NBER WORKING PAPER SERIES

THE WELFARE ECONOMICS OF DEFAULT OPTIONS IN 401(K) PLANS

B. Douglas Bernheim
Andrey Fradkin
Igor Popov

Working Paper 17587
http://www.nber.org/papers/w17587

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2011

Previously circulated as "The Welfare Economics of Default Options: A Theoretical and Empirical Analysis of 401(k) Plans." We would like to thank participants at the 2010 CESifo Venice Summer Institute Conference on Behavioural Welfare Economics, the 2011 ECORE Summer School (UCL, Louvain-la-Neuve), the June 2011 D-TEA Paris Meetings, the 2012 ASSA Winter Meetings (Chicago), the Public Economics Seminar at UC Berkeley, and the PIER Seminar at the University of Pennsylvania, as well as anonymous referees, for helpful comments. The first author has benefited immeasurably from numerous conversations with Antonio Rangel concerning the topic of behavioral welfare economics, which have spanned many years. The first author also acknowledges financial support from the National Science Foundation through grants SES-0752854 and SES-1156263. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Welfare Economics of Default Options in 401(k) Plans
B. Douglas Bernheim, Andrey Fradkin, and Igor Popov
NBER Working Paper No. 17587
November 2011, Revised October 2014
JEL No. D03,D14,D60,D91,J26

ABSTRACT

Default contribution rates for 401(k) pension plans powerfully influence workers' choices. Potential causes include opt-out costs, procrastination, inattention, and psychological anchoring. We examine the welfare implications of defaults under each of these theories. We show how the optimal default, the magnitude of the welfare effects, and the degree of normative ambiguity depend on the behavioral model, the scope of the choice domain deemed welfare-relevant, the use of penalties for passive choice, and other 401(k) plan features. Depending on which theory and welfare perspective one adopts, virtually any default contribution rate may be optimal. Still, our analysis provides reasonably robust justifications for setting the default either at the highest contribution rate matched by the employer or – contrary to common wisdom – at zero. We also identify the types of empirical evidence needed to determine which case is applicable.

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

Andrey Fradkin
National Bureau of Economic Research
1050 Massachusetts Ave.
Cambridge, MA 02138
afradkin@gmail.com

Igor Popov
Department of Economics
Stanford University
Stanford, CA 94305-6072
iapopov@stanford.edu

A data appendix is available at:
http://www.nber.org/data-appendix/w17587

1 Introduction

Starting with Madrian and Shea (2001), several studies have found that changing the default contribution rate for a 401(k) pension plan has a powerful effect on employees' contributions,[1] particularly compared with conventional policy instruments such as capital income taxes. Yet default provisions have received far less attention and, with few exceptions, the critical task of evaluating their *welfare effects* has been almost entirely ignored. That task poses two types of conceptual challenges. First, the cognitive mechanisms behind default effects are poorly understood, and there are competing explanations. Second, most explanations involve non-standard theories that render traditional normative tools inapplicable.

This paper analyzes the welfare effects of 401(k) default contribution options quantitatively, using reasonably parameterized models fit to data reflecting responses on the critical behavioral margins. To our knowledge, it is the first study to provide practical guidance concerning both the normative importance of default options and the nature of welfare-optimal policies. We consider multiple theories of default effects involving opt-out costs, sophisticated and naive time inconsistency, inattentiveness, and psychological anchoring. To accommodate the non-standard elements of these theories, we employ Bernheim and Rangel's (2009) framework for behavioral welfare analysis. In that framework, inconsistencies in choice translate into a quantifiable degree of normative ambiguity, which one can either accept or reduce/resolve by refining the set of choices deemed welfare relevant.

While welfare is our main focus, our estimated models are of independent interest. In a conventional model, unrealistically large opt-out costs (averaging thousands of dollars) are required to rationalize default effects. Non-standard (behavioral) theories potentially resolve this puzzle, and the data appear to favor an explanation involving anchoring effects.

For models with frame-dependent weighting (i.e., the two flavors of time inconsistency plus inattentiveness), we obtain five main findings. First, even if one treats all decision

[1] See also Choi et. al (2002, 2003, 2003, 2006), Beshears et. al. (2008), and Carroll et. al. (2009). Bronchetti et. al. (2011) describe a related context in which no default effect is observed.

frames as welfare-relevant, the degree of ambiguity concerning the normative effects of default rates is small over the pertinent range. This is surprising because as-if opt-out costs average thousands of dollars in the "naturally occurring" decision frame, and different welfare perspectives discount those costs to widely differing degrees. For our model of sophisticated time-inconsistency, the explanation is that as-if opt-out costs are small on average among the workers who actually incur those costs by opting out; hence, discounting *incurred* opt-out costs to a greater or lesser degree makes relatively little difference. For the other models, different explanations apply.

Second, the welfare-optimal default rate tends to coincide with the cap on employer matching contributions. The match cap induces a convex kink-point in the workers' opportunity sets, and hence creates a point of accumulation in the distribution of ideal contribution rates. When that effect is large, incurred opt-out costs dominate other considerations, and minimizing the opt-out frequency (a rule of thumb advocated by Thaler and Sunstein, 2003) by setting the default equal to the match cap maximizes welfare. In contrast, without matching provisions, optimal default rates tend toward the center of the distribution of worker preferences, despite theoretical reasons to anticipate that the optimum would lie either at the lowest or highest possible contribution rate (again because those are points of accumulation in the distribution of ideal contribution rates).

Third, when a 401(k) plan includes a generous employer match, the welfare stakes are substantial. The loss from setting a default rate at zero rather than at the welfare-optimal rate can run as high as two or more percent of earnings, representing a substantial fraction of the potential surplus generated by the 401(k) plan. Without matching provisions, the stakes are much smaller. Following the Thaler-Sunstein opt-out-minimization criterion yields small welfare losses even when it is suboptimal; hence it is a reasonable rule of thumb.

Fourth, we investigate the use of penalties for passive choice. Previous theoretical research has shown that it is sometimes best to compel active decision making through a large penalty or an extreme default. We ask whether defaults and penalties for passive choices should be used *in combination*, for example, by setting a moderate penalty along with an attractive default. We find that welfare is a double-peaked function of the size of the penalty, so that the optimum *either* involves an attractive default with no penalty, *or* a penalty so extreme it renders the default virtually irrelevant.

Fifth, we investigate the effects of policies that alter the context of opt-out choices. For models of time inconsistency, we examine precommitment opportunities that would enable workers to decide, in advance, whether to compel active choice. The potential benefits of such opportunities is a theme in the literature on time inconsistency, which often assumes that the forward-looking perspective is normatively "correct." Recognizing the validity of other perspectives, we show that precommitment opportunities have a previously unrecognized down-side: they create *substantial ambiguity* concerning the welfare effects of default policies. Intuitively, workers will commit to opting out even if they subsequently perceive enormous costs in the moment. Surprisingly, precommitment opportunities have the *opposite* effect on welfare ambiguity if workers exhibit naive time inconsistency, in that they virtually eliminate welfare losses in *all* evaluation frames, not just the forward-looking ones. Thus, assuming workers are naively time-inconsistent, the case for precommitment opportunities is especially strong, even if the correct frame of evaluation is unclear.

Related questions arise in the context of models with inattention: if one takes them literally and evaluates welfare from the fully attentive perspective, the best policy is plainly one that maximizes attentiveness. Surprisingly, our analysis points to the same prescription even if one instead remains agnostic about the cognitive processes underlying choice, and hence about the correct frame for evaluation.

In contrast, for models with anchoring, unless one restricts the welfare-relevant domain, the degree of normative ambiguity is substantial, and welfare analysis is only modestly informative. A possible restriction is to evaluate outcomes from the perspective of a "neutral" frame; i.e., one in which choices would be free from the effects of anchors. (To be clear, that perspective is then used to evaluate choices in all frames, including those for which anchors are present.) We find that aggregate worker welfare evaluated from the perspective of the neutral frame does not vary much with the default rate. Because higher default rates increase contributions and thereby create costs for employers and the government, it follows that the socially optimal default rate is zero.

Our findings concerning welfare are therefore conditional: they depend both on the behavioral model and, to varying degrees, on the decision frame(s) deemed welfare-relevant. We emphasize that one could in principle resolve these ambiguities through additional empirical investigation. We do not attempt such resolutions here because the required data are currently unavailable. However, we set the stage for such analyses by highlighting the models' empirically distinguishable implications concerning frame dependence, and by clarifying the types of evidence concerning cognition that might provide objective rationales for evaluating welfare from the perspective of one decision frame rather than another.

Thaler and Sunstein (2003) were the first to comment on the welfare effects of default options, though not in the context of a formal model. They proposed that companies should set defaults to minimize opt-out frequencies, offering as justification a principle of *ex post* validation. As noted above, our findings shed light on the performance of that rule of thumb. Only one prior study has addressed these issues formally: Carroll, Choi, Laibson, Madrian, and Metrick (2009), henceforth CCLMM, who assume that default effects arise from procrastination by sophisticated time-inconsistent workers.[2] They also adopt a particular perspective on welfare – that "true well-being" is governed by "long-run" preferences. In their setting, with a high degree of time inconsistency, the optimal policy is to force active decisions, e.g., by setting an extreme default contribution rate. With a low degree of time inconsistency, it is better either to set the default at the center of the distribution of preferred savings rates or to skew it toward either end of that distribution, depending on whether population heterogeneity with respect to desired saving is low or high, respectively.

While CCLMM's analysis is an important first step toward understanding the welfare

[2] The working paper version of CCLMM also studied naive time-inconsistent workers.

effects of default options, it is limited in several respects. First, it is not quantitative. It enumerates several possibilities but provides no guidance as to which applies in practice; nor does it gauge the the welfare costs associated with suboptimal defaults. Second, it examines only a single behavioral theory of default effects. Because opt-out costs alone can generate such effects, a non-standard theory may not be needed, and if one is needed, considerations other than time inconsistency may come into play. Third, as noted above, CCLMM adopt a single welfare perspective, the "long-run criterion." That choice is controversial (see Bernheim, 2009). Those who favor it argue that it reflects the decision maker's true preference purged of "present bias." Yet people may overintellectualize temporally distant choices and properly appreciate experiences only "in the moment." Fourth, CCLMM's simple model omits factors that may significantly impact optimal default rates, such as caps on employer matching contributions and bounds on employee contributions.[3] Fifth, CCLMM do not examine some interesting policy alternatives, such as the combined use of defaults and penalties for passive choice. The current paper addresses each of these limitations.

In the next section, we put forth our framework for analysis, including models of default effects and welfare criteria. Section 3 explains how we parametrized the models. Section 4 uses the models to investigate welfare, and provides some theoretical results that illuminate and extend our numerical findings. Section 5 provides some concluding remarks. Proofs of theorems and other supplemental materials appear in an online Appendix.

2 Analytic framework

2.1 The basic model with costly opt-out

We use x to stand for the *total contribution rate* of a worker newly eligible to participate in a 401(k) plan; it equals the sum of employer and employee contributions, divided by earnings (exclusive of the employer contribution), and lies between 0 and some maximum, $\overline{x}$. The

[3] These factors create points of accumulation in the distribution of ideal contribution rates. CCLMM explicitly assume that the distribution of ideal contribution rates is atomless.

plan's default provisions imply a total contribution rate of d.

We focus on the worker's initial ("period 0") choice between accepting the default and opting out to some $x \neq d$. This choice matters for three reasons: (1) opt-out entails costly effort (e); (2) x determines current 401(k) saving and the default for the next period;[4] and (3) x determines the amount of residual cash, z, available for near-term consumption and non-401(k) saving. Normalizing the worker's total earnings to unity, we write:

$$z = 1 - \tau(x), \tag{1}$$

where τ captures employer matching provisions and the tax deductibility of contributions. Usually, τ is an increasing, piecewise-linear function with one or more convex kink-points at the values of x that exhaust the employer match or move the worker between tax brackets.[5]

We assume that, in period 0, the individual acts as if he maximizes the utility function

$$u(e) + V(x, z) \tag{2}$$

The function $u(\cdot)$ captures the disutility of effort, e. As a normalization, we assume $u(0) = 0$. One can think of V as a state evaluation function: it accounts for the effects of current 401(k) saving (x), the default contribution rate for the next period (also x), and residual cash (z) on future consumption and hence continuation utility. For notational simplicity, we suppress the dependence of u and V on parameters pertaining to preferences and conditions constraining future choices (such as initial assets and future interest rates), except where it is important to be explicit.

Although our depiction of the worker's decision problem may strike the reader as static, we construe it as dynamic, precisely because V serves as a "reduced form" that encompasses the effects on utility of current choices through their impact on all subsequent decisions; see the Appendix for formal details. Obviously, one cannot use a reduced-form utility function

[4] In principle, these two effects are separable (e.g., upon electing a contribution rate of 3%, the default for the next period could change to 4%), but in practice they always go hand-in-hand (in the same example, the new default would be 3%).

[5] It is also natural to assume that $\tau(0) = 0$ and $\tau(\overline{x}) < 1$.

to analyze the effects of an intervention that would be expected to change the reduced form, but we do not encounter that problem. Fixing the initial contribution rate, x, opportunities after period 0 do not depend upon the initial default d. Because our formulation captures the dependence of V on x, a change in d leaves V unaffected. This observation simplifies our task because it means we can estimate V and treat it as a fixed utility function; there is no need to estimate underlying intertemporal preferences using data on consumption trajectories.

We assume that opting out entails effort e', and write the resulting disutility as $\gamma \equiv -u(e') < 0$. In period 0, the worker chooses x to maximize (2) subject to (1), plus the additional constraint that $e = 0$ for $x = d$, and $e = e'$ for $x \neq d$. To find the solution, we first solve for the worker's "ideal point," x^*, by maximizing (2) subject to (1), ignoring the opt-out costs. The gain in the worker's continuation utility from choosing x^* rather than d is given by

$$V(x^*, 1-\tau(x^*)) - V(d, 1-\tau(d)) =: \Delta(d)$$

The worker opts out if and only if that gain exceeds the effort cost:

$$\Delta(d) \geq \gamma. \tag{3}$$

2.2 Models with frame-dependent weighting

Next we examine a class of models characterized by *frame-dependent weighting*, in which the worker acts as if he places greater relative weight on opt-out costs in some psychological decision frames than in others.[6] All of these models assume that, conditional on opting out, the worker maximizes $V(x, z)$ subject to $z = 1 - \tau(x)$, and therefore chooses x^*. However, instead of (3), the opt-out condition is

$$\Delta(d) \geq D(f)\gamma, \tag{4}$$

[6] In behavioral economics, the phrase "decision frame" refers to an aspect of a decision problem that may affect what is chosen without altering the chooser's opportunity set.

where f denotes the decision frame, and $D(f)$ is a frame-dependent weight.[7] For each of the models described below, we assume that, with existing institutional arrangements, workers normally make opt-out decisions in a *naturally occuring* decision frame, f^*.

Sophisticated time inconsistency. To introduce time inconsistency, we assume that a worker makes the opt-out decision either "in the moment," which we call the *contemporaneous frame*, $f = 0$, or in advance (as a commitment), which we call the *forward-looking frame*, $f = -1$. *Sophistication* means that he correctly anticipates his future actions and properly assesses the continuation value function. As in the standard model of *quasihyperbolic discounting* (see, e.g., Laibson, 1997), we assume the worker attaches the weight $\beta \in (0,1)$ to all future consequences, maximizing $\beta\left[u(e) + V(x,y)\right]$ in the forward-looking frame, but maximizing $u(e) + \beta V(x,y)$ in the contemporaneous frame. The opt-out condition for this model corresponds to (4), with $D(-1) = 1$ and $D(0) = \beta^{-1}$.

For existing institutions, the contemporaneous frame is naturally occurring ($f^* = 0$). If this model is correct, the frequency with which workers opt out should differ if, instead, they make the decision in advance ($f = -1$). We know of no direct evidence on that point.

Naive time inconsistency. We introduce naive time inconsistency by assuming that a worker's choice depends on two aspects of the decision frame: first, whether it is contemporaneous or forward-looking; second, the degree of sophistication it elicits. If a worker *correctly* anticipated his near-term opt-out choices, he would assess his continuation value as $V(x^*, 1 - \tau(x^*))$ when opting out from d to x^*, and as $V(d, 1 - \tau(d))$ when sticking with the default. A naive worker is overly optimistic about his subsequent actions: he places the weight $\kappa(f) \in [0,1]$ on the continuation value he would receive if he subsequently switched to x^* after a very brief delay (when optimal according to his *current* standards), and the weight $1 - \kappa(f)$ on his actual continuation value. Thus, he acts as if his continuation value

[7] Implicit in this formulation is the assumption that the initial period 0 frame, f, does not affect the decisions after period 0, so that V does not depend on f. That is, the *direct* psychological influence of the *initial* frame is temporary: it may influence the period 0 allocation between x and z, but not subsequent choices given (x,z).

from choosing the default is

$$\kappa(f)\left[\max\{V(x^*,1-\tau(x^*))-\gamma,V(d,1-\tau(d))\}\right]+(1-\kappa(f))V(d,1-\tau(d)).$$

The parameter $\kappa(f)$ measures the worker's degree of naivete.[8] With $\kappa(f)=1$, the worker is always certain he will reoptimize after minimal delay. For $\kappa(f)=0$, we have sophisticated time inconsistency. We will assume there is, in principle, some way to frame the decision problem so that actual future consequences are made explicit and transparent to the worker, in which case $\kappa(f)=0$ for that frame.[9]

With this formulation of naivete, the opt-out condition corresponds to (4), with

$$D(f)=\frac{\beta^{-1}-\kappa(f)}{1-\kappa(f)}$$

if f entails making the opt-out decision contemporaneously, and $D(f)=1$ if f entails making that decision in advance (see the Appendix). For any given value of β, naifs and sophisticates are equally inclined to opt-out when making decisions in advance, but naifs are less inclined to opt-out contemporaneously (because $(\beta^{-1}-\kappa(f))\,/\,(1-\kappa(f))>\beta^{-1}$). With contemporaneous framing, if $\kappa(f)$ is close to unity (high naivete), $D(f)$ may be extremely large even if β^{-1} is not.

We assume the naturally occurring frame for existing institutions is naivete-promoting and contemporaneous. If this model is correct, then the frequency with which workers opt out should be higher if, instead, they make the decision in advance or under conditions that render the future course of action transparent. We know of no direct evidence on that point.

Inattentiveness. To model inattention, we assume the worker behaves as if he attends to the task of selecting a 401(k) contribution rate if and only if the choice is sufficiently consequential, in the sense that the stakes exceed some threshold, $\chi(f)$, which may depend

[8] This formulation is related to the notion of a *partially naive hyperbolic agent*; see, e.g., O'Donoghue and Rabin (2001) and Della Vigna and Malmendier (2004).

[9] For example, imagine a decision frame in which responsibility for all future actions is transferred to an automaton programmed to act exactly as the worker would act, and the worker is provided with a detailed and accurate account of the automaton's choice mapping. More generally, Bernheim (2014) argues that any case of "biased beliefs" implicitly invokes this type of frame dependence.

on the frame.[10] If he does not attend, he ends up with a "status quo" bundle.[11] Because the identity of the status quo may affect the outcome even when it has no effect on the worker's opportunity set, we treat it as part of the decision frame, f.

With naturally occurring institutions, the default contribution rate, d, governs the status quo. However, it would be improper to treat d as an aspect of the decision frame, because it also affects the worker's opportunity set by determining the effort required to obtain any given contribution rate. Thus we distinguish between the status quo, which determines the outcome if the worker fails to attend, and the default, which determines which outcomes require effort. This distinction is not merely conceptual: fixing any given status quo, it is possible to vary the schedule relating options to required effort, for example by selectively adding or removing red tape depending on which alternative the worker wishes to elect.

According to our model of inattentiveness, the worker attends and opts out if and only if

$$\Delta(d) \geq \chi(f) + \gamma, \tag{5}$$

To put this inequality in the form of condition (4) simply take $D(f) = \frac{\chi(f)}{\gamma} + 1$.

We assume an inattentive decision frame is naturally occurring for existing institutions. If this model is correct, the frequency with which workers opt out should be higher if, instead, choice framing draws their attention to retirement planning. The evidence on that point is both limited and mixed.[12] Choices should also respond to changes in the status quo, even when the schedule relating options to required effort is held fixed. That hypothesis is testable, but we know of no evidence that speaks to it.

[10] Our approach to inattention contrasts with that of Sims (2003) and the literature that followed from his work, in that we do not model inattention as a rational response to information processing constraints.

[11] In the context of our model, the term "bundle" refers to a vector (x, z, e), specifying 401(k) contributions, income not contributed to the worker's 401(k) plan, and effort.

[12] According to Carroll et. al. (2009), a survey of unenrolled workers that drew attention to 401(k) issues did not increase enrollment among those who responded. Yet Karlan et. al. (2010) show that saving decisions are sensitive to attentiveness manipulations in a related context.

2.3 A model with anchoring

A default may also influence decisions through the power of suggestion; it may, for example, provide a salient starting point for a worker's thinking,[13] or a perceived "stamp of approval." To model this mechanism, we assume the default d not only impacts the worker's opportunity set as above, but also establishes a frame, $f = d$, involving a *psychological anchor* that inclines him toward choosing $x = f$. Thus, as in our model of inattention, the default plays two roles that are in principle separable (for the same reasons), only one of which is properly considered a framing effect.

Formally, the worker acts as if he maximizes[14]

$$u(e) + V(x, z, f) \tag{6}$$

for $f \in [0, \overline{x}]$.[15] Plainly, the worker's associated ideal point, $x^*(f)$, now depends on the frame, as does Δ. Otherwise, (3) still governs the opt-out decision.

Here, the naturally occurring frame corresponds to the employer's default contribution rate. If this model is correct, choices should respond to changes in the psychological anchor, even when the schedule relating options to required effort is held fixed. That hypothesis is testable, but we know of no evidence that speaks to it. In Section 3, we separate the framing and opportunity-set effects empirically through additional identifying assumptions.

2.4 Welfare framework

We use the framework for behavioral welfare analysis developed by Bernheim and Rangel (2009), henceforth BR (see also Bernheim, 2009, 2014).[16] Within this framework, one

[13] A series of studies have documented the importance of anchoring effects in the laboratory; see, for example, Ariely, Loewenstein, and Prelec (2003).

[14] This formulation is not meant to suggest that the default directly affects "true well-being." Indeed, comparisons of $V(x, z, f)$ and $V(x', z', f')$ are meaningful only if $f = f'$ (because f merely parameterizes ordinal preferences over (e, x, z) bundles).We interpret (6) simply as an analytic device for recapitulating the dependence of a choice mapping on a decision frame f.

[15] In principle, one could allow for negative or arbitrarily large default frames, even though these are not institutionally permissible. However, if sufficiently extreme defaults would have no marginal influence on choice, the bounds are inconsequential.

[16] When applying the BR framework to a particular model, we limit consideration to the choice domain encompassed by the model. Stepping outside the domain of the model, behavior may exhibit other non-

derives a quantitative welfare criterion directly from the choice mapping, which summarizes an individual's selections from all possible opportunity sets, conditional on framing.

Welfare-relevant choices. In the BR framework, one starts by "pruning" the domain of the choice mapping, eliminating choices that are not deemed welfare-relevant. (The criterion can then be applied to all choice situations, including ones that were pruned, and hence played no rule in the criterion's construction.) To avoid paternalistic judgments, BR advocate limiting such pruning to choices that are demonstrable mistakes, in the sense that the decision maker misunderstands the available options, an occurrence Bernheim (2009, 2014) calls "characterization failure." To understand the logic of pruning, suppose someone must choose between x and y. In frame A, he correctly recognizes x and y, and chooses y. In frame B, he mistakes y for z, and chooses x. Only the first of these choices is a suitable guide for a policy maker choosing between x and y on his behalf.

The welfare criterion. If an individual's welfare-relevant choices are internally consistent,[17] we can proceed as if they reveal "true preferences." However, there may be no objective basis (or only a controversial one) for resolving inconsistencies by limiting the welfare-relevant domain. An important advantage of the BR framework is that it permits one to conduct quantitative welfare analysis even in those cases. The framework evaluates welfare according to P^*, the *unambiguous choice relation*: xP^*y iff y is never chosen when x is available. BR argue that any choice-based welfare criterion should have a particular set of properties, and show that this one uniquely meets that requirement. Welfare analysis involving P^* exploits the coherent aspects of choice that are present in virtually all behavioral models, while expressing the incoherent aspects of choice as ambiguity (incompleteness).

Under the following restrictive conditions, applying P^* is equivalent to treating the decision maker as a collection of individuals, one for each decision frame, and using a "multi-self Pareto criterion": (a) the welfare-relevant domain is the Cartesian product of the collection

standard patterns; e.g., the worker might exhibit a general rather than context-specific tendency to make present-biased choices. We acknowledge that consideration of all non-standard choice patterns on an unlimited choice domain would yield greater normative ambiguity.

[17] By "internally consistent," we mean that they satisfy the Weak Axiom of Revealed Preference.

of possible choice sets and a set of possible frames, and (b) the individual behaves as if he maximizes some well-behaved utility function within each frame (see Bernheim and Rangel, 2009, Theorem 3). These two requirements, which we call the *multi-self conditions*, are satisfied for the models of time inconsistency and anchoring described above,[18] but not for our model of inattention.[19] When they are satisfied, each decision frame offers a comprehensive and coherent perspective on welfare, and the "best" choice from the perspective of any welfare-relevant frame is unimprovable according to P^*.

Aggregate equivalent variation. The BR framework yields a generalized notion of equivalent variation, which accommodates the normative ambiguity associated with internally inconsistent choice patterns as follows. For a change from policy p to p', EV_A is the smallest increment to income with p such that the bundle obtained with p is unambiguously chosen over the bundle obtained with p'. Similarly, EV_B is the largest increment to income with p 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_B and no more than EV_A.

In this study, we aggregate EV_A and EV_B over workers. Aggregation is valid in standard welfare economics: because equivalent variation is a monotonic transformation of utility, one can find Pareto optima by maximizing the weighted sum of EVs.[20] Maximizing the *simple* sum of EVs treats a dollar as equally valuable no matter who receives it. The following new result, which we use later, shows that these principles generalize as long as the unambiguous choice relation is transitive (which it is for the models considered here):

[18] Condition (a) is not, however, satisfied more generally for models of time inconsistency. One cannot pair a frame specifying that all discretion is exercised at time t with opportunity sets involving distinct consumption alternatives prior to t. The models of time inconsistency described in Section 2.2 avoid that consideration only because nothing is consumed in period -1.

[19] Condition (a) is not satisfied because one cannot pair a decision frame that establishes a particular alternative as the status quo with opportunity sets that exclude that alternative.

[20] If the opportunity set is not lower hemicontinuous in the amount of compensation, then EV need not be a *strictly* monotonic transformation of utility. In that case, the set of alternatives that maximize aggregate EV contains at least one Pareto optimum, but all the maximizers need not be Pareto optima. An analogous technical qualification appears in Theorem 1.

Theorem 1: *Suppose P^* is transitive. Consider any non-negative weights λ_{Ai} and λ_{Bi} for all individuals i such that $\sum_i (\lambda_{Ai} + \lambda_{Bi}) = 1$. Let X_M denote the set of alternatives that maximize $\sum_i (\lambda_{Ai} EV_{Ai} + \lambda_{Bi} EV_{Bi})$ within a set X. Then at least one element of X_M is a weak generalized Pareto optimum within X.*[21]

2.5 Application of the welfare framework to the models

To measure an equivalent variation, one must specify the baseline environment in which the equalizing compensation is received. Throughout, we take that environment to be one in which each worker obtains, at no cost, his ideal point, x^*.[22] With this baseline, equivalent variations are often *negative*, which means they measure efficiency losses.

For each of our models, the first step in calculating EV_A and EV_B is to evaluate equivalent variation from the perspective of an arbitrary frame f, which may differ from the decision frame. Then one maximizes (for EV_A) or minimizes (for EV_B) equivalent variation over the frames deemed welfare-relevant.

Time inconsistency. To avoid repetition, we allow for naivete from the start and treat sophistication as a special case. From the perspective of frame f, the equivalent variation is the value of m that satisfies

$$\begin{aligned}\beta V(x^*, 1 + m - \tau(x^*)) = \beta\kappa(f) \max\{V(x^*, 1 - \tau(x^*)) - \gamma, V(d, 1 - \tau(d))\} \\ + \beta(1 - \kappa(f))V(d, 1 - \tau(d))\end{aligned} \tag{7}$$

[21] In the BR framework, x is said to be a weak generalized Pareto optimum in X if there is no y in X such that $yP_i^* x$ for all individuals i. To be clear, the sets that maximize weighted sums of the form $\sum_i (\lambda_{Ai} EV_{Ai} + \lambda_{Bi} EV_{Bi})$ may not contain all weak generalized Pareto optima. To illustrate, suppose the multiself conditions are satisfied. Then the set of alternatives that maximize any weighted sum of equivalent variations evaluated in any welfare-relevant decision frame contains a weak generalized Pareto optimum. If a given frame is not used to evaluate EV_A or EV_B, the "best" choice from the perspective of that frame may be unrelated to the choices that maximize any weighted sum of EV_As and EV_Bs.

[22] Two clarifying remarks are in order. First, throughout our analysis, we hold the baseline contribution rate fixed at x^* as we vary the equalizing compensation, even though the worker's ideal point actually changes. As an alternative, we have also calculated equivalent variations for a baseline environment in which each worker can costlessly elect his ideal point taking the equalizing compensation into account. The computations are far more involved, but the results are virtually identical. Second, for our anchoring model (in which x^* depends on f), we must specify the baseline frame that determines x^*.

if the worker chooses the default, and

$$\beta V(x^*, 1 + m - \tau(x^*)) = \beta V(x^*, 1 - \tau(x^*)) - b(f)\gamma \tag{8}$$

if he opts out, where $b(f) = 1$ if f involves a contemporaneous perspective, and $b(f) = \beta$ if f involves a forward-looking perspective.[23]

Some studies advocate evaluating welfare based solely on forward-looking choices, on the grounds that people suffer from "present bias" and "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, the forward-looking frame is the problematic one.[24] Absent an objective basis for adjudicating between these perspectives, there is an argument for remaining agnostic and respecting both.

There is also a case for evaluating welfare based solely on sophisticated choices (those with $\kappa(f) = 0$), on the grounds that naive choices involve characterization failure. However, caution is warranted. Models are simply lenses through which we interpret and rationalize choice patterns, and a variety of models can usually account for the same patterns. If we treat our model of naivete as an as-if representation that may happen to fit the choice data rather than as a literal depiction of cognitive processes, the argument for ignoring supposedly naive choices is no longer compelling. One may then wish to remain agnostic and respect all perspectives, regardless of how our model labels them.

Suppose we deem *sophisticated forward-looking choices* (and nothing else) welfare relevant. Then we evaluate EV using (7) and (8) with $f = -1$ (so that $b(f) = \beta$) and $\kappa(f) = 0$. Because β factors out of both formulas, they are the same as those used to compute EV for the basic model.

[23] These equations define EV because the worker would choose the bundle $(0, x^*, 1 + m - \tau(x^*))$ over, respectively, $(0, d, 1 - \tau(d))$ or $(e', x^*, 1 - \tau(x^*))$ for larger values of m, and conversely for smaller values.

[24] That said, Bernheim and Rangel (2009) develop a formal justification for conducting welfare analysis based on the forward-looking (or "long run") perspective that does not invoke normatively arbitrary notions of "bias;" see their Theorem 11.

Suppose we deem *all sophisticated choices* (and nothing else) welfare-relevant. Then we calculate EV_A from the perspective of a frame that yields the highest value of m, which is necessarily forward-looking. Accordingly, we use (7) and (8) with $f = -1$ (so that $b(f) = \beta$) and $\kappa(f) = 0$. Similarly, we calculate EV_B from the perspective of a frame that yields the lowest value of m, which is necessarily contemporaneous. Accordingly, we use (7) and (8) with $f = 0$ (so that $b(f) = 1$) and $\kappa(f) = 0$.

Suppose we deem *all choices* welfare-relevant. Then we calculate EV_A from the perspective of the frame that yields the highest value of m, which is plainly a forward-looking frame that leaves future consequences implicit (thereby achieving $b(f) = \beta$ and some $\kappa(f) < 1$). We calculate EV_B from the perspective of the frame that yields the lowest value of m, which is plainly a contemporaneous frame that makes future consequences explicit (thereby achieving $b(f) = 1$ and $\kappa(f) = 0$). Accordingly, the equations that identify EV_B are the same as in the previous paragraph.

Inattention. From the perspective of frame f, the equivalent variation is the value of m that satisfies

$$V(x^*, 1 + m - \tau(x^*)) = V(d, 1 - \tau(d)) - n(f)\chi(f) \tag{9}$$

if the worker chooses the default, and

$$V(x^*, 1 + m - \tau(x^*)) = V(x^*, 1 - \tau(x^*)) - \gamma - n(f)\chi(f) \tag{10}$$

if he opts out, where $n(f)$ equals 1 or -1 depending on whether f specifies $(0, x^*, 1 + m - \tau(x^*))$ or the alternative as the status quo bundle.

There is clearly a case for evaluating welfare based solely on fully attentive choices (those with $\chi(f) = 0$), on the grounds that inattentive choices involve characterization failure. However, caution is warranted for the same reasons mentioned in the context of naive time inconsistency. Absent better evidence on cognitive activity in "attentive" and "inattentive" frames, there is also an argument for remaining agnostic and respecting all choices.

Suppose we deem *fully attentive choices* (and nothing else) welfare relevant. Then we evaluate EV using (9) and (10) with $\chi(f) = 0$, which are the same formulas used to compute

EV for the basic model.

Suppose we deem *all choices* welfare-relevant. We calculate EV_A from the perspective of the frame that yields the highest value of m, which is plainly the least attentive frame in which either $(0, d, 1 - \tau(d))$ (for those who do not opt out) or $(e', x^*, 1 - \tau(x^*))$ (for those who opt out) is the status quo bundle. We calculate EV_B from the perspective of the frame that yields the lowest value of m, which is plainly the least attentive frame (i.e., one that maximizes $\chi(f)$) in which $(0, x^*, 1 + m - \tau(x^*))$ is the status quo bundle.

Anchoring. From the perspective of frame f, the equivalent variation is the value of m that satisfies

$$V(x^*(f'), 1 + m - \tau(x^*(f')), f) = V(d, 1 - \tau(d), f)$$

if the worker chooses the default, and

$$V(x^*(f'), 1 + m - \tau(x^*(f')), f) = V(x^*(d), 1 - \tau(x^*(d)), f) - \gamma$$

if he opts out, where the f' is the framing used for the baseline decision.[25]

Concerning possible restrictions on the welfare-relevant domain, the following possibility merits consideration: if one could show that workers erroneously regard defaults as useful information, then settings with explicit defaults would potentially involve characterization failure. In that case it might be appropriate to evaluate welfare from the perspective of a neutral, anchorless frame, f^N – for example, one in which an active 401(k) election is a precondition of employment.

Suppose instead we deem *all choices* welfare relevant. We calculate EV_A from the perspective of the frame $f = 0$ when $d < x^*(f')$, and from the perspective of the frame $f = \overline{x}$ when $d > x^*(f')$. We calculate EV_B from the perspective of the frame $f = \overline{x}$ when $d < x^*(f')$, and from the perspective of the frame $f = 0$ when $d > x^*(f')$. For demonstrations of these assertions, see the proof of Theorem 3.

Summary. Table 1 summarizes the frames used for measuring EV_A and EV_B for each of

[25] The choice of f' is arbitrary; we take it to be the neutral frame, f^N, defined below.

these theories. It also includes comments concerning relationships among these measures that follow from our calibration strategy, for reasons we explain below.

3 Estimation and calibration

Our main goal is to make quantitative statements concerning welfare for plausibly parametrized models. In this section, we discuss those models and describe their derivation.

3.1 Data

Following Madrian and Shea (2001), Choi et al. (2006), Beshears et al. (2008), and others, we analyze data describing distributions of 401(k) contribution rates for *recently eligible* employees at various companies before and after changes in default contribution rates. To avoid confounding factors (including any ancillary consequences of establishing automatic enrollment), we restricted attention to companies that switched between regimes with strictly positive default rates and did not change their 401(k) plans in other important ways. Three of the companies examined in the aforementioned references satisfied these criteria.[26] Those papers provide details concerning each of the three companies and their retirement plans. We summarize the salient details in Table 2.[27]

3.2 The basic model with costly opt-out

3.2.1 Identification strategy and specification

To conduct welfare analysis for the basic model, one must extract two types of information from the data: (1) the value workers derive from 401(k) contributions, and (2) the level of "as-if" opt-out costs that rationalize observed choices. Ignoring opt-out costs for the moment,

[26] Specifically, the data we use are the disaggregated distributions of contribution rates underlying Figure 3 in Beshears et al. (2008) and Figures 2B and 2C in Choi et al. (2006). The data for all three companies used in our analysis cover employees with similar tenure; they were generally eligible for several months to a little more than a year.We thank Brigitte Madrian for her generous help in providing these distributions.

[27] As shown in Table 2, Company 2 *initially* switched from a default of zero. Unexpectedly, our model performed equally well in fitting distributions for zero and strictly positive default rates. Company 3 also initially operated with a default of zero. We discarded those data because, when the company implemented automatic enrollment, it applied the policy retroactively to workers hired under the original regime.

the first type of information is embedded in the demand curve for 401(k) contributions. Because our data do not permit us to estimate that curve directly, we employ an indirect approach. Saez (2009) showed that it is possible to recover the elasticity of taxable income with respect to tax rates from the degree of bunching at kink points in a progressive income tax schedule: greater bunching implies greater responsiveness to the difference in effective prices around the kink point. An analogous kink in a worker's opportunity set appears at the contribution rate that exhausts the employer's matching contributions. Greater bunching implies that the demand for 401(k) contributions responds more elastically to the difference in the effective price of contributions around the kink point, and hence that the inframarginal benefits of those contributions are a smaller multiple of the marginal benefits.

Conditional upon knowing the value of 401(k) contributions, one can extract the second type of information (the level of as-if opt-out costs) from the degree of bunching at the default contribution rate: greater bunching implies that workers are willing to forgo greater value to avoid the costs of opt-out. Because higher opt-out costs also dampen the elasticity of demand for 401(k) contributions without altering their marginal benefits, the identification of the two types of information must be simultaneous.

Formally, our approach is to fit distributions of 401(k) contributions to a model that, while parsimonious, nevertheless allows for a broad range of possibilities concerning the key inputs for our welfare calculations – the elasticity of demand for 401(k) contributions and the distribution of as-if opt-out costs. We specify the indirect utility function as follows:

$$V(x, z, \alpha, \rho) = \rho \ln(x + \alpha) + \ln(z). \tag{11}$$

For notational brevity, we will use θ to denote the vector of preference parameters, (α, ρ). Intuitively, ρ governs the overall division of resources, while α governs the sensitivity of employee contributions to price (through the current employer matching rate). With $\alpha = 0$, V is Cobb-Douglas, expenditure shares are fixed, and employee contributions are unresponsive to a temporary price change. In contrast, a temporary reduction in price increases optimal employee contributions if $\alpha > 0$, and reduces it if $\alpha < 0$.[28]

We model the worker's budget constraint as follows: $z = 1 - \frac{(1-t)x}{1+M}$ for $x \leq x_M$ and $z = \overline{Z} - (1-t)x$ for $x \geq x_M$, where x_M is the total contribution rate at the match cap,[29] t is the marginal personal tax rate, M is the matching rate, and $\overline{Z} = 1 + (1-t)x_M \left(1 - \frac{1}{1+M}\right)$. We interpret $\overline{Z} - 1$ as the "virtual income" implicit in the kinked budget constraint. Because most workers fell into the 15% or 25% marginal tax brackets, we assume $t = 0.2$.

Intuitively, ρ is identified from the overall level of contributions, while α is identified from the degree of bunching in the distribution of contributions at the match cap. To allow for heterogeneity in tastes, we assume $\rho = \max\{\widetilde{\rho}, 0\}$, and that the CDF for the random variable $\widetilde{\rho}$, denoted F, is normal with mean μ^i, where i denotes the firm, and variance σ^2. Thus we allow average contributions to differ across firms. We treat α as common to all workers.

We also allow for heterogeneous as-if opt-out costs by assuming that γ is distributed according to a CDF, Φ, that takes the following form:

$$\Phi(\gamma) = \begin{cases} \lambda_1 + (1-\lambda_1)(1 - e^{-\lambda_2 \gamma}) \text{ for } \gamma \geq 0 \\ \\ 0 \text{ for } \gamma < 0 \end{cases}$$

Accordingly, λ_1 is the fraction of workers who act as if opt-out is costless. We take the distributions of $\widetilde{\rho}$ and λ to be independent.

Henceforth, we will use $\psi \equiv (\alpha, \sigma, \lambda_1, \lambda_2)$ to denote the values of underlying parameters that are assumed to be the same across all firms.

3.2.2 Estimation method

Workers at firm $i \in \{1, ..., I\}$ pick r, the employee contribution rate, from a discrete set $R^i \equiv \{0, 0.01, 0.02, ..., \overline{r}^i\}$, and the employer matches contributions at the rate M^i up to r_M^i. Defining $x_k^i \equiv 0.01\left[(k-1) + M^i \min\{k-1, 100r_M^i\}\right]$ and $K^i \equiv 100\overline{r}^i + 1$, the worker selects

[28] We could also allow for price responsiveness by relaxing the restriction that the elasticity of substitution between x and z is unity. However, the data are insufficiently rich to permit us to identify both the elasticity of substitution and α.

[29] For example, if the employer provides a 50% match on contributions up to 6% of income, then $x_M = 0.09$.

x from $X^i = \{x_1^i, x_2^i, ..., x_K^i\}$.[30]

For any fixed α and firm i, we partition the range of ρ into intervals, $B_1^i(\alpha) = [0, \rho_1^i(\alpha)]$, $B_2^i(\alpha) = [\rho_1^i(\alpha), \rho_2^i(\alpha)]$,..., $B_{K^i}^i(\alpha) = [\rho_{K^i-1}^i(\alpha), \infty]$, such that a worker with no opt-out costs is willing to choose $x_k^i \in X^i$ iff $\rho \in B_k^i(\alpha)$. With opt-out cost γ and default d, that worker opts out iff

$$\gamma \leq \rho \left[\ln(x_k^i + \alpha) - \ln(d + \alpha)\right] + \left[\ln(1 - \tau^i\left(x_k^i\right)) - \ln(1 - \tau^i\left(d\right))\right] \equiv \Gamma_k^i(\alpha, \rho, d)$$

A worker at firm i chooses $x_k^i \neq d$ with probability

$$\Pr{}_i(x_k^i \mid \psi, \mu^i, d) = \int_{B_k^i(\alpha)} \Phi\left(\Gamma_k^i(\alpha, \max\{0, \widetilde{\rho}\}, d)\right) dF(\widetilde{\rho}), \tag{12}$$

and chooses $x_k^i = d$ with the residual probability.

Firm i operates in S^i regimes, with default d^{is} in regime s. We observe N_k^{is}, the number of workers with $r = 0.01(k-1)$ at firm i in regime s. We have no data on workers' characteristics; the distribution of ρ subsumes such factors.[31] The total log-likelihood is:

$$\sum_{i=1}^{I} \sum_{s=1}^{S^i} \sum_{k=1}^{K^i} N_k^{is} \log\left[\Pr{}_i(x_k^i \mid \psi, \mu^i, d^{is})\right].$$

To estimate the parameters, we maximize the log-likelihood.

3.2.3 Estimates, interpretation, and fit

Table 3 contains parameter estimates, which are reasonably precise. Appendix Figure A.1 depicts the fitted and actual distributions of contribution rates under each default regime for each company. The model performs well, reproducing spikes at 0%, the default option, the maximum matchable contribution rate, and the overall cap (though predictably missing smaller spikes at 10%).

The mean utility weights mirror contributions: for companies 1, 2, and 3, respectively, the mean ideal contribution rates are 9.58%, 4.77%, and 6.51%, while the medians are 11%,

[30] So, for example, if $\bar{r}^i = 0.15$, $r_M^i = 0.06$, and $M^i = 0.5$, then a worker chooses from the set $X^i = \{0, 0.015, ..., 0.075, 0.09, 0.1, ..., 0.17, 0.18\}$.

[31] Data on worker characteristics would allow us to compute the welfare effects of defaults for separate subgroups, but it would not alter our analysis of aggregate welfare.

3%, and 5%. The standard deviation of ρ reflects considerable heterogeneity among workers. An estimated 40% of workers act as if opt-out costs are negligible.

The estimate of λ_2, the as-if opt-out cost distribution parameter, is less reasonable. Among the 60% of workers with positive opt-out costs, the mean of γ is $\frac{1}{\lambda_2} = 0.0847$, and the median is $\frac{\ln(2)}{\lambda_2} = 0.0587$. The monetary equivalent of a utility penalty γ is given by $v(\gamma)$, the solution to $V(0, 1-v(\gamma), \theta) = V(0, 1, \theta) - \gamma$. For our estimates, $v(0.0847) = 0.0812$ and $v(0.0587) = 0.0567$. If we construe the data as representing decisions taken over the first year of eligibility during which a worker earns \$40,000, the monetary equivalent of γ is more than \$3,200 at the mean and more than \$2,200 at the median. Yet it is difficult to believe that the typical employee would turn down a payment of a hundred dollars, let alone several thousand, to avoid making an active 401(k) election. Thus one can reconcile observed choices with opt-out costs of a reasonable magnitude only by introducing the types of behavioral considerations discussed above.

Why does the basic model require enormous opt-out costs to rationalize observed behavior? Intuitively, for those who would contribute even without matching provisions and tax deductibility, the EV associated with actual contributions must be very large. Extremely high opt-out costs are then required to explain why many such workers stop contributing when the default rate falls from 3% to 0%. DellaVigna (2009) reached a similar conclusion based on a back-of-the-envelope calculation concerning the value of matching contributions, which he placed at \$1,200 (for a worker earning \$40,000).

3.3 Models with frame-dependent weighting

We specify and parametrize our models of frame-dependent weighting in the same way as the basic model, except that we interpret Φ as the distribution of $D(f^*)\gamma$ rather than of γ.[32] Unfortunately, our data do not allow us to identify $D(f^*)$ and γ separately.

In light of our findings for the basic model, our main motivation for considering these alternatives is to reconcile opt-out behavior with more plausible assumptions about opt-out

[32] Recall that f^* is the naturally occurring frame.

costs. We therefore calibrate them by specifying reasonable opt-out costs, and determining the other parameters as residuals. Specifically, we assume that the mean of γ is one percent of the mean of $D(f^*)\gamma$ – in other words, the equivalent of roughly \$25 to \$30 for the typical person. We make this plausible but arguably extreme assumption in part to be conservative: by assuming that the distribution of γ is concentrated near zero, we effectively bound the range of possibilities.

For sophisticated time inconsistency, this approach implies $\beta^{-1} = D(f^*) = 100$, or equivalently $\beta = 0.01$. Typical estimates of β from the literature are much larger. It follows that sophisticated time inconsistency is likely not the main explanation for observed opt-out behavior. We comment below on the implications of using more reasonable values of β (which, like the basic model, imply implausibly large opt-out costs).

For naive time inconsistency, the same approach implies $D(f^*) = \frac{\beta^{-1}-\kappa(f^*)}{1-\kappa(f^*)} = 100$. Based on typical estimates in the literature, we set $\beta = 0.75$ and compute $\kappa(f^*)$ as a residual. The resulting value, $\kappa(f^*) = 0.997$, indicates near-perfect naivete. While we use this value, we question its plausibility, because workers would presumably learn from numerous failures to follow through on intentions over the course of a year.

For inattentiveness, the same approach implies $D(f^*) = \frac{\chi(f^*)}{\gamma} + 1 = 100$. Taken literally, this equation means we are assuming $\chi(f^*)$ is proportional to γ. While that assumption can be criticized, it is of no great consequence; any other assumption yielding a distribution of γ similarly concentrated near zero will yield comparable numerical results.

We also assume that the naturally occurring frame, f^*, is associated with greatest naivete or least attentiveness (depending on the model) potentially deemed welfare relevant, so that $\kappa(f^*) \equiv \max_f \kappa(f) \equiv \kappa_{\max}$, and $\chi(f^*) \equiv \max_f \chi(f) \equiv \chi_{\max}$. These assumptions strike us as reasonable both institutionally and in light of the enormous as-if opt-out costs implied by the estimated model.[33]

Table 1 summarizes the relationships among our measures of equivalent variation that

[33] Indeed, we have seen that our parametrized model of naive time inconsistency involves almost perfect naivete in the naturally occurring frame.

follow from this calibration strategy. Starting with sophisticated time inconsistency, EV_B^S respects choices in the naturally occurring frame, and is therefore the same as EV for the basic model. In contrast, EV_A^S discounts as-if opt-out costs by 99 percent (compare equation (8) with $f = 0$ and with $f = -1$).

Now consider naive time inconsistency. If we admit all choices, EV_A^{N1} – which we evaluate from the perspective of a maximally naive forward-looking frame – is approximately zero, irrespective of the default. In that frame, the worker is unwilling to pay much to start out with one default rather than another because he acts as if he expects to adjust his contribution rate to his ideal point at low cost with virtually no delay (see equation (7)). In contrast, EV_B^N is only slightly less than EV_A^S for sophistication. The EV_B formulas for naivete and sophistication are the same, and are identical to the EV_A^S formula, except that they inflate γ by β^{-1} (compare equations (8) with $b(f) = 1$ to (8) with $b(f) = \beta$). For the case of naivete, we assume $\beta = 0.75$, so $\beta^{-1} = 1.33$. Because we take the values of γ to be relatively small, inflating them by 33 percent has little effect on the resulting equivalent variation. In contrast, for the case of sophistication, we assume $\beta^{-1} = 100$, so the difference between EV_B^S and EV_A^S is much larger.

If instead we treat all sophisticated choices (and nothing else) as welfare-relevant, EV_A^{N2} coincides with EV_A^S. This equivalence follows from two observations: first, with $\kappa(f) = 0$, the formulas are the same, and second, because both calculations reflect forward-looking perspectives, β factors out (hence the difference in assumed values has no effect). EV_B^N is unaffected by this domain restriction. Because the $EV_A^S - EV_B$ is extremely small (see the last paragraph), so is the range of normative ambiguity for this case ($EV_A^{N2} - EV_B^N$).

For our model of inattentiveness, $\chi(f)$ appears in both (9) and (10), so it impacts *all* of the decisions that potentially define that equivalent variation, regardless of whether workers actually incur opt-out costs. Because a change in the default rate does not alter the set of workers for whom $\chi(f)$ factors into the calculation of EV, the curves relating EV to the default rate will, to a reasonable approximation, appear as parallel shifts of the EV_A^S curve (compare (9) and (10) to (7) and (8) with $\kappa(f) = 0$ and $b(f) = \beta$). The shift is upward for EV_A^I (which treats the alternative to the status quo as the baseline) and downward for EV_B^I (which treats the status quo as the baseline). Because our calibration entails large values of $\chi_{\max}$, these shifts are substantial.

3.4 The anchoring model

To introduce anchoring, we assume that any given default frame shifts the latent utility weight $\widetilde{\rho}$ toward the value that would rationalize the default as an optimal choice. Anchoring effects of this type can produce bunching at the default option. However, unlike switching costs, which tend to create a trough in the distribution of choices by sweeping out density near the default, anchoring tends to pull all choices toward the default, thereby creating a spike without a neighboring trough. It is therefore possible to identify the effects of opt-out costs and anchoring separately from the shape of the distribution of contribution rates around the default option.

Formally, for any given values of α and d, let ρ^* denote the value of ρ for which $x^*(\alpha, \rho^*) = d$.[34] With anchoring, we assume the worker acts as if his utility weight is

$$\rho = \begin{cases} \max\{0, \min\{\widetilde{\rho} + \zeta, \rho^*\}\} \text{ if } \widetilde{\rho} \leq \rho^* \\ \\ \max\{\widetilde{\rho} - \zeta, \rho^*\} \text{ if } \widetilde{\rho} \geq \rho^* \end{cases}$$

where $\zeta \geq 0$ is a constant. Thus, the anchor shifts a worker's as-if utility weight by ζ toward the value that rationalizes the default, but not beyond. The default then becomes the as-if ideal point for all individuals with $\widetilde{\rho} \in \{\rho^* - \zeta, \rho^* + \zeta\}$.

We estimate this model in the same way as the basic model, except that (12) becomes

$$\Pr{}_i(x_k^i \mid \psi, \mu_i, d) = \begin{cases} \int_{\rho_{k-1}^i(\alpha)-\zeta}^{\rho_k^i(\alpha)-\zeta} \Phi\left(\Gamma_k^i(\alpha, \max\{0, \widetilde{\rho} + \zeta\}, d)\right) dF(\widetilde{\rho}) \quad \text{if } x_k^i < d \\ \\ \int_{\rho_{k-1}^i(\alpha)+\zeta}^{\rho_k^i(\alpha)+\zeta} \Phi\left(\Gamma_k^i(\alpha, \max\{0, \widetilde{\rho} - \zeta\}, d)\right) dF(\widetilde{\rho}) \quad \text{if } x_k^i > d. \end{cases}$$

[34] In the case of $d = 0$ it is the largest such value. In the case where d coincides with the match cap, it is the nearest such value to the worker's $\widetilde{\rho}$ parameter.

The anchoring model fits the data slightly better than the basic model (see Appendix Figure A.2). As shown in Table 2, the estimates of α, μ_1, μ_2, μ_3, and σ are similar. The estimate of ζ reflects a large as-if anchoring effect that can shift the utility weight by roughly two-thirds of its standard deviation. Significantly, the estimated as-if opt-out cost distribution changes dramatically. Only 10.9% of workers act as if opt-out cost are negligible. Moreover, the estimate of λ_2 increases by almost two orders of magnitude, reducing the implied value of $v(\gamma)$ to 0.00134 at the mean, and 0.00093 at the median. For an employee earning \$40,000 per year, the implied monetary equivalent of γ is \$54 at the mean and \$37 at the median, which strikes us as reasonable. In our view, anchoring therefore emerges as the most plausible explanation for bunching at the default option.

In Section 2.5, we mentioned the possibility of evaluating welfare for the anchoring model from the perspective of a "neutral" anchorless frame. Formally, we model the neutral frame by setting $\zeta = 0$, but we acknowledge that a more complete understanding of anchoring effects would be required to justify that assumption.[35]

4 Welfare analysis

While our findings are mainly quantitative, we also derive theoretical results that provide either additional insight or some assurance of generality. Those results require some additional technical assumptions; see the Appendix for details.

4.1 Welfare analysis in models with frame-dependent weighting

Our analysis of models with frame-dependent weighting focuses on five issues: the degree of normative ambiguity; the nature of the optimal default; the size of the stakes; the desirability of using penalties to encourage active decision making; and the desirability of ensuring that

[35] This assumption is arguably justified if (a) workers act as if $\zeta = 0$ when an active 401(k) election is a precondition of employment (so that no default contribution rate is specified), (b) the presence of a default contribution rate causes the worker to ignore information he himself characterizes as pertinent (regardless of frame), and (c) no such distraction occurs under the policy regime described in (a). In that case, it is arguable that the worker correctly characterizes his alternatives only in the neutral frame.

workers make opt-out decisions within particular contexts.

1. The degree of ambiguity. A potential concern about any welfare analysis that admits multiple perspectives is that the results may be highly ambiguous, and hence of little value. In that case, to perform a discerning evaluation, one would need to adopt (and justify) some strong refinement of the welfare-relevant domain. In effect, that is the approach adopted by CCLMM, who embrace the forward-looking frame in a model with time inconsistency.

The need for a refinement may seem apparent from our parametrized model of sophisticated time inconsistency: intuitively, whether or not one heavily discounts as-if opt-out costs averaging thousands of dollars would seem highly consequential. Surprisingly, our first main finding is that the decision frame used for welfare evaluation makes only a modest difference over the pertinent range of default options. As a result, the degree of ambiguity is relatively small, and one can reach useful conclusions concerning welfare without taking a potentially controversial stand on the "correct" welfare perspective.

Figure 1 shows various versions of aggregate EV_A and EV_B, both expressed as fractions of the typical worker's income, for each of the three firms and all potential default employee contribution rates.[36] The figure assumes that workers make opt-out choices in the naturally occurring frame; we simulate and evaluate those choices using our parametrized models.

Beginning with sophisticated time inconsistency, a striking feature of the figure is that the scope of ambiguity concerning welfare, $EV_B^S - EV_A^S$, is rather small – generally less than half a percent of income (except at low default rates for company 1). Moreover, the frame used to evaluate welfare has no impact on the EV-maximizing default rate.[37] The explanation for this surprising finding is straightforward: for the range of default rates considered, the population of opt-outs is dominated by workers whose opt-out costs are relatively small or zero. Therefore, heavily discounting *incurred* as-if opt-out costs makes relatively little

[36] It is worth keeping in mind that the variables in our models, x and d, are *total* contribution rates (they include the employer match), whereas the horizontal axis in Figure 1 measures the *employee* contribution rate. The same observation applies to the figures that follow.

[37] Using a more empirically plausible value of β, 0.75 (which of course implies unreasonably large values of γ), leaves the EV_B^S unchanged, but shifts the EV_A^S curve downward, closing roughly three quarters of the gap between the two curves.

difference. With a more extreme default (e.g., 70% of earnings), matters change: because nearly all workers opt out, including those with very high opt-out costs, $EV_A^S - EV_B^S$ becomes quite large (see the Appendix). But as a practical matter such extreme default rates are not relevant, because they exceed statutory limits on contributions.

Figure 1 also presents results for naive time inconsistency. As explained in Section 3.3, EV_A^{N1}, which treats all choices as welfare-relevant, is close to zero at all default rates; EV_A^{N2}, which treats only sophisticated choices as welfare-relevant, is identical to EV_A^S; and EV_B^N, which pertains to both cases, is indistinguishable from EV_A^S given the scale of the graph. Thus, if one limits the welfare-relevant domain to sophisticated choices (both contemporaneous and forward-looking), the degree of normative ambiguity is negligible. If one treats all choices as welfare-relevant, normative ambiguity is greater, but only because the default is virtually inconsequential in the evaluation frame for EV_A^{N1}.

Next we turn to inattention. For the reasons explained in Section 3.3, EV_A^I and EV_B^I (not shown in Figure 1) are both roughly parallel to EV_A^S, with EV_A^I much higher and EV_B^I much lower, in each case by roughly 4.5 percent of income; see the Appendix. Thus, although the degree of ambiguity implied by the magnitude of $EV_A^I - EV_B^I$ is substantial, it is of little or no practical consequence if our objective is to compare one default rate to another.

2. The optimal default. In Figure 1, all measures of equivalent variation (including EV_A^{N1}, which appears flat) are maximized for a default employee contribution rate equal to the match cap (6%) at all three companies, and the same is true of EV_A^I and EV_B^I.[38] What accounts for this finding?

A possible explanation is that the optimal default is largely determined by opt-out minimization, as Thaler and Sunstein's rule of thumb assumes. Figure 1 shows that the opt-out frequency is indeed minimized at the EV-maximizing rate. Achieving a low opt-out fre-

[38] What then of the finding in Carroll et al. (2009) that an extreme default is optimal from the forward-looking perspective when sophisticated time inconsistency is sufficiently severe, as we assume it is in our parametrized model? The result still holds, but only for defaults substantially outside the range considered, where the evaluation frame matters to a much greater degree. For each company, EV_A^S reaches a global maximum for default rates above 90% (see the online appendix).

quency is clearly advantageous from a welfare perspective because it avoids the costs associated with forcing workers to make adjustments. More generally, low opt-out is achieved by setting a default that lies at a point of accumulation in the distribution of ideal worker contribution rates. Generally, those points of accumulation will include the minimum and maximum contribution rates, and any rates corresponding to convex kink points in workers' opportunity sets (such as the match cap). One would therefore expect to see a tendency for the EV-maximizing default rate to coincide with one of those values.

With small opt-out costs, the preceding conjecture holds with generality; we will prove it for sophisticated time inconsistency.[39] Formally, define $\mathcal{A} \subset [0, \overline{x}]$ to contain 0, $\overline{x}$, and all convex kink points in the workers' opportunity sets; also assume that γ and θ are distributed independently so that we can change the distribution of γ without altering that of θ. Let H^{γ} and H^{θ} be the associated CDFs, and assume that the support of H^{γ} is $[0, \overline{\gamma}]$.

Theorem 2: *Fix a frame of evaluation, $f \in \{-1, 0\}$, for the model of sophisticated time inconsistency. Consider a sequence of CDFs H_k^{γ} with $\overline{\gamma}_k \to 0$ and mean γ_k such that $\gamma_k/\overline{\gamma}_k > e^*$ for all k and some $e^* > 0$.*[40] *The EV-maximizing default rates, d_k^*, converge to a point in $\mathcal{A}$.*[41]

Theorem 2 is of interest in part because it provides a potential justification for the historically prevalent practice of setting defaults at zero. Though it identifies a connection between EV-maximization and opt-out minimization, it does not imply that the two are always the same. Indeed, the divergence between the EV-maximizing and opt-out minimizing default rates can be arbitrarily large. To understand why, note that the variation in the overall opt-out frequency over default rates is primarily driven by those with low opt-out

[39] Avoiding incurred opt-out costs becomes even more important as those costs rise, but additional considerations arise that could overturn the result in principle, even though that does not occur in our simulations.

[40] The critical property is that the right tail of the distribution of γ not be too thick, which we ensure here in a simple way by placing a lower bound on the ratio of the mean to the maximum.

[41] Because Theorem 2 provides a potential justification for setting an extreme default, it seems reminiscent of a result due to Carroll et al. (2009). However, in that paper, an extreme default is used to maximize active decision making (i.e., opt-out); here, it is optimal for precisely the opposite reason.

costs (because their opt-out decisions are more sensitive to the default rate). But those are precisely the workers for whom the default rate is least important. Accordingly, if opt-out costs are correlated with ideal points, the Thaler-Sunstein rule can be highly sub-optimal.

To explore these issues in greater depth, we examine the implications of altering the matching provisions. Throughout, we report results for EV_A^S and EV_B^S only, recognizing that results for EV_A^{N2}, EV_B^N, EV_A^I, and EV_B^I are all either identical or extremely similar to those for EV_A^S. First, we remove the matching provisions for the current period.[42] Figure 2 displays our results. The EV_A^S-maximizing default now differs considerably across companies: it is 13% for company 1, 0% for company 2, and 4% for company 3 (mirroring the median ideal contribution rates reported in Section 3.1). The EV_B^S-maximizing default rates are similar: 12% for company 1, 1% for company 2, and 5% for company 3. With one exception, these rates do not coincide with the minimum or maximum contribution rates.[43] Nor do they generally coincide with the opt-out-minimizing defaults, which are 15% for company 1 and 0% for companies 2 and 3. The gap between EV-maximizing and opt-out minimizing defaults is particularly large for company 3.

Second, we simulate choices and evaluate welfare effects with match caps other than 6%. Focusing again on the models of sophisticated time-inconsistency, Figure 3(a) plots the EV_A^S-maximizing default employee contribution rate as a function of the match cap for the three companies, while Figure 3(b) plots the EV_B^S-maximizing defaults. In each case, the EV-maximizing default coincides with the match cap for intermediate values, but they differ for low and high match caps, in some cases dramatically (and in those cases the EV-maximizing default rate does not typically equal either the minimum or maximum contribution rate).

[42] To simulate choices, we simply change the worker's current opportunity constraint to $z = 1 - tx$. Our reduced-form approach does not allow us to simulate the effects of changes in future matching provisions on V. We note that our analysis may overstate the responsiveness of contributions to the current match rate. By assuming V is differentiable, we attribute all of the bunching at x_M to the kink in the current period's budget constraint. Part of that bunching may be due to a kink in V, because (a) the current choice is somewhat persistent, and (b) future matching creates a kink in the future opportunity set at x_M. If, however, the costs of switching arise from a new employee's lack of familiarity with his employer's benefits procedures, they may decline rapidly with tenure, in which case any induced kink in V would be minor.

[43] In contrast, for EV_A^{N1}, which appears flat in the figure, the optimal default rate is the maximum rate for company 1, and zero for companies 2 and 3.

Thus, setting the default rate equal to the match cap is desirable in some instances, but not in others. Note also the similarity between Figures 4(a) and 4(b): the frame of evaluation does not have much bearing on the welfare-optimal policy.

3. The stakes. To assess the economic importance of defaults, we compare the aggregate EV for three alternatives: zero, the maximum rate, and the EV-maximizing rate. We focus on EV_A^S and EV_B^S because all but one of the other EV measures are either identical or extremely similar to EV_A^S. For the remaining measure, EV_A^{N1}, the stakes are plainly trivial.

With matching provisions in place, setting the default optimally rather than at zero raises EV_A^S by 0.99%, 0.32%, and 0.54% of earnings for companies 1, 2, and 3, respectively; for EV_B^S, the changes are 2.29%, 0.61%, and 0.98%. Similarly, setting the default optimally rather than at the maximum rate raises EV_A^S by 0.42%, 0.67%, 0.58% of earnings for companies 1,2, and 3, respectively; for EV_B^S, the changes are 0.57%, 1.17%, and 0.93%. Using a back-of-the-envelope calculation, we place the net value of the opportunity to make 401(k) contributions at roughly 7%, 2%, and 2.75% of earnings for companies 1, 2, and 3, respectively.[44] Thus, the use of a suboptimal default rate can dissipate a substantial fraction – typically 15 to 30 percent – of the potential economic benefits created by 401(k) plans.

The welfare costs of using a suboptimal default are considerably smaller without a match, because there is no"free money" at stake. Setting the default optimally rather than at zero raises EV_A^S by 0.49%, 0%, and 0.01% of earnings for companies 1, 2, and 3, respectively; for EV_B^S, the changes are 1.00%, 0.01%, and 0.17%. Similarly, setting the default optimally rather than at the maximum rate raises EV_A^S by 0.02%, 0.58%, and 0.29% of earnings for companies 1, 2, and 3, respectively; for EV_B^S, the changes are 0.08%, 1.02%, and 0.61%. Some of these percentages are small simply because the optimal default is close to either zero or the maximum contribution rate. Setting the worst default – zero for company 1, and the maximum rate for companies 2 and 3 – continues to dissipate a substantial fraction of the value derived from 401(k) participation.

[44] These ballpark figures represent total employer contributions plus somewhere between 20% and 40% of employee contributions, a rough estimate of the tax advantages.

Although opt-out minimization is typically sub-optimal for these companies in the absence of matching provisions, the welfare costs of following the Thaler-Sunstein rule of thumb are fairly small: for companies 1, 2, and 3, respectively, the losses are 0.02%, 0%, and of 0.01% of earnings based on EV_A^S, and 0.08%, 0.01%, and 0.17% based on EV_B^S.

4. Penalties for passive choice. CCLMM raise the possibility that, with a sufficient degree of sophisticated time inconsistency, and evaluating welfare from the perspective of the forward-looking decision frame, it may be optimal to compel active decision making by setting an extreme default or a large penalty for passive choice. However, they do not ask whether a firm could beneficially employ penalties and defaults *in combination*, for example, by setting a moderate penalty along with an attractive default.

We address this issue by simulating decisions made in the naturally occurring frame when sophisticated time-inconsistent workers face both defaults and penalties for passive choices. Then we optimize simultaneously over both instruments, evaluating welfare according to EV_A^S, as in CCLMM.[45] We find that the optimal penalty is enormous (roughly 55% of earnings) and the default is essentially inconsequential. We can overturn the latter result by assuming that some small fraction of the population, η, never makes an active decision. As we increase η from zero, the optimum changes sharply from a policy with an extreme penalty and a largely inconsequential default to one with no penalty and attractive default. Figure 4 illustrates why this result holds. Fixing a default of 6%, it graphs average EV_A^S for company 1 against the size of the penalty (measured as a fraction of earnings) with η ranging from 0.25% to 1.25%.[46] Each curve has two local maxima, one at zero and one at a massive penalty. Varying η simply determines which is the global optimum. Thus, the availability of a penalty either does not change the optimal default problem, or renders it essentially irrelevant.

5. Contexts for opt-out decisions. The potential benefit of precommitment opportunities

[45] Recall, that maximizing EV_A^S is equivalent, or nearly so, to maximizing several welfare measures for other models.

[46] For these calculations, we assume that the matching provision at employer 1 is in effect.

is an important theme in the literature on time inconsistency.[47] Trivially, from the forward-looking perspective, sophisticated forward-looking choices are necessarily ideal. However, as we show next, such opportunities can have a previously unrecognized down-side: to the extent one wishes to remain agnostic concerning the proper welfare standard, they can introduce *substantial normative ambiguity*, increasing EV_A but *reducing* EV_B.

Figure 5 and Figure 6 are identical to Figures 1 and 2, except that we assume workers make decisions in a forward-looking frame.[48] Focusing initially on sophisticated time inconsistency, we see gaps between EV_A^S and EV_B^S ranging to 4% of earnings. Moreover, from the perspective of the forward-looking frame, the choice of the default is of little consequence, but from the perspective of the contemporaneous frame, the opt-out minimizing default rate is strongly preferred. Thus, reducing welfare ambiguity emerges as a potential reason to prefer policies that force workers to make opt-out decisions in the contemporaneous frame.

The explanation for this increase in welfare ambiguity is straightforward: shifting decisions from the contemporaneous frame to the forward-looking frame increases welfare from the perspective of the forward-looking frame (EV_A^S) by bringing the decision frame and the evaluation frame into alignment, but decreases welfare according to the contemporaneous frame (EV_B^S) by creating misalignment. The ambiguity is large because the vast majority of workers – even those who have extremely high as-if opt-out costs in the naturally occurring frame – opt out when making decisions in the forward-looking frame (see Figures 5 and 6) . Whether one steeply discounts those costs is therefore enormously consequential.

In sharp contrast, for naive time inconsistency, Figures 5 and 6 also show that offering precommitment opportunities has the *opposite* effect on welfare ambiguity; indeed, it results in trivial welfare losses for *all* evaluation frames.[49] Intuitively, with precommitments and

[47] For an exception, see Bernheim, Ray, and Yeltekin (2013), who demonstrate how external commitment devices can undermine the effectiveness of internal self-control mechanisms.

[48] For our model of naive time inconsistency, the worker makes identical decisions in all forward-looking frames, regardless of the degree of naivete, $\kappa(f)$.

[49] Compared with Figures 1 and 2, EV_B^N increases sharply, and is approximately zero for all default rates. EV_A^{N1} also increases because the change aligns the decision frame with the frame of evaluation, but this difference is not noticeable because the values are already close to zero in Figures 1 and 2. We do not mean to suggest, however, that shifting the opt-out choice from the naturally occurring frame to a forward-

low values of γ, virtually all workers end up with their ideal points. Moreover, they incur little cost in the process, even from the perspective of the contemporaneous frame, which inflates γ only by the factor $\beta^{-1} = 1.33$. Thus, creating opportunities for precommitments emerges as the best policy even if the correct frame of evaluation is unclear. When such opportunities are available, the choice of the default is largely inconsequential.

Analogous issues arise with inattentiveness. From the perspective of a fully attentive frame, the best policy is plainly one that makes workers fully attentive. Surprisingly, the same conclusion follows even if one instead remains agnostic about the underlying cognitive processes, and hence about the correct frame for evaluation. While there is substantial ambiguity with respect to the measurement of equivalent variation when all choices are treated as welfare-relevant, switching decisions from an inattentive frame to a fully attentive one shifts both the EV_A^I curve and the EV_B^I curve upward.[50] The reason is simple: the EV_A^I and EV_B^I curves are a roughly constant vertical distance from the EV_A^S curve, which we have seen shifts upward when moving from Figures 1 and 2 to Figures 5 and 6. Thus, presenting workers with opt-out choices in a frame that induces "as-if" fully attentive behavior emerges as the best policy even if the correct frame of evaluation is unclear. With that policy, the choice of the default is largely inconsequential.[51]

The following theorem underscores the generality of the preceding results.

Theorem 3: *Assuming all choices are deemed welfare-relevant:*

(i) *For the model of sophisticated time inconsistency, shifting the opt-out decision from the naturally occurring frame to the forward-looking frame weakly increases* EV_A^S *and*

looking one unambiguously increases welfare. Plainly, it reduces welfare evaluated from the perspective of the naturally occurring frame, and therefore cannot create a generalized Pareto improvement if choices in that frame are deemed welfare-relevant (given that this model satisfies the multi-self conditions). However, the reduction in welfare according to that perspective is tiny.

[50] Compare Appendix Figures A.5 and A.6 on the one hand, with Figures A.3 and A.4 on the other. This does not mean that a shift from the naturally occurring frame to a fully attentive one produces a generalized Pareto improvement. If all choices are deemed welfare-relevant, then a change in the decision frame cannot produce an unambiguous increase in welfare according to P^*.

[51] Because the EV_A^I and EV_B^I curves are roughly parallel to the EV_A^S curve, and because the latter is close to a flat line in Figures 5 and 6, so are the former.

weakly reduces EV_B^S (in each case strictly if $\Delta(d) \in (\gamma, \beta^{-1}\gamma)$).

(ii) *For the model of naive time inconsistency, shifting the opt-out decision from the naturally occurring frame to a forward-looking one weakly increases EV_A^{N1} (strictly if $\Delta(d) \in (\gamma, \beta^{-1}\gamma)$); it also weakly increases EV_B^N (strictly if $\Delta(d) \in \left(\beta^{-1}\gamma, \frac{\beta^{-1}-\kappa(f^*)}{1-\kappa(f^*)}\gamma\right)$) for all workers except those with $\Delta(d) \in (\gamma, \beta^{-1}\gamma)$ and possibly $\Delta(d) = \gamma$.*

(iii) *For the model of inattention, shifting the opt-out decision from the naturally occurring frame to a fully attentive one weakly increases both EV_A^I and EV_B^I (in each case strictly if $\Delta(d) \in (\gamma, \gamma + \chi(f^*))$)*

Parts (i) and (iii) establish the generality of our conclusions for sophisticated time inconsistency and inattention, respectively. For naive time inconsistency, part (ii) implies that the upward shift in the EV_A^{N1} curve is completely general, and that the EV_B^N curve shifts upward whenever $\kappa(f^*)$ is close to unity and the distribution of $\beta^{-1}\gamma$ is concentrated near zero, so that the number of workers with $\Delta(d) \in \left(\beta^{-1}\gamma, \frac{\beta^{-1}-\kappa(f^*)}{1-\kappa(f^*)}\gamma\right)$ is large relative to the number with $\Delta(d) \in [\gamma, \beta^{-1}\gamma)$.

4.2 Welfare analysis with anchoring

Because our parametrized model of anchoring involves low opt-out costs, one can gain insight into the results reported below by studying the special case where those costs are zero:

Theorem 4: *Assume $\gamma = 0$. EV evaluated from the perspective of the frame f is maximized at $d = f$, non-decreasing for $d < f$, and non-increasing for $d > f$. Assuming all choices are deemed welfare-relevant: (i) every default rate is a weak generalized Pareto optimum; (ii) if x_0 is the baseline default contribution rate, then EV_A^A is non-increasing in d provided $x^*(d) < x_0$ and non-decreasing in d provided $x^*(d) > x_0$, while EV_B^A is non-decreasing in d provided $x^*(d) < x_0$ and non-increasing in d provided $x^*(d) > x_0$.*

It follows immediately that, if all choices are deemed welfare-relevant, one cannot say that any default is unambiguously better than any other. Also notice the sharp conflict between potential welfare perspectives: for any given worker, EV_A and EV_B move in opposite directions as the default changes. Indeed, part (ii) implies that the graph of EV_A^A and EV_B^A for any single worker resembles a horizontal hourglass.[52]

Figure 7 (which assumes actual matching provisions) and Figure 8 (which assumes no match) graph aggregate EV_A^A and EV_B^A for the three companies as functions of the default. Because the curves are fairly flat, it may be tempting to infer that the choice of a default is, at most, only modestly consequential. However, that inference is incorrect. As indicated in Table 1, EV_A^A is evaluated in the frame $f = 0$ for some workers, and in the frame $f = \overline{x}$ for others. An increase in the default shifts workers from the second group to the first. Thus, when we compare EV_A^A for different defaults, the composition of evaluation frames differs. Similar statements hold for EV_B^A. From any single fixed perspective, the default rate is in fact highly consequential, and welfare implications differ dramatically across perspectives. To illustrate, each panel of Figures 7 and 8 also shows aggregate EV from the perspective of frames $f = 0$ and $f = \overline{x}$. As Theorem 3 suggests, the first of these decreases monotonically while the second increases monotonically. Consider company 1 in Figure 8. Moving from $d = 0$ to $d = 0.15$, EV_A^A and EV_B^A, neither of which is evaluated from the perspective of a fixed frame, change by -0.07% and $+0.62\%$ of income, respectively – a relatively modest conflict. In contrast, EV falls by 2.35% from the perspective of $f = 0$, and increases it by 2.89% from the perspective of $f = \overline{x}$. Thus, the range of ambiguity concerning the welfare effects of this change exceeds five percent of income.[53] Similar statements hold for the other panels in Figures 7 and 8.[54]

[52] This statement assumes there is some intermediate d for which $x^*(d) = x_0$; otherwise, both curves are monotonic.

[53] If we calculated equivalent variations using an extreme baseline default rate (either 0 or $\overline{x}$) rather than $x^*(f^N)$, EV_A^A and EV_B^A would each be evaluated from the perspective of a single frame, and consequently would be as steeply sloped as the curves for EV evaluated from the perspective of $f = 0$ and $f = \overline{x}$ shown in the figures. That is why the flatness of the EV_A^A and EV_B^A curves in the figures is potentially misleading.

[54] As noted in the previous section, a policy maker who wishes to "play it safe" might consider setting the default to maximize the lowest value of equivalent variation across all evaluation frames. Here that strategy will yield a different answer depending on the values chosen for workers' baseline contribution rates. Because that choice is fundamentally arbitrary, the "play it safe" strategy is misguided.

Accordingly, one cannot make precise welfare statements in this setting without restricing the set of choices deemed welfare-relevant. One possibility is to admit choices only if they are made in the neutral frame, f^N, which eliminates the influence of anchors. The resulting measure of equivalent variation, EV_N^A, is also shown in Figures 7 and 8. Strikingly, all the EV_N^A curves are rather flat. With a match (Figure 7), EV_N^A varies between -0.52% and -0.91% for company 1, between -0.70% and -0.87% for company 2, and between -0.70% and -0.92% for company 3. While it is maximized at a default rate equal to the match cap at all three companies, the welfare loss from setting a default of zero is only 0.39% of earnings for company 1, 0.15% for company 2, and 0.22% for company 3. Results without a match (Figure 8) are generally similar.[55]

So far, we have limited our discussion to employee welfare. Because higher defaults increase contributions, they obviously impose costs on employers (through matching) and the government (through taxes). We are unable to measure the present value those effects. However, if worker welfare is largely unaffected by the default contribution rate (as is the case for EV_N^N in Figures 7 and 8), then plainly $d = 0$ emerges as the social optimum.

4.3 An observation concerning the Pareto criterion

We have treated the task of selecting a default as a matter of *de novo* policy design. From the perspective of an employer with existing employees, it is actually a matter of policy *reform.* As Feldstein (1976) noted, the problem of reform differs from that of *de novo* design in that it involves a starting point. When creating a 401(k) plan, an employer may wish to ensure that no worker is made worse off (the "Pareto improvement criterion"). In this section, we show that a plan meets this criterion if and only if $d = 0$.[56] This observation provides another potential justification for setting the defaults to zero.

It is trivial to verify the preceding claim for the standard model with opt-out costs.[57]

[55] The optimal default rates differ, however: EV_N^A is maximized at the contribution limit for company 1, at 0 for company 2, and at 10% for company 3.

[56] While the Pareto *improvement* criterion is discerning in this context, the simple Pareto criterion is not. With sufficient heterogeneity across workers, all defaults are Pareto efficient.

[57] With $d = 0$, no worker can be worse off because each has the option not to contribute without incurring

However, additional considerations arise with frame-dependent weighting or anchoring, and the principle is not completely general. Still, under some additional technical assumptions (see the Appendix), the Pareto improvement criterion implies that a 401(k) plan must have a default of zero and, with frame-dependent weighting, that the frame in which workers make the opt-out decision (f_D) must belong to the set of welfare-relevant frames that are *least* conducive to contributing, in the sense that $D(f_D) = D_M$, where D_M is the maximum value of $D(f)$ within the welfare-relevant domain.[58]

Theorem 5: *Regardless of whether the welfare-relevant domain is unrestricted or restricted to any subset of frames, offering a 401(k) plan in the current period creates a weak generalized Pareto improvement over not offering a plan in the current period if and only if* $d = 0$ *and, for the cases of as-if time inconsistency and inattentiveness,* $D(f_D) \geq D_M$*. Furthermore, for those same cases, setting* f_D *such that* $D(f_D) = D_M$ *creates a weak generalized Pareto improvement over setting* f_D *such that* $D(f_D) > D_M$ *(in each case along with* $d = 0$*).*

5 Concluding remarks

We have investigated the welfare effects of 401(k) plan defaults under various assumptions about the sources of default effects, using the welfare framework proposed by Bernheim and Rangel (2009). Our main results are summarized in the introduction. Naturally, the paper leaves many important questions unanswered. More empirical research is required to distinguish between the choice patterns associated with the various theories of default effects, and to justify potential restrictions on the welfare-relevant domains. Other explanations for

opt-out costs; however, with $d > 0$, any worker who ideally prefers to contribute zero is necessarily worse off, either because he contributes d or because he contributes zero and incurs the opt-out cost. Note, however, that as a general matter, if $x^*(\theta)$ has full support on $[0, \overline{x}]$ (which we assume) every feasible d is Pareto optimal.

[58] For our models of time inconsistency and anchoring, we assume that the welfare-relevant domain contains either all of the choices for a given frame or none of them. For our model of inattentiveness, we assume that, if the welfare-relevant domain contains a choice problem in which the status quo receives an as-if utility bonus of χ, then for all X and $x \in X$, it contains a choice problem wherein X is the opportunity set and x, the status quo, receives the same as-if utility bonus.

default effects may also merit exploration. For example, we interpret opt-out costs in our models as pertaining to implementing decisions, rather than to reaching decisions. Costly decision making is notoriously difficult to model, as one is quickly drawn into an infinite regress: determining whether a problem is worth solving requires the individual to solve a more difficult problem which in turn may or may not be worth solving, and so forth. We leave such matters to future studies.

References

[1] Ariely, Dan, George Loewenstein, and Drazen Prelec (2003), "Coherent Arbitrariness: Stable Demand Curves without Stable Preferences," *Quarterly Journal of Economics* 118(1), 73-105.

[2] Bernheim, B. Douglas (2009), "Behavioral Welfare Economics," *Journal of the European Economic Association* 7(2-3), 267–319.

[3] Bernheim, B. Douglas (2014). *Simple Solutions for Complex Problems in Behavioral Economics.* Clarendon Lectures. Oxford University Press, forthcoming.

[4] Bernheim, B. Douglas, and Antonio Rangel (2007), "Toward Choice-Theoretic Foundations for Behavioral Welfare Economics," *American Economic Review Papers and Proceedings* 97(2), 464-470.

[5] Bernheim, B. Douglas, and Antonio Rangel (2008), "Choice-Theoretic Foundations for Behavioral Welfare Economics," In Andrew Caplin and Andrew Schotter (eds.), *The Methodologies of Modern Economics*, Oxford University Press.

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

[7] Bernheim, B. Douglas, Debraj Ray, and Sevin Yeltekin (2013), "Poverty and Self-Control," NBER Working Paper No. 18742.

[8] Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian (2008), "The Importance of Default Options for Retirement Savings Outcomes: Evidence from the United States," in Stephen J. Kay and Tapen Sinha, eds., *Lessons from Pension Reform in the Americas*, Oxford: Oxford University Press, 59-87.

[9] Bronchetti, Erin Todd, Thomas S. Dee, David B. Huffman, and Ellen Magenheim, "When a Nudge Isn't Enough: Defaults and Saving Among Low-Income Tax Filers," NBER Working Paper No. 16887, March 2011.

[10] 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), pp. 1639-74.

[11] Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick (2002). "Defined Contributions Pensions: Plan Rules, Participant Decisions, and the Path of Least Resistance," in James Poterba, ed., *Tax Policy and the Economy*, Cambridge, MIT Press, pp. 67-113.

[12] Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick (2003), "Passive Decisions and Potent Defaults," NBER Working Paper 9917.

[13] Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick (2004), "For Better or for Worse: Default Effects and 401(k) Savings Behavior," in David A. Wise, ed., *Perspectives on the Economics of Aging*, Chicago: University of Chicago Press, 81-121.

[14] Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick (2006), "Saving for Retirement on the Path of Least Resistance," *Behavioral Public Finance: Toward a New Agenda*, Russell Sage, Ed McCaffrey and Joel Slemrod, eds., 304-351.

[15] DellaVigna, Stefano (2009), "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature* 47(2), 315-372.

[16] Della Vigna, Stefano, and Ulrike Malmendier (2004), "Contract Design and Self-Control: Theory and Evidence," *Quarterly Journal of Economics* 119, 353-402.

[17] Feldstein, Martin (1976), "On the Theory of Tax Reform," *Journal of Public Economics* 6, 77-104.

[18] Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman (2010), "Getting to the Top of Mind: How Reminders Increase Saving," NBER Working Paper No. 16205.

[19] Laibson, David (1997), "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics* 112(2), 443-478.

[20] 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.

[21] O'Donahue, Ted, and Matthew Rabin (1999), "Doing it Now or Later," *American Economic Review* 89(1), 103-24.

[22] O'Donahue, Ted, and Matthew Rabin (2001), "Choice and Procrastination," *Quarterly Journal of Economics* 116, 121-160.

[23] Saez, Emmanuel (2009), "Do Taxpayers Bunch at Kink Points," *AEJ: Economic Policy* 2(3), 180-212.

[24] Thaler, Richard, and Cass R. Sunstein (2003), "Libertarian Paternalism," *American Economic Review Papers and Proceedings* 93(2), 175-179.

Table 1: Details concerning welfare analysis for various models

Model	Welfare-relevant domain	EV_A	EV_B
Sophisticated time inconsistency	All choices	Notation: EV_A^S Frame: forward-looking	Notation: EV_B^S Frame: contemporaneous Same as *EV* for basic model
Naive time inconsistency (1)	All choices	Notation: EV_A^{N1} Frame: forward-looking, implicit Approximately zero	Notation: EV_B^N Frame: contemporaneous, explicit Slightly less than EV_A^S
Naive time inconsistency (2)	Choices with future consequences explicit	Notation: EV_A^{N2} Frame: forward-looking, explicit Same as EV_A^S	Notation: EV_B^N Frame: contemporaneous, explicit Same as naive time inconsistency (1)
Inattentiveness	All choices	Notation: EV_A^I Frame: inattentive, status quo≠baseline Same "parallel" upward shift of EV_A^S	Notation: EV_B^I Frame: inattentive, status quo =baseline Large "parallel" downward shift of EV_A^I
Anchoring	All choices	Notation: EV_A^A Frame: either $f = 0$ or $f = \bar{x}$	Notation: EV_B^A Frame: either $f = 0$ or $f = \bar{x}$

Table 2: Description of the companies

Parameter	Company 1	Company 2	Company 3
Default regimes	3%, 6%	0%, 3%, 6%	3%, 4%
Matching rate	1	0.5	0.5
Maximum matchable contribution	0.06	0.06	0.06
Contribution limit	0.15	0.15	Up to 25%
Dates Observed	2002 - 2003	1997 - 2001	1998 - 2002
Industry	Chemicals	Insurance	Food

Source: Beshears et al. (2008) for Company 1, and Choi et al. (2006) for Companies 2 and 3".

Table 3: Parameter Estimates

Parameter	Description of Parameter	Basic Model	Model with Anchoring
α	Retirement saving shift parameter	0.1340	0.1027
		(0.0023)	(0.0680)
μ_1	Mean utility weight, company 1	0.2150	0.2155
		(0.0079)	(0.0263)
μ_2	Mean utility weight, company 2	0.1313	0.1260
		(0.0016)	(0.0419)
μ_3	Mean utility weight, company 3	0.1570	0.1487
		(0.0023)	(0.0214)
σ	Standard deviation of utility weight	0.0910	0.1222
		(0.0005)	(0.0369)
λ_1	Fraction of employees with zero opt-out costs	0.4011	0.1094
		(0.0021)	(0.0422)
λ_2	Opt-out cost distribution parameter	11.8100	747.2000
		(0.1600)	(199.4000)
ζ	Anchoring parameter		0.0785
			(0.0209)
Log Likelihood		-2.8250E+05	-2.8050E+05

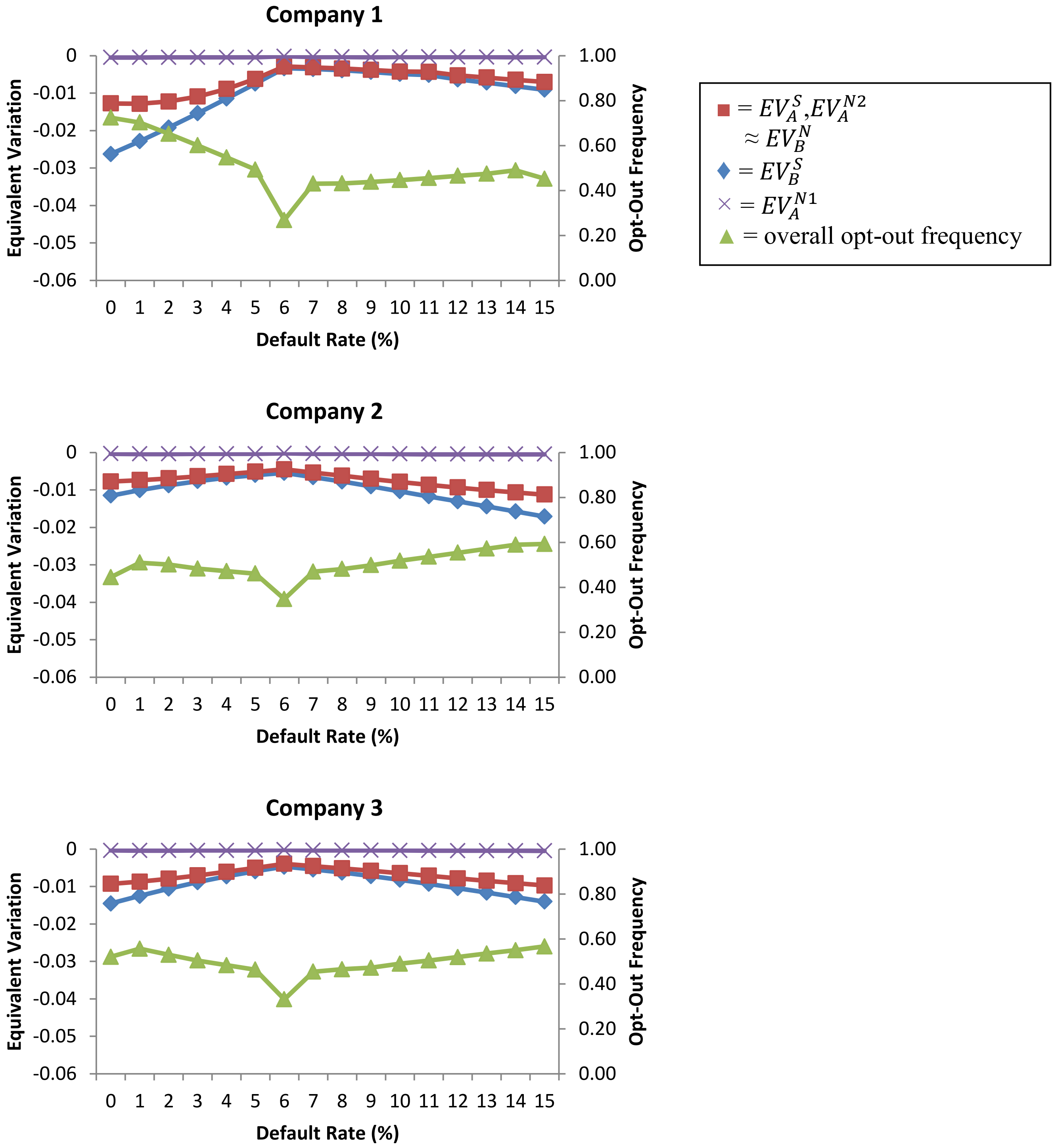


Figure 1: Average equivalent variations and opt-out frequencies, with decisions made in the naturally occurring frame, and with an employer match. We plot EV_A and EV_B for sophisticated time inconsistency, and naïve time inconsistency treating (1) all choices as welfare-relevant, and (2) only choices with explicit future consequences as welfare-relevant.

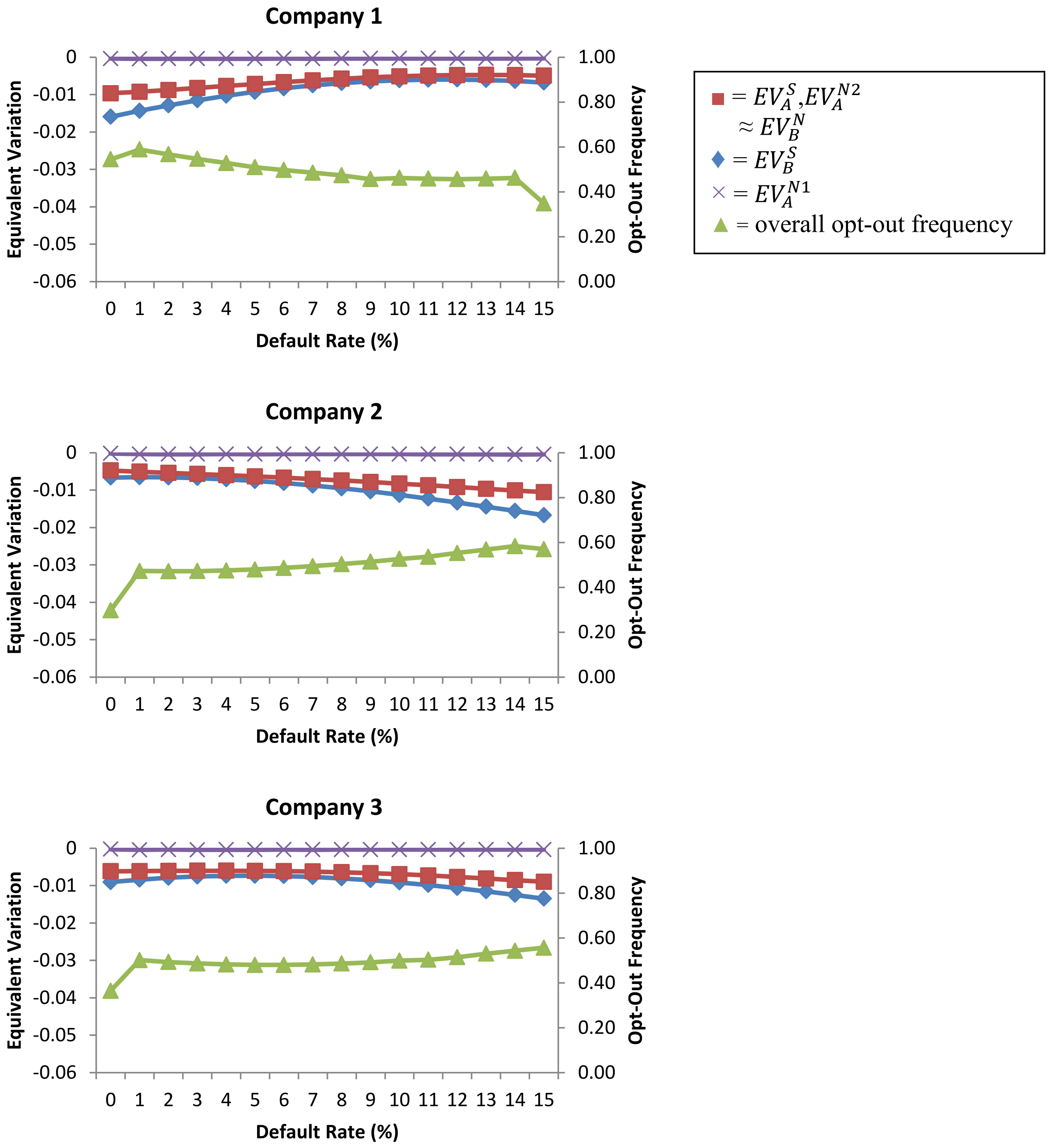


Figure 2: Average equivalent variations and opt-out frequencies, with decisions made in the naturally occurring frame, without an employer match. We plot EV_A and EV_B for sophisticated time inconsistency, and naïve time inconsistency treating (1) all choices as welfare-relevant, and (2) only choices with explicit future consequences as welfare-relevant.

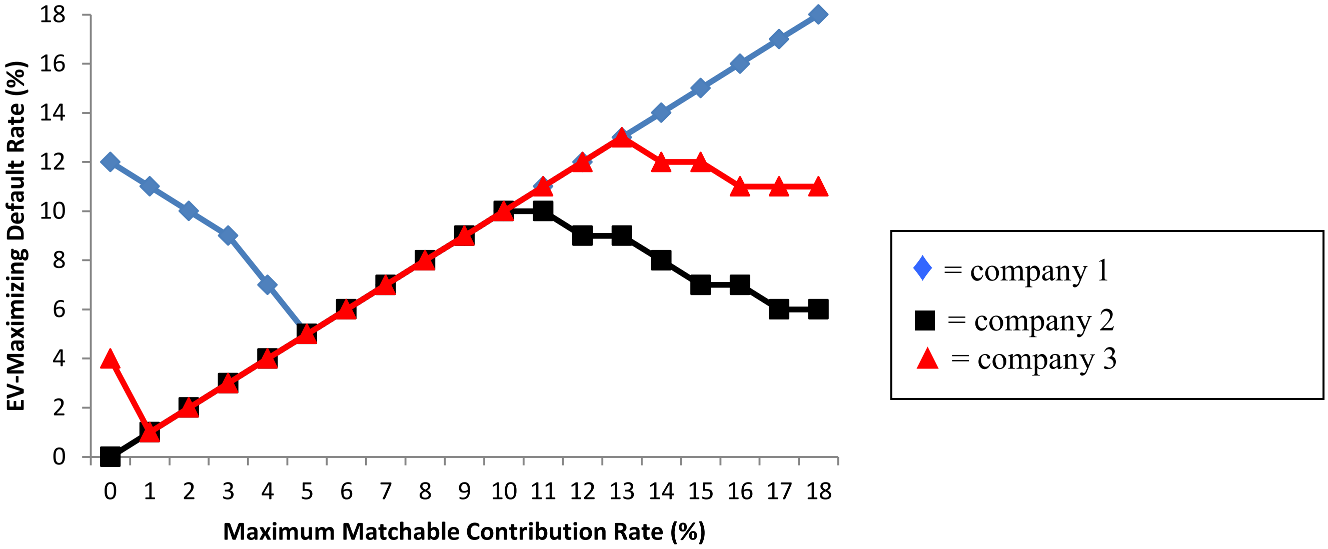


Figure 3(a): Average EV_A^S-maximizing default rate versus maximum matchable employee contribution rate, with decisions made in the naturally occurring frame, and with an employer match.

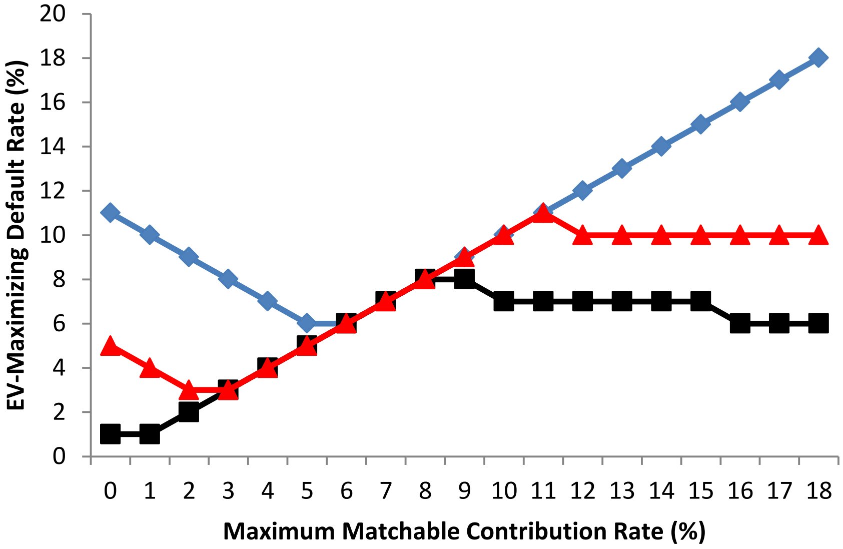


Figure 3(b): Average EV_B^S-maximizing default rate versus maximum matchable employee contribution rate, with decisions made in the naturally occurring frame, and with an employer match.

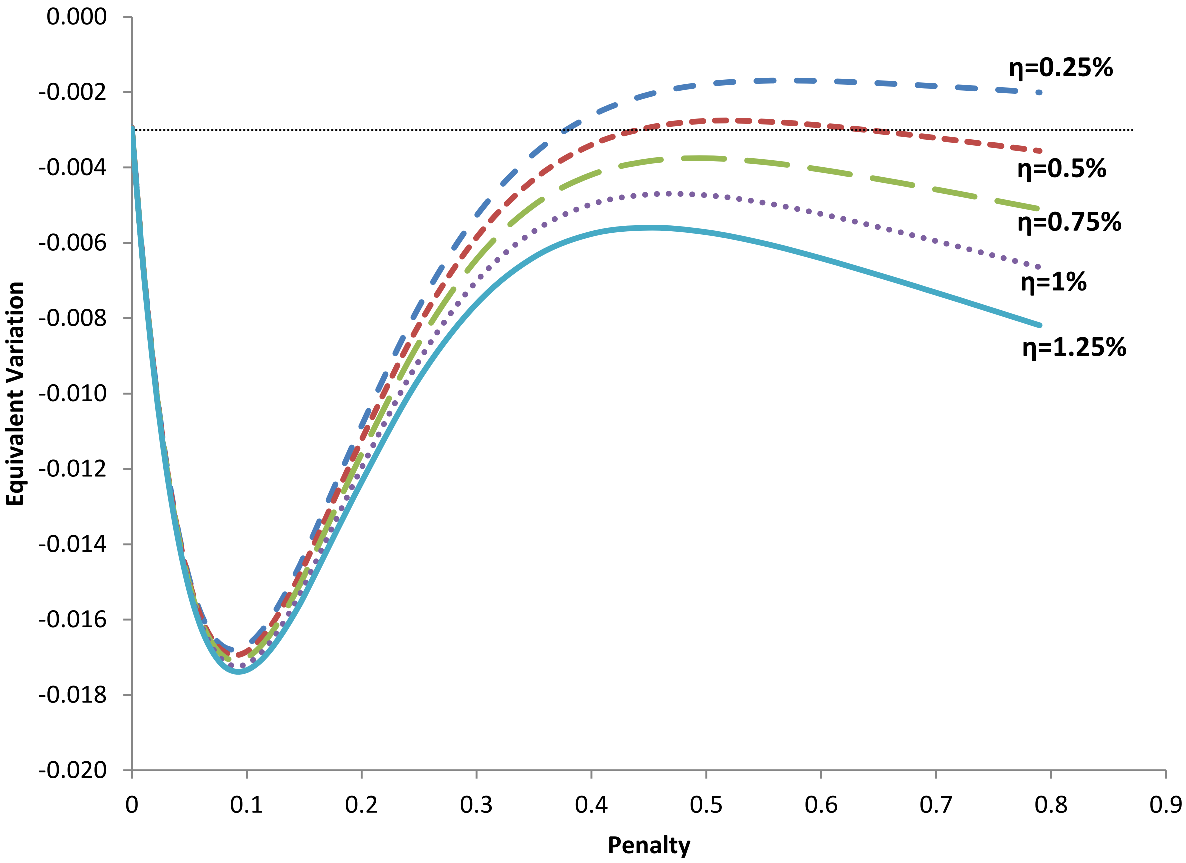


Figure 4: Average EV_A^Sas a function of the penalty for inactive choice, with the default rate fixed at 6 percent. Based on company 1, with decisions made in the naturally occurring frame, and with an employer match. Each line corresponds to a different value of η, the fraction of the population that never makes an active decision.

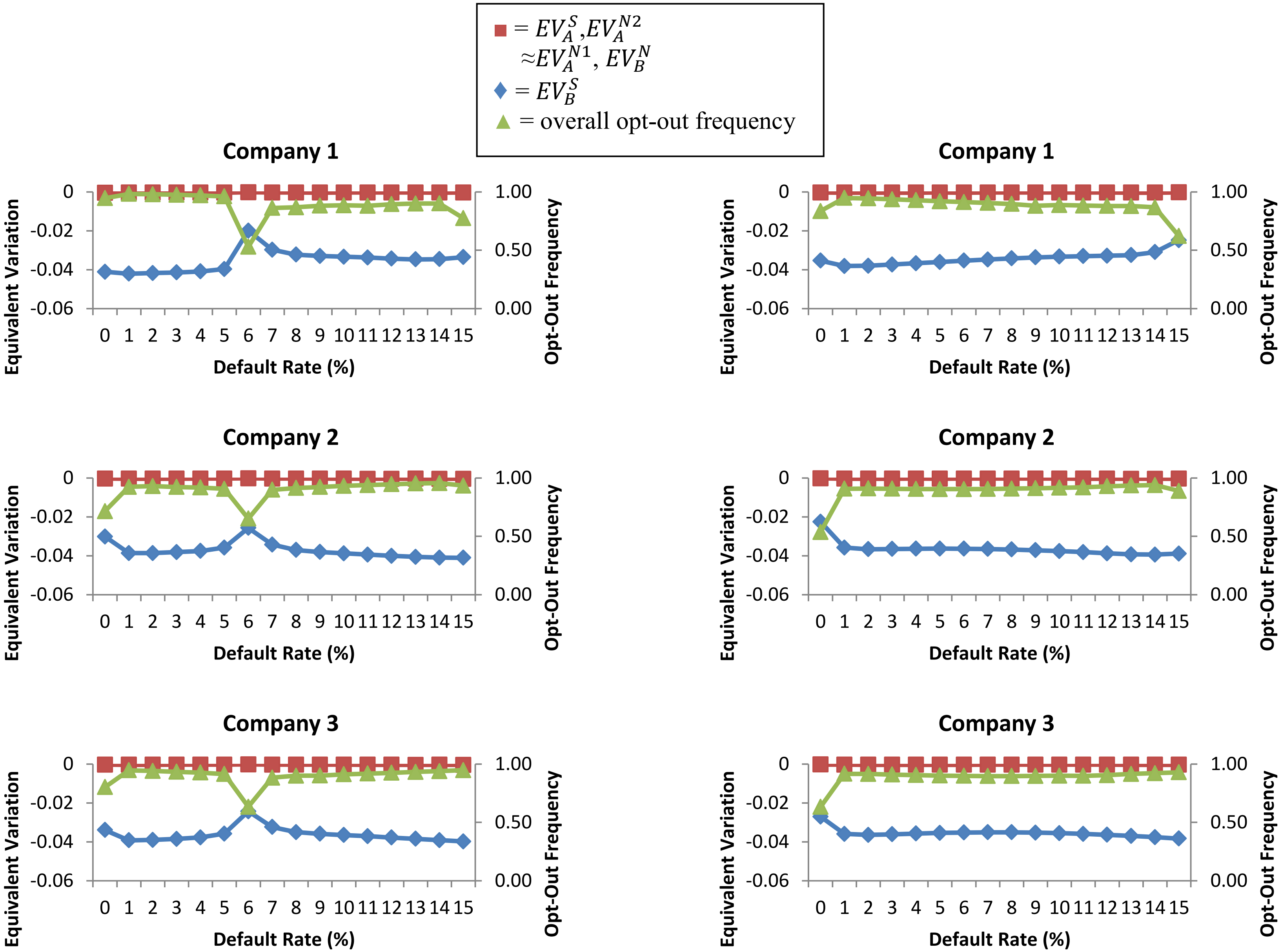
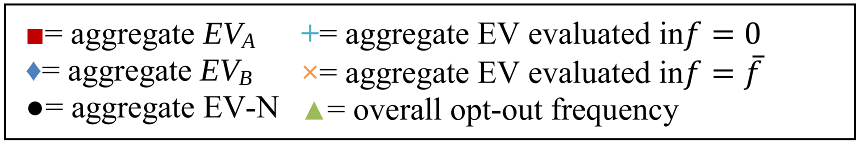


Figure 5: Average equivalent variations and opt-out frequencies, with decisions made in the alternative frame, and with an employer match. We plot $EV_A$and EV_B for models of sophisticated and naïve time inconsistency.

Figure 6: Average equivalent variations and opt-out frequencies, with decisions made in the alternative frame without an employer match. We plot $EV_A$and EV_B for models of sophisticated and naïve time inconsistency.


■= aggregate EV_A
◆= aggregate EV_B
●= aggregate EV-N
+= aggregate EV evaluated in $f = 0$
×= aggregate EV evaluated in $f = \bar{f}$
▲= overall opt-out frequency

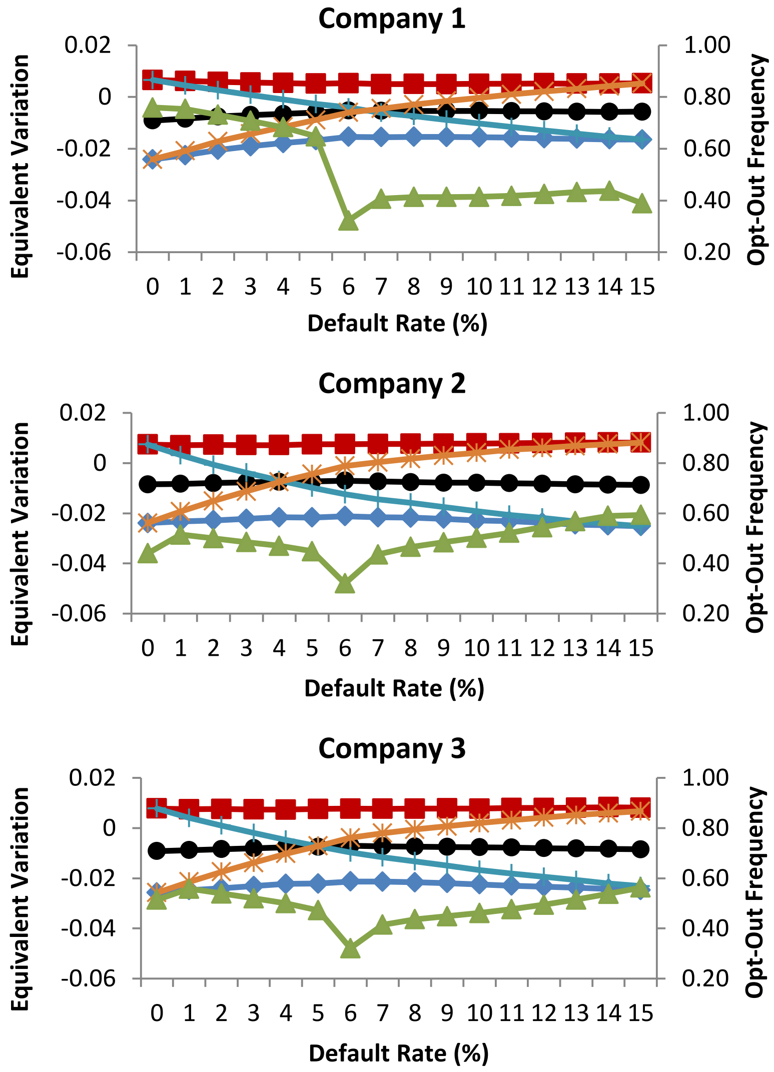


Figure 7: Average equivalent variation and opt-out frequency, with anchoring, and with an employer match. We separately evaluate EV for each employee in the most favorable, least favorable, lowest, highest, and neutral frames.

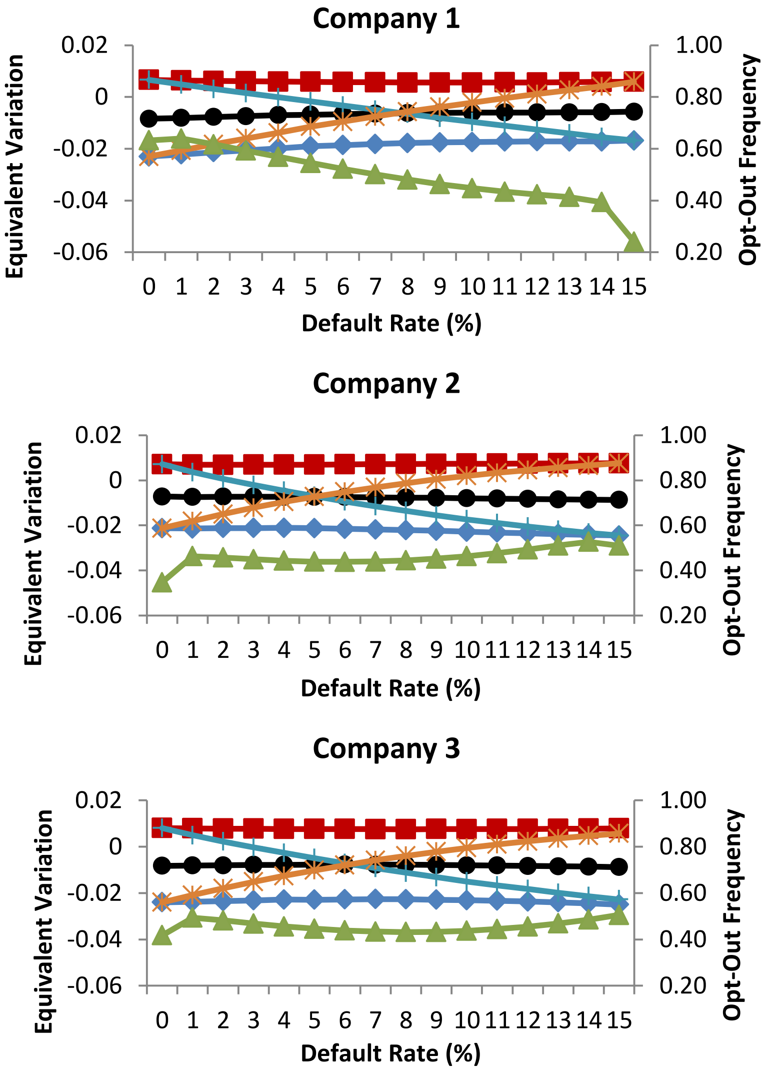


Figure 8: Average equivalent variation and opt-out frequency, with anchoring without an employer match. We separately evaluate EV for each employee in the most favorable, least favorable, lowest, highest, and neutral frames.