E[u_i(d_1) - u_i(d_0)|I$$

Proof: We know that $W(d_1) - W(d_0) = E[v_i(d_1) - v_i(d_0)]$. We partition individuals into the four groups (PA, AP, PP and AA) and apply the law of iterated expectations to express the change in welfare as a probability-weighted sum over these four groups. As before, $a_i(d) \leq 0 \implies x_i(d) = d$ and 2) $a_i(d) > 0 \implies x_i(d) = x_i^* = \arg\max u_i(x)$. In the PA group, $a_i(d_1) > 0$ so $v_i(d_1) = u_i(x_i^*) - \pi_i\gamma_i$,and $a_i(d_0) \leq 0$, so $v_i(d_0) = u_i(d_0)$. Thus $E[v_i(d_1) - v_i(d_0)|PA] = E[u_i(x^*) - u_i(d_0) - \pi_i\gamma_i|PA]$. Proceeding similarly for the other four groups and substituting in the resulting expressions yields the desired result. ■

Proposition 2: *Let X be any interval in $\mathbb{R}$. If d^* represents an interior solution to the optimal default problem, the following first-order condition is satisfied:*

$$\begin{aligned}
0 = W^{'}(d^*) \quad &= \quad E[(1-\pi_i)\gamma_i | a_i(d^*) = 0,\ u_i'(d^*) < 0]\, f_{a|u'<0}(0)\, F_{u'}(0) \\
&- \quad E[(1-\pi_i)\gamma_i | a_i(d^*) = 0,\ u_i'(d^*) > 0]\, f_{a|u'<0}(0)\ (1 - F_{u'}(0)) \\
&+ \quad E\left[u'(d^*)\,|\, a_i(d^*) < 0\right]\ F_{a;d^*}(0)
\end{aligned}$$

where $f_{a|u'>0}$ is the probability density function of $a_i(d^*)$ conditional on $u_i'(d^*) > 0$; $F_{u'}$ is the cumulative density function of $u_i'(d^*)$; and, as above, $F_{a;d^*}$ is the cumulative density function of $a_i(d^*)$.

Proof: One can obtain this result by direct calculation of the derivative of the welfare function, as divided into active and passive choosers in Lemma 1 (i.e. expressing the expectations as integrals and applying Leibniz rule). One can also obtain the result by plugging in $d_1 = d_0 + \Delta d$ in Proposition 1, taking the limit as Δd approaches zero, plugging in the definitions of the primitives, and noting that the PA and AP groups now both have $a_i(d) = 0$, which implies that $u_i(x^*) - u_i(d) = \gamma_i$ by construction. ■

Proposition 3 Suppose that there exists a penalty default $d_p \in X$.

(3.1) *There exists a threshold $\underline{\pi} \in [0, 1)$ such that $\pi_i \leq \underline{\pi}$ for all i implies d_p maximizes social welfare.*

Proof: We will prove the existence of a threshold $\underline{\pi} \in [0, 1)$ such that when $\pi_i \leq \underline{\pi}$, $W(d^p) \geq W(d)$ for any d.

Let $X^A \subset X$ be the subset of X such that for any $d \in X^A$, $P(a_i(d) \leq 0) > 0$.

Let $d \in X$ be an arbitrary default. We know $W(d^p) \geq W(d)$ is trivially true when d is also a penalty default, i.e. $d \notin X^A$ as then $W(d) = W(d_p)$ for any π. Next suppose $d \in X^A$, so $p(PA) > 0$. Let $\tilde{\pi}(d) = \sup_{i \in PA(d)} \pi_i$ be the largest possible value of π_i for the PA group for default d. We know from Equation (7) that

$$W(d_p) - W(d) \geq p(PA)\{E[u_i(x^*) - u_i(d)|PA] - \tilde{\pi}(d)E[\gamma_i|PA]\} \tag{13}$$

The RHS of this expression is a continuous and strictly monotonically decreasing function of $\tilde{\pi}(d)$ (so long as $E[\gamma_i|PA] > 0$, which must be true because PA individuals choose passively). When $\tilde{\pi}(d) = 0$, the RHS of this expression is weakly positive because $u_i(x^*) \geq u_i(d)$ for all i.[27] When $\tilde{\pi}(d) = 1$, the RHS is strictly negative because $u_i(x^*) - u_i(d) < \gamma_i$ for all individuals that are passive at d, which is the PA group in this situation. The Intermediate Value Theorem then implies there is a value of $\tilde{\pi}(d)$, such that we know that the expression on the RHS of (13) is 0. Denoting this threshold by $\underline{\pi}(d)$, we have that $W(d_p) - W(d) \geq 0$ when $\pi_i \leq \underline{\pi}(d)$ for all i. The result then follows from letting $\underline{\pi} = \inf_{d \in X^A} \underline{\pi}(d)$, so that $\pi_i < \underline{\pi}$ implies $W(d_p) - W(d) \geq 0$ for any d. ■

(3.2) *There exists a threshold $\overline{\pi} \in (0, 1]$ such that $\pi_i \geq \overline{\pi}$ for all i implies d_p* minimizes *social welfare.*

Proof: The proof is analogous to the proof of (3.1). For any default d, let $\hat{\pi}(d) = \inf_{i \in PA(d)} \pi_i$. Using equation (7) and a similar Intermediate Value Theorem argument to the above we derive that there is a threshold $\overline{\pi}(d)$, such that $\pi_i \geq \overline{\pi}(d)$ implies $W(d_p) - W(d) \leq 0$. The result then follows from letting $\overline{\pi} = \sup_{d \in X^A} \overline{\pi}(d)$. ■.

Proposition 4 *Suppose that $X = [x_{min}, x_{max}] \subseteq \mathbb{R}$ and that:*

(A4.1) *As-if costs γ_i are distributed independently of x_i^*.*

(A4.2) *Preferences are given by $u_i(x) = u(x - x_i^*)$ for some map $u : \mathbb{R} \to \mathbb{R}$, with*

[27] The expression is *strictly* positive if we presume x_i^* is a *unique* maximum for every individual. Under this assumption, we know that the penalty default d_p is the *uniquely* optimal default.

$u'(0) = 0$, $u'' < 0$ and $u(c) = u(-c)$ for any c.

(A4.3) $\quad$ *x_i^* follows a single-peaked and symmetric distribution about some mode x^m.*

Under these conditions, there exists a threshold $\overline{\pi} \in (0, 1]$ such that $\pi_i \geq \overline{\pi}$ for all i implies that the optimal default is the default that minimizes opt-outs.

Proof: We provide the proof of the theorem for the case when $\pi_i = 1$ for all i. It is straightforward to show that if the theorem holds when $\pi_i = 1$ for all i, it must hold for sufficiently high π_i.

Starting from the case where $\pi_i = 1$ for all i, we first prove that $W'(x^m) = 0$, $W'(d) > 0$ for $d < x^m$, and $W'(d) < 0$ for $d > x^m$, which implies that W has a unique global maximum at x^m. We then prove that opt-outs are minimized under x^m. We start by letting $d \in X$ be some default.

Step 1: Characterizing the first and second derivative of $W(d)$.

Let $W_\gamma(d) = E[v_i(d)|\gamma_i = \gamma]$. By (A4.1) we know that $W(d) = \int_\gamma W_\gamma(d) f(\gamma) d\gamma$. To prove our result, it therefore suffices to prove that for any fixed γ, $W'_\gamma(d) = 0$ if $d = x^m$, and $W''_\gamma(d) < 0$ always.

We first introduce some notation involving the function $u()$. Without loss of generality $u(0) = 0$. Taking γ as given, by (A4.2) there is some unique value ξ such that $u(\xi) = u(-\xi) = \gamma$. Note that when $x^* = d - \xi$, utility at the default is given by $u(d - x^*) = u(d - (d - \xi)) = u(\xi) = \gamma$, and similarly when $x^* = d + \xi$, $u(d - (d + \xi)) = \gamma$. By (4.2), an individual is active when $x_i^* \leq d - \xi$ or $x_i^* \geq d + \xi$.

We next characterize $W'_\gamma(d)$. For illustrative purposes, suppose $\pi_i = \pi$ is homogeneous for all i, which is true when $\pi_i = 1$ for all i. Welfare for people with given γat d is given by

$$W_\gamma(d) = \int_{x^*=-\infty}^{d-\xi} -\pi\gamma f(x^*) dx^* + \int_{x^*=d-\xi}^{d+\xi} u(d - x^*) f(x^*) dx^* + \int_{x^*=d+\xi}^{\infty} -\pi\gamma f(x^*) dx^*,$$

Where $f(x^*)$ is the pdf of x_i^*. Note that $f(x^*)$ does not depend on γ by (A4.1). Differentiating the above with respect to d and applying $u(d - x_i^*) = \gamma$ at $x_i^* = d - \xi$ or $d + \xi$, we obtain

$$W_\gamma'(d) = \gamma(1 - \pi)[f(d - \xi) - f(d + \xi)] + \int_{x^*=d-\xi}^{d+\xi} u'(d - x^*)f(x^*)dx^* \tag{14}$$

This is an analogue of Proposition 2 for some fixed γ, with the added structure of (A4.2). When $\pi = 1$ (A4.4), the first term of this expression, which corresponds to the PA and AP groups, vanishes, leaving only the PP group, which we now split into those with $x^* < d$ and those with $x^* > d$:

$$W_\gamma'(d) = \int_{x^*=d-\xi}^{d} u'(d - x^*)f(x^*)dx^* + \int_{x^*=d}^{d+\xi} u'(d - x^*)f(x^*)dx^*. \tag{15}$$

Step 2: For any constant ζ, $f(d + \zeta) \geq f(d - \zeta) \iff x^m \geq d$.

Suppose $x^m \geq d$ and take a constant ζ. If $x^m > d + \zeta > d - \zeta$, the result immediately follows from the assumption in (A4.3) that $f()$ is single-peaked. If $d + \zeta \geq x^m \geq d > d - \zeta$ take a constant c such that $d + \zeta - x_m = x_m - c$. By symmetry about x^m, $f(c) = f(d + \zeta)$. We know that $c < x_m$, because $x_m - (d + \zeta) \leq 0$. We also know that $c \geq d - \zeta$, because we presumed $x_m \geq d$. We then have $x^m \geq c \geq d - \zeta$. The single-peaked assumption then implies $f(d + \zeta) = f(c) \geq f(d - \zeta)$.

Supposing $x^m < d$ and proceeding analogously proves the converse.

Step 3: $x^m \geq d \iff W_\gamma'(d) \geq 0$.

Starting from equation (15), note that by (A4.2) the first term is positive ($u' > 0$ when $x^* < d$) and the second term is negative ($u' < 0$ when $x^* < d$). We can compare the signs of the two terms in the previous expression by re-writing this equation, using the symmetry of the utility function, as:

$$W_\gamma'(d) = \int_{x^*=d-\xi}^{d} u'(d - x^*)[f(x^*) - f(\tilde{x})]dx^*$$

where $\tilde{x} = 2d - x^*$, so that $d - x^* = -(d - \tilde{x})$. We know from symmetry that when $d = x^m$, $f(x^*) = f(\tilde{x})$, so $W'(x^m) = 0$.

As $u'(d - x^*) > 0$ in the range of integration we use above. When $x^m > d$, the result in Step 3 implies that $f(x^*) \geq f(\tilde{x})$ for $x^* \in [d-\zeta, d]$, so we know that $W_\gamma'(d) \geq 0$. When $x^m < d$, the result in step 2 implies that $f(x^*) \leq f(\tilde{x})$ for $x^* \in [d-\zeta, d]$, and we know that $W_\gamma'(d) \leq 0$.

Step 3 proves that there is a unique global maximum of W at x^m.

Step 4: Setting $d = x^m$ minimizes opt-outs.

Let the frequency of opt-outs be given by $A(d) = P(a_i(d) > 0)$. Using $\xi = u^{-1}(\gamma)$ from before and letting F be the cdf of x_i^*, we know that

$$A(d) = F(d-\xi) + 1 - F(d+\xi)$$

Taking a derivative with respect to d, we have that

$$A'(d) = f(d-\xi) - f(d+\xi).$$

Setting $d = x^m$, it is straightforward to verify using (A4.3) that $A'(d) = 0$ if $d = x^m$, $A'(d) < 0$ if $d < x^m$, and $A'(d) > 0$ if $d > x^m$, which is sufficient to prove that x^m minimizes $A(d)$. ■

Proposition 5 *In the model with internalities, suppose that*

(A5.1) *For all i, $u_i(x) = -\frac{\alpha}{2}(x - x_i^a)^2$ with $\alpha > 0$.*

(A5.2) *Normative preferences are given similarly by $u_i(x) + m_i(x) = -\frac{\alpha}{2}(x - x_i^*)^2$.*

(A5.3) *The error in active choice $x_i^a - x_i^*$ is independent of x_i^a and γ_i.*

Then the marginal social welfare effect of a change in the default is given by $W_0'(d) + \mu X'(d)$, where $W_0(d)$ denotes social welfare without internalities (see Equation (6)), $\mu = E[\mu_i]$, and $X(d) = E[x_i(d)]$.

Proof Step 1: (A5.1) and (A5.2) imply that the internality $m(x)$ is linear.

Note that $u_i'' = -\alpha$ under (A5.1). By (A5.1) we can write

$$u_i(x) = \frac{u''(0)}{2}(x - x_i^a)^2.$$

By (A5.2) we can write

$$u_i(x) + m_i(x) = \frac{u''(0)}{2}(x - x_i^*)^2.$$

Subtracting the previous expression from this one and simplifying we obtain

$$m_i(x) = -u''(x_i^* - x_i^a)x + \frac{u''}{2}(x_i^{*2} - x_i^{a2}). \tag{16}$$

The second term is a constant with respect to x, and may therefore be safely ignored.

Step 2: Proving the result.

The result essentially follows from Equations (10) and the following equation from the text

$$\frac{\partial E[x_i(d)]}{\partial d} = E\left[x_i^a - d|PA\right]\ P(PA) + E\left[d - x_i^a|AP\right]\ P(AP) + P(PP) \tag{17}$$

Specifically, apply the linear internality to this equation to obtain:

$$\begin{aligned} W'(d) &= W_0'(d) + E\left[\mu_i(x_i^a - d)\,|\,PA\right]\ P(PA) \\ &- \quad E\left[\mu_i(x_i^a - d)\,|\,AP\right]\ P(AP) \\ &+ \quad E\left[\mu_i\,|\,PP\right]\ P(PP). \end{aligned}$$

Next, note that $\mu_i = m_i'(x) = -u''(x_i^* - x_i^a)$ by (16). Applying (A5.3) then implies that we can pull out the $E[\mu_i]$ terms.

$$W'(d) = W_0'(d) + \mu\{E[x^a - d|PA]P(PA) - E[x^a - d)|AP]P(AP) + P(PP)\}$$

Noting that the term inside curly brackets is the expression for $X'(d)$ in Equation (17), we obtain the desired result. ■

B Relationship to the Axiomatization of Masatlioglu and Ok (2005)

Masatlioglu and Ok (2005) provides an axiomatic characterization of a model very similar to the fixed as-if cost model we use. Their paper seeks to rationalize status quo bias; recall that we showed in Section 1.1 that giving extra utility to the status quo is the same as having a fixed cost of not choosing the status quo (see Section 1.1). The representation of choices used by Masatlioglu and Ok (see their equations (3) and (4)) is isomorphic to our own (see our equation (2), and Section 1.1), with one exception: the fixed as-if cost could depend on the default in their model. Whether and to what extent γ depends on d is difficult to test empirically, but we know of no evidence suggesting that it does. Nevertheless, here we discuss further the implications of our restriction that γ does not depend on d by relaxing it and examining welfare.

Consider a model that is identical to our baseline model except that the fixed cost is a function of d for each individual, denoted $\gamma_i(d)$. It is straightforward to show that the derivative from Proposition 2 becomes

$$\begin{aligned} 0 = W'(d) = & \quad E\left[\pi_i \gamma_i'(d) \right| AA] P(AA) \\ & + \ E\left[(1-\pi_i)\gamma_i(d) \right| PA] P(PA) \\ & - \ E\left[(1-\pi_i)\gamma_i(d) \right| AP] P(AP) \\ & + \quad E[u'(d)|PP]P(PP). \end{aligned} \tag{18}$$

This expression is identical to the expression in Proposition 2 except for the first term. In our basic model, individuals that are always active for a change in the default do not experience any change in their welfare. When the fixed costs depend on d and $\pi_i > 0$, changing the default can affect the welfare of these decision-makers because . The analogue of equation (5) is also straightforward to derive for this model.

First, we note that the argument in Proposition 3 (see the proof above) for active choices being optimal for sufficiently low π is unaffected by this addition. When as-if costs are not normative, forcing active choices still leads all individuals to receive x_i^* without incurring any costs. Whether forcing active choices *minimizes* welfare for sufficiently high π is unclear. The difficulty is that the penalty default d^p could in principle have a lower fixed cost ($\gamma(d^m)$) than other defaults, which can make the penalty default relatively more attractive than some other defaults.

We know by the same logic as Proposition 4 (proof above) that the last three terms of (18) will all be zero under (A3.1)-(A3.3) when we minimize opt-outs, and that ignoring the changes in $\gamma(d)$ for active choosers we would get to a global optimum by minimizing opt-outs when π_i is sufficiently high for all individuals. The additional term in Equation (18) therefore implies that minimizing opt-outs will not be optimal in general when the change in $\gamma(d)$ for a marginal change in the default is zero. Intuitively, if increasing the default from the opt-out minimizing default would reduce the cost incurred by active decision-makers, we know the aggregate effect on all other decision-makers is zero (by Proposition 3), so such an increase in the default would be an improvement on minimizing opt-outs. For a more extreme example, suppose there is a default d^* such that $\gamma_i(d^*) = 0$ for all i. Such a default is obviously the optimal default regardless of the π_i's.[28]

To summarize, our result that active choices are desirable when default effects are purely driven by behavioral frictions survives the extension implied by the model of Masatlioglu and Ok (2005). Minimizing opt-outs will still be a good rule of thumb when default effects are real costs and the dependence between the costs and the default is not too strong, but if the costs vary strongly with the default it may be possible to improve on the opt-out minimization rule of thumb.

C Variable Opt-Out Costs

Thus far we have assumed that as-if opt-out costs are constant (for a given individual) and do not depend on which non-default option the decision-maker selects. An alternative behavioral model is that defaults "pull" decision-makers towards options near the default in addition to making them more likely to select the default itself. For example, defaults may serve as an anchor (Example 1.2.7).

Ultimately, the question of whether defaults effects can be better described by including variable as-if costs in the model is an empirical question. Empirical evidence, reviewed in Section 1.2, regularly finds that increases in the default can affect choices far away from the default, suggesting that fixed costs are likely present. A variable costs model alone, such as a model of anchoring and adjustment where a higher default tends to lead to higher $x_i(d)$, would not predict, for example, that the fraction of individuals who contribute nothing to their pension would increase when the default rate of contribution is increased. Whether adding variable costs gives the model *additional* explanatory power relative to the fixed-cost-only model is more difficult to test. One possibility is to look closely at choices around the default. The fixed costs model with no variable cost predicts a "hole" in the observed

[28]When $\pi = 0$, both the active choice policy and the default d^* are optimal defaults.

distribution of choices around the default, whereas adding variable costs model predicts a "hill" around the default when fixed costs are sufficiently low. Still, given that both fixed and variable costs are plausibly heterogeneous, separately identifying these two components of decision-makers' revealed preferences without strong assumptions about distributions of the two costs is difficult. Here, we show how the inclusion of a variable costs affects the conclusions of our main analysis, especially the desirability of active choices versus minimizing opt-outs.

We focus on the case where X is a real interval. Suppose that instead of (1), individual behavior is given by

$$x_i(d) = \arg\max_{x \in X}\; u_i(x_i) - c_i(x_i - d) - \gamma 1\{x_i \neq d\}. \tag{19}$$

For simplicity, we will assume that u_i is single-peaked, with $u_i'(x_i^*) = 0$ and $u_i'' < 0$ everywhere. For this extension, we assume that the as-if cost associated with choosing a non-default option increases the further the chosen option is from d, so that $c_i'(x_i - d) \geq 0$ when $x_i - d > 0$, and $c_i'(x_i - d) \leq 0$ when $x_i - d < 0$. The as-if cost function is twice differentiable, with $c'' \geq 0$. We normalize $c(0)$ to zero. In this model individuals choose the default when passive, or $\tilde{x}(d) = \arg\max u_i(x_i) - c_i(x_i - d)$ when active. The individual is active if $\tilde{a}_i(d) \equiv [u_i(\tilde{x}_i(d)) - c_i(\tilde{x}_i(d) - d)] - u_i(d) - \gamma_i > 0$.

Similar to before, welfare is given by

$$w_i(x) = u_i(x) - \rho_i c_i(x - d) - \pi_i \gamma_i 1\{x_i \neq d\}, \tag{20}$$

where ρ_i denotes the normative relevance of variable costs $c_i(\cdot)$ and π_i the normative relevance of fixed costs as before. Indirect utility and social welfare are also defined similarly to before.

Given any change in the default, we can divide individuals into four groups as before, except now these groups are based on $\tilde{a}_i(d)$. Taking a derivative of the welfare function with respect to d, we have that the necessary condition from Proposition 2 becomes, with the addition of variable costs,

$$
\begin{aligned}
0 \;=\; W'(d) \;=\;& E\left[\rho_i c_i' + (1-\rho_i) c_i' \tfrac{c_i''}{c_i''-u_i''} \middle| AA\right] P(AA) \\
+\;& E\left[(1-\rho)c + (1-\pi_i)\gamma_i \middle| PA\right] P(PA) \\
-\;& E\left[(1-\rho)c + (1-\pi_i)\gamma_i \middle| AP\right] P(AP) \\
+\;& E[u'(d)|PP]P(PP).
\end{aligned}
\tag{21}
$$

where all components involving $c_i(\cdot)$ are evaluated at $x = \tilde{x}(d)$.

Adding variable costs changes this expression in two ways. First, the always-active choosers (AA) are affected by a change in the default. The sign of the welfare effect on an always-active chooser is positive if and only if $x_i^* < d$. For an individual with $x_i^* < d$, we will have that $x_i^* < x_i(d) < d$, and an increase in the default makes it costlier to choose an option close to x_i^*. The $\rho_i c_i'$ term of the welfare effect for members of the AA group in Equation 20 corresponds to the direct welfare effect of increasing this cost. Such an individual also increases x_i in response to this change in costs: it is straightforward to show that $\tilde{x}_i'(d) = \frac{c_i''}{c_i''-u_i''} \in [0, 1)$. The second term of the welfare effect for the AA group corresponds to the welfare impact of this change in behavior.[29] As before, when as-if costs are fully normatively relevant for all individuals, $\rho_i = 1$, and the envelope theorem eliminates the indirect welfare effect from the behavioral response. However, when $\rho_i < 1$, the individual over-reacts to the increase in costs, reducing their welfare. The opposite intuition applies when $x_i^* > d$; such individuals in the AA group are made better off by an increase in the default. The second addition to the welfare calculation is the extra variable cost incurred by marginally active decision-makers in the PA and AP groups. As it changes welfare discretely when the individual switches between choosing actively and choosing passively, this component affects welfare in exactly the same fashion as the fixed cost.

Our key result that forcing active choice is optimal when default effects are driven purely by behavioral frictions will still be true in this model, but properly examining an active choice policy requires subtle reasoning here. In this model, setting an extreme default so that everyone opts out will not necessarily be equivalent to forcing active choices directly.

[29]Note that the behavioral response is $\tilde{x}'(d) = 0$ when costs are linear, i.e. $c'' = 0$.

One might naturally suppose that when forcing active choices, the planner sets no anchor, which eliminates the variable costs, whereas when a penalty default acts as an anchor, the variable costs will matter for behavior and welfare. Suppose there is a policy that forces decision-makers to make active choices and eliminates variable costs (i.e. it does not set an anchor). It is straightforward to show that such a policy will be globally optimal when π_i is sufficiently small for all individuals (regardless of ρ_i), exactly as in Proposition 3. However, whether such a policy becomes extremely undesirable when π_i and ρ_i are sufficiently high is not clear in this model, because the policy that forces active choices also eliminates the variable costs and this can improve welfare. Conversely, a penalty default will surely minimize welfare when π_i and ρ_i are sufficiently high, but due to the large distortions on active choices it may have through the variable costs, it may not be optimal when π_i and ρ_i are large.

By a very similar procedure to the one we use in Proposition 4, one can show that minimizing opt-outs is optimal when π_i and ρ_i are sufficiently large, under some regularity conditions. Specifically, we could maintain Assumptions (A4.1)-(A4.3), and add the assumption that the variable cost function is the same for all individuals, $c_i(x_i - d) = c(x_i - d)$. Under these assumptions minimizing opt-outs will still be globally optimal when default effects are driven by real components of individual welfare.

D Additional Details from Empirical Application

This Appendix provides additional results for our empirical application. First, we show in Figure 4 how the marginal internalities from Figure 3 map to the mean optimal savings rate, $E[x_i^*]$. At $\mu = 0$, the mean savings rate corresponds to the observed savings rate when we simulate the model under the active choice policy, which is a 7 percent contribution not including the employer matching contributions, or just over 9 percent when we add in the match. As μ reaches larger values, the optimal savings rate increases and approaches the maximum 15% contribution asymptotically. Interestingly, as μ increases, the optimal default in Figure 3a approaches the maximum contribution more quickly than the mean optimal savings rate does. This finding might seem counter-intuitive at first, but it occurs because there is a mass of individuals contributing 15%, and the mass grows as μ increases. Like the 6% default before, this mass point is an attractive default because it gives a large number of people their exactly ideal option and it leads few people to opt out (see BFP Theorem 2).

Next, we show some results from two other employers, labeled "Company 1" and "Company 2" in BFP, along with additional details on the parameters used to calculate welfare in the model. In the model used here, the difference in distributions of contributions at different companies is used to identify differences in μ_ρ, which governs overall preferences over savings rates. Different companies also have different matching contribution rates. All other parameters are the same across companies. A complete table of parameter values is contained in Table 2.

Figure 5 repeats Figures 2 and 4 in the body of the text for Company 1. Figure 6 does the same for Company 2. We can see that apart from relatively minor differences, we obtain the same results for all three companies. The most noticeable difference is actually that the higher, 100 percent match rate at company 1 makes defaults lower than 6 percent much less desirable, which is intuitive.

Figure 4: Marginal Internalities and Mean Optimal Contribution Rates

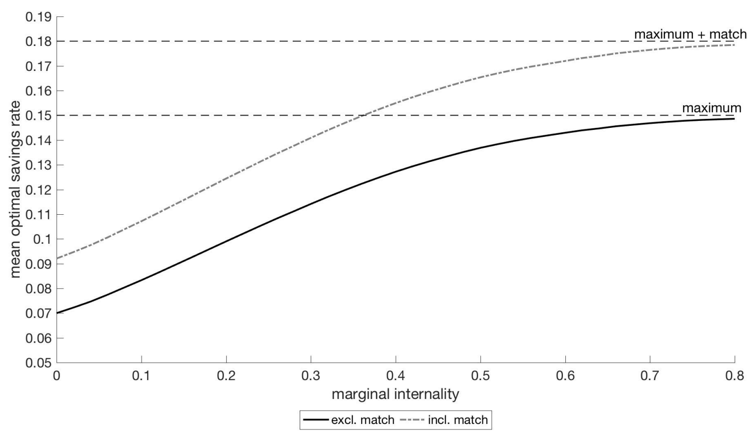
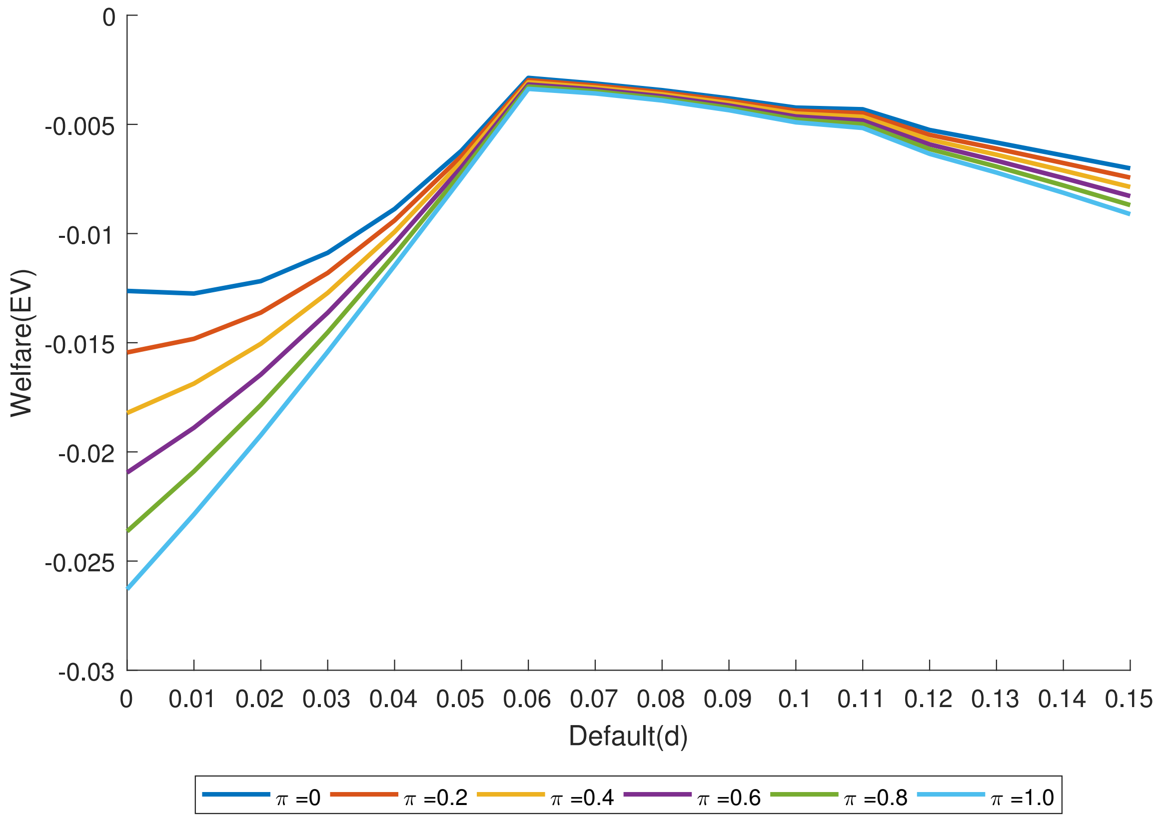


Figure 5: Results for Company 1

(a) Equivalent variation over defaults, by π



(b) Active Choices versus Minimizing Opt-Outs

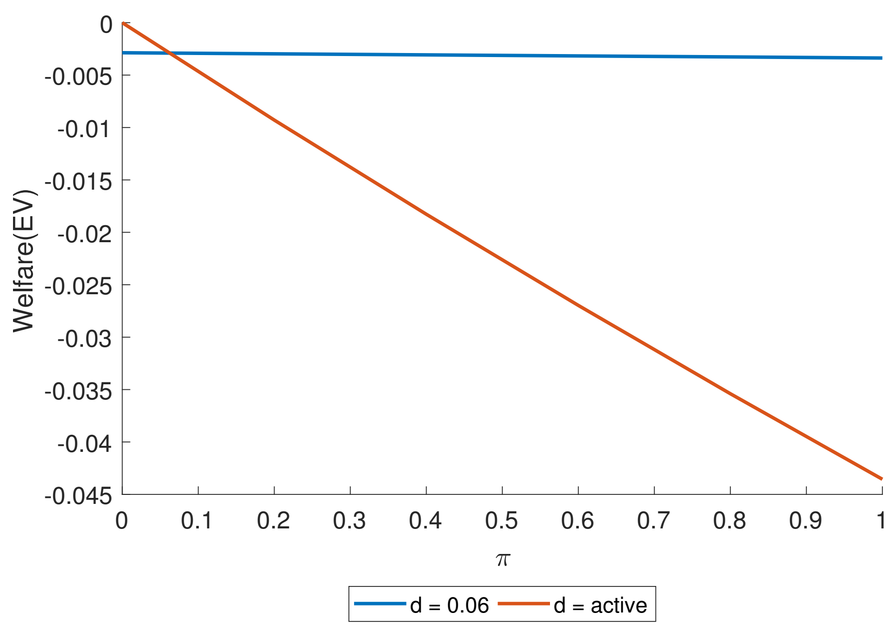
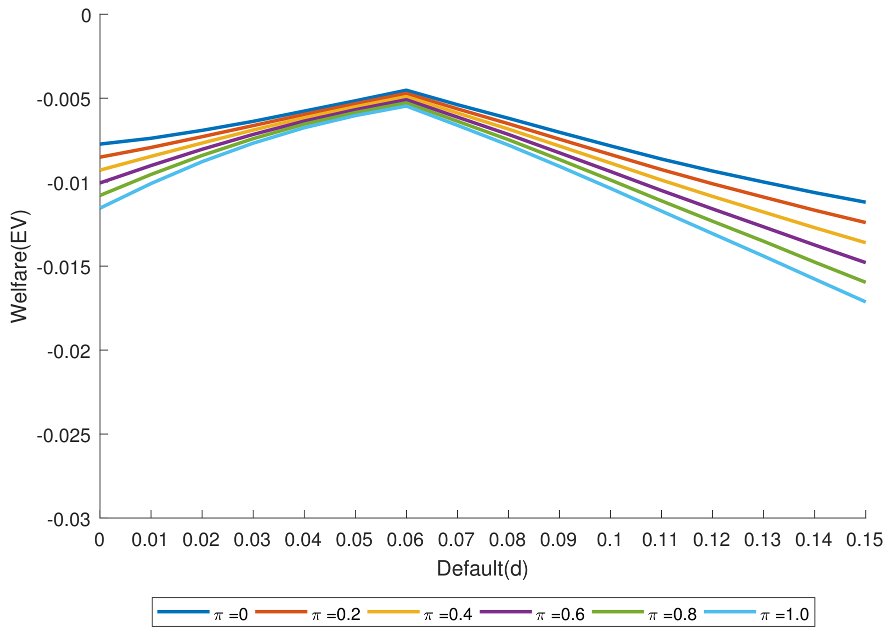


Figure 6: Results for Company 2

(a) Equivalent variation over defaults, by π



(b) Active Choices versus Minimizing Opt-Outs

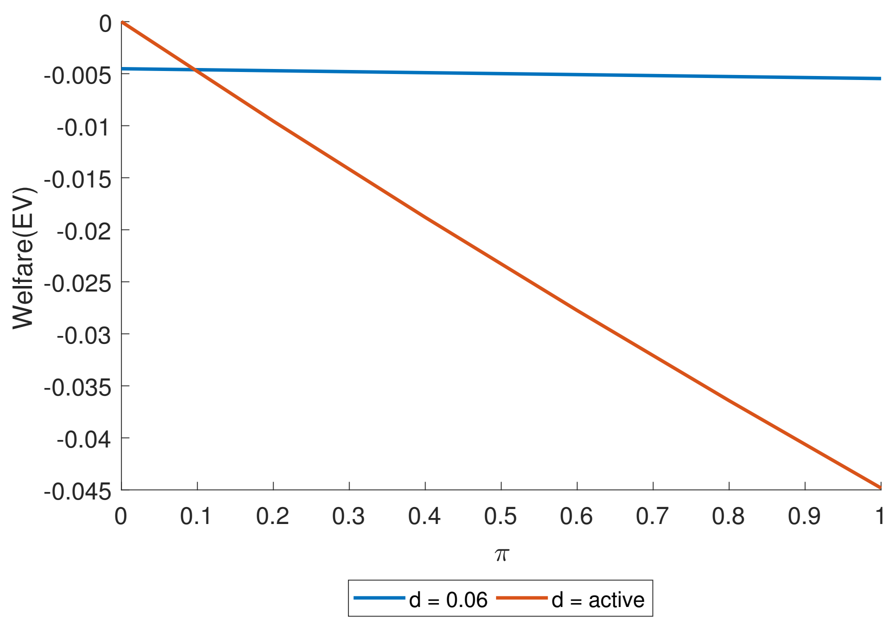


Table 2: Model Parameters and Plan Characteristics

Parameter	Value
Mean savings utility weight, μ_ρ, company 1	0.2150
Mean savings utility weight, μ_ρ, company 2	0.1313
Mean savings utility weight, μ_ρ, company 3	0.1570
Standard deviation of savings utility weight, σ	0.0910
Savings shift parameter, α	0.1340
Fraction with zero as-if costs, λ_1	0.4011
As-if costs distribution parameter, λ_2	11.81
Maximum matched contribution (all companies)	0.06
Employer match rate, company 1	1.0
Employer match rate, company 2	0.5
Employer match rate, company 3	0.5

Note: this table reports the parameter values we use in our empirical illustration. The parameter values come from Table 2 of Bernheim, Fradkin and Popov (2015), for the "basic model."