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, 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 July 2013
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 theory using the framework for behavioral welfare economics developed by Bernheim and Rangel (2009). 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. In some settings, non-participation emerges as the optimal default, contrary to common wisdom.

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

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

Andrey Fradkin
Department of Economics
Stanford University
579 Serra Mall
Stanford, CA 94305-6072
afrad@stanford.edu

An online 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 the distribution of contributions among relatively new employees.[1] The magnitude of that effect dwarfs those of more conventional policy instruments such as capital income taxes. Yet in comparison, default effects have received far less attention. Moreover, with the exceptions discussed below, the critical task of evaluating the *welfare effects* of default options has been almost entirely ignored. That task poses two separate types of conceptual challenges. First, the cognitive mechanisms behind default effects are poorly understood, and there are a number of competing explanations (as detailed below). Second, several of those explanations involve violations of standard choice axioms that render traditional normative tools inapplicable. Thus, an alternative welfare framework is required.

In this paper, we analyze the welfare effects of 401(k) default contribution options quantitatively, using reasonably parameterized models that have been calibrated to data reflecting responses on the critical behavioral margins. Thus we are in a position to provide some practical guidance concerning both the normative importance of default options and the nature of welfare-optimal policies. As far as we know, this is the first paper to provide such an analysis. Instead of focusing on a single theory of default effects, we consider multiple alternatives involving opt-out costs, frame-specific weighting (a term we use for a class of choice processes that encompasses sophisticated time inconsistency, naive time inconsistency, and inattentiveness as special cases), and psychological anchoring. To address the conceptual problems arising from the non-standard elements of these theories, we employ the framework for behavioral welfare analysis developed by Bernheim and Rangel (2009). In that framework, inconsistencies in choice (e.g., across decision frames) translate into a quantifiable degree of normative ambiguity, which one can either accept or reduce/resolve

[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.

by refining the set of choices deemed welfare relevant (e.g., by adopting the perspective of a single frame).

While welfare is our main focus, our calibration results are of some independent interest. In a conventional model, unrealistically large opt-out costs (averaging thousands of dollars for the typical worker) are required to rationalize observed default effects. Thus we provide an empirical justification for examining behavioral theories. Moreover, the data appear to favor an explanation involving anchoring effects: once the model is extended to allow for those effects, the calibrated opt-out cost distribution becomes reasonable.

With respect to models with frame-dependent weighting (i.e., the two flavors of time inconsistency plus inattentiveness), our main findings are as follows. First, the welfare implications of varying the default rate over the pertinent range are only modestly sensitive to the decision frame used for evaluation. Accordingly, even if one treats all decision frames as welfare-relevant, the degree of normative ambiguity is small. This finding is surprising because, as noted above, as-if opt-out costs average thousands of dollars in the "naturally occurring" decision frame; thus, the degree to which those costs are discounted in alternative frames would seem highly consequential. The explanation is that, while as-if opt-out costs are large on average for the entire population, they are small on average among workers who actually incur those costs by opting out; hence, evaluating welfare from the perspective of choice frames that discount those costs to varying degrees makes relatively little difference.

Second, there is a strong tendency for the welfare-optimal default rate to coincide with the cap on employer matching contributions, even when that cap differs significantly from the desired contribution rate of the typical worker. The explanation is that 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. Because that effect is quantitatively large, the magnitude of incurred opt-out costs dominates other considerations; as a result, the default rate that minimizes the opt-out frequency (a rule of thumb advocated by Thaler and Sunstein, 2003) is welfare-optimal in these cases. In contrast, when matching

provisions are absent, optimal default rates track the center of distribution of worker preferences reasonably closely. That result holds even though there are theoretical reasons to anticipate that the optimum would lie either at the lowest or highest allowable contribution rate (again because those are points of accumulation in the distribution of ideal contribution rates, and hence setting the default rate equal to one of those values would minimize incurred opt-out costs).

Third, when a 401(k) plan includes a reasonably 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 several percent of earnings, representing nearly one-third of the potential surplus generated by the 401(k) plan. However, without matching provisions, the stakes are considerably smaller. Notably, the welfare losses that result from following the Thaler-Sunstein opt-out-minimization criterion are quite small even when that criterion prescribes a suboptimal default rate; hence it emerges as a reasonable rule of thumb in general.

Fourth,we investigate the use of penalties for passive choice. As discussed below, previous theoretical research has shown that it is sometimes better to compel active decision making (either through a large penalty or by setting an extreme default) than to encourage passive choice by setting an attractive default. We examine whether a firm could beneficially employ defaults and penalties for passive decision making *in combination*, for example, by setting a moderate penalty along with an attractive default. We find that the optimum *either* involves an extreme penalty (so that the default is virtually irrelevant), *or* an attractive default with no penalty (so that active decisions are not compelled). This finding reflects the fact that welfare is a double-peaked function of the size of the penalty, with the one peak at zero and the other at a level high enough to compel active choice.

Fifth, we investigate the value of precommitment opportunities – i.e., a policy that allows each worker to decide, in advance, whether he (or she) will be subject to a large penalty for passive choice, rather than one that imposes such penalties indiscriminately. While the

potential benefit of precommitment opportunities is an important theme in the literature on time inconsistency, we show that they can have a previously unrecognized down-side, in that they create *substantial ambiguity* concerning the welfare effects of alternative default policies. The explanation is related to the first result mentioned above: with precommitments, the set of workers who opt out will include those who act as if opt-out costs are enormous when making the same decisions without precommitment.

With respect to models with anchoring, our main results are as follows. When all welfare perspective are admitted, the degree of normative ambiguity is extremely high, both with respect to the identity of the optimal default, and with respect to the magnitudes of welfare gains and/or losses. Thus, in contrast to our models with frame-dependent weighting, welfare analysis is largely uninformative absent a restriction on the welfare-relevant choice domain. One possible candidate is to perform evaluations in a "neutral" frame; i.e., one in which choices are free from the effects of anchors. Significantly, we find that aggregate worker welfare in the neutral frame is almost entirely independent of 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.

As mentioned above, the previous literature on the normative implications of default options is extremely limited. To our knowledge, Thaler and Sunstein (2003) were the first to comment on this topic, though not in the context of a formal model. Instead, 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 previous study has attacked these issues in the context of a formal model: Carroll, Choi, Laibson, Madrian, and Metrick (2009), henceforth CCLMM. The analysis in that paper assumes that default effects arise from procrastination by sophisticated time-inconsistent decision makers.[2] The authors also adopt a particular perspective on welfare – that "true well-being" is governed by "long-run"

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

preferences. They show that, with a high degree of time inconsistency, the optimal policy is to force active decisions, e.g., by setting an extreme default contribution rate. In contrast, 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 (assuming population heterogeneity with respect to desired saving is low), or to skew it toward either end of that distribution (if population heterogeneity is high).

While CCLMM's analysis represents an important first step toward understanding the normative implications of default options, it is limited in a number of respects. First, it is not quantitative. While it enumerates several possibilities for optimal default policies, it provides no guidance as to which applies in practice; nor does it gauge the magnitude of the welfare costs associated with setting suboptimal default options. Second, CCLMM consider only a single behavioral theory of default effects: i.e., that workers have time-inconsistent preferences. The need for a non-standard theory is not self-evident, inasmuch as opt-out costs are by themselves sufficient to generate default effects. Furthermore, to the extent non-standard behavioral tendencies contribute to the magnitude of those effects, considerations other than time inconistency may be paramount. Third, as noted above, CCLMM adopt a single welfare perspective, involving what they and others have termed the "long-run" criterion. That choice is controversial (see Bernheim, 2009). Those who favor the long-run criterion as a welfare standard for quasihyperbolic discounting often argue that it reflects the decision maker's true preference purged of "present bias." Yet it is equally possible that people overintellectualize their choices when making decisions at arms length, and that they properly appreciate experiences only "in the moment." Fourth, CCLMM's stylized theoretical model omits factors that potentially play important roles in determining optimal defaults, such as caps on employer matching contributions and bounds on employee contributions (which create points of accumulation in the distribution of ideal contribution rates).[3] Fifth, CCLMM do not examine some interesting policy alternatives, such as the

[3] CCLMM explicitly assume that the distribution of ideal contribution rates is atomless.

combined use of defaults and penalties for passive choice, or the introduction of opportunities to precommit to active choice through self-imposed penalties. The current paper usefully supplements CCLMM's analysis by addressing each of these limitations.

The remainder of the paper is organized as follows. Section 2 sets forth the models of default effects, and Section 3 details their calibration. Section 4 briefly summarizes the Bernheim-Rangel framework and discusses its application to the problem at hand. Section 5 uses the calibrated models to investigate welfare, and provides some theoretical results that illuminate and extend our numerical findings. Section 6 provides some concluding remarks. Proofs of theoretical results appear in an online appendix.

2 Models of default effects

2.1 The basic model with costly opt-out

Consider an individual who has recently become eligible to participate in his employer's 401(k) plan. His *total contribution rate* $x \in [0, \overline{x}] \equiv X$ is defined as the sum of employer and employee contributions, divided by earnings (exclusive of the employer contribution). The plan has a default employee contribution rate that implies a total contribution rate of $d \in X$. We focus on the initial choice (in "period 0") between (a) accepting the default and (b) expending costly effort to opt out by selecting $x \in X \backslash d$.

As of period 0, the worker cares only about his opt-out effort level e and his (possibly state-contingent) future consumption trajectory, c, which encompasses not only goods but also effort subsequently expended to change contribution rates.[4] Period 0 preferences correspond to a utility function $u(e, \omega) + U(c, \theta)$, where ω and θ are (potentially overlapping) parameter vectors and $u(0, \omega) = 0$. Let $u(e', \omega) \equiv -\gamma \leq 0$, where e' is the fixed effort level required under prevailing rules to opt out of the default. The period 0 choice of x matters because it determines the worker's current 401(k) saving, his default for the next period,[5]

[4] The elements of c are potentially indexed by both time and states of nature. All consumption other than e takes place after period 0.

[5] In principle, these first two effects are separable (e.g., upon electing a contribution rate of 3%, the

and cash available for near-term consumption and non-401(k) saving (z), all of which impact his subsequent opportunity set for c.

Choosing c to maximize U subject to future opportunity constraints (parameterized by a vector π) for fixed x and z yields an optimal continuation consumption correspondence $C(x, z, \theta, \pi)$. We assume that, given x, π does not depend on d.[6] Defining the indirect utility function $V(x, z, \theta, \pi) = U(c, \theta)$ for $c \in C(x, z, \theta, \pi)$, we can we treat the worker's short-term problem as one of maximizing

$$W(e, x, z, \omega, \theta, \pi) = u(e, \omega) + V(x, z, \theta, \pi) \tag{1}$$

over e, x, and z, subject to two constraints. The first constraint is that $x \neq d$ requires $e \geq e'$. The second constraint pertains to the worker's budget. Specifically, a choice $x \in [0, \overline{x}]$ yields

$$z = 1 - \tau(x), \tag{2}$$

where τ reflects deductibility of contributions as well as employer matching provisions. Generally, τ is an increasing function with $\tau(0) = 0$ and $\tau(\overline{x}) < 1$. With a non-increasing matching rate and a match cap, τ will be a piecewise-linear function with a finite number of convex kink-points.

Fixing $e = 0$ and maximizing (1) subject to (2), ignoring the costs of opting out, yields an "ideal point" $x^*(\theta)$. The worker opts out of d to $x^*(\theta)$ iff

$$\Delta(\theta, d, \pi) \equiv V(x^*(\theta), 1 - \tau\left(x^*(\theta)\right), \theta, \pi) - V(d, 1 - \tau(d), \theta, \pi) \geq \gamma. \tag{3}$$

Notice that d enters only through the period 0 opportunity constraint for (e, x, z) bundles; any choice of x renders the initial d subsequently irrelevant.[7] That observation allows us to

default for the next period could change to 4%), but in practice they always go hand-in-hand (in the previous example, the new default would be 3%).

[6] Future default rates depend on the initial default rate only indirectly through the initial contribution rate (which in practice establishes a new default).

[7] This property hinges on the assumed absence of any relation between the future opportunity set (parameterized by π) and the initial default rate d, given the initial contribution rate x.

implement our framework empirically by estimating the reduced-form valuation function V rather than the primitive utility function U, which we can accomplish with more limited data, and to simplify the analysis of optimal defaults by working with reduced-form preferences over (e, x, z) bundles rather than primitive preferences over (e, c) bundles.[8] In taking this approach, it is possible that we will either (a) impose structure on V that is inconsistent with the underlying optimization problem, or (b) fail to impose structure implied by that problem. With respect to (a), our assumptions concerning V are modest and largely innocuous.[9] With respect to (b), we are skeptical of the prospects for deriving helpful properties of sufficient generality; in any event, empirical analysis adds appropriate structure by fitting V to data, and our theoretical analysis yields useful insights without additional structure.

2.2 Models with frame-dependent weighting

A number of psychological mechanisms can amplify the default effects resulting from opt-out costs. In this section, we will discuss a class of mechanisms characterized by *frame-dependent weighting*. The defining characteristic of these mechanisms is that the worker acts as if he places more weight on the effort costs of opt-out (relative to future consequences) in some psychological decision frames than in others. Specifically, he acts as if he maximizes

$$W(e, x, z, \omega, \theta, \pi, f) = \delta(f)u(e, \omega) + V(x, z, \theta, \pi),$$

where f is the decision frame and $\delta(f)$ is the weighting function.[10] Rather than (3), the opt-out criterion then becomes

$$\Delta(\theta, d, \pi) \geq \delta(f)\gamma. \tag{4}$$

[8] Without knowing anything about the correspondence C, we can conclude that the bundle (e, c) for $c \in C(x, z, \theta)$ is chosen over (and hence revealed preferred to) (e', c') for $c' \in C(x', z', \theta)$ from the observation that (e, x, z) is chosen over (e', x', z').

[9] We explicitly acknowledge a potential exception in Section 5.1.

[10] Implicit in this formulation of frame-dependent weighting is the assumption that the initial period 0 frame, f, does not affect the continuation consumption trajectory, $C^*(x, z, \theta, \pi)$ (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).

Notice that, for any *fixed* decision frame, a model with frame-specific weighting is observationally equivalent to our simple model of costly opt-out.

With existing institutional arrangements, there is a frame f^* under which workers normally make the opt-out decision. We will call f^* the *naturally occurring frame*, and normalize the function u so that $\delta(f^*) = 1$.

Below we show that psychological mechanisms involving sophisticated time inconsistency, naive time inconsistency, and attentiveness all involve frame-dependent weighting.

Sophisticated time inconsistency. To introduce sophisticated time inconsistency, we assume that a worker can choose to opt out either in a *contemporaneous frame* $f = 0$ wherein he makes that choice "in the moment," or a *forward-looking frame* $f = -1$ wherein he decides whether to opt out (and commits to that choice) one period in advance. With *present bias*, we have $\delta(-1) < \delta(0) = 1$. For existing institutions, the contemporaneous frame is naturally occurring ($f^* = 0$).

If default effects are attributable to sophisticated time inconsistency, then the frequency with which workers opt out should differ between the two frames. We know of no direct evidence on that point.

Naive time inconsistency. Introducing naive time inconsistency through this framework requires a bit more structure. As an approximation, suppose that the utility received from electing (x, z) accrues at a *constant rate*, given by the function $V(x, z, \theta, \pi)$.[11] Then the net utility cost of defaulting to d rather than electing $x^*(\theta)$ accumulates at the rate $\Delta(\theta, d, \pi)$ (defined above). In any given decision frame f, the worker will hold (probabilistic) beliefs about the point in time at which he will switch from d to $x^*(\theta)$ assuming he does not do so immediately.[12] We can write the expected costs of not switching immediately according to those beliefs as $D(f)\Delta(\theta, d, \pi)$, where $D(f)$ reflects the amount of time the worker expects to pass before he switches. As with sophisticated time inconsistency, the decision frame also

[11] One does not need to assume that the utility accrual rate is constant forever. It need only be (roughly) constant through the most distant date at which the individual thinks he is likely to switch from d to $x^*(\theta)$ if he does not do so immediately.

[12] If, for example, the value of γ is determined randomly each period, the date of switching will be uncertain.

affects the weight placed on present versus future consequences, which we will write here as $\widehat{\delta}(f)$. Accordingly, in frame f, the worker will opt out immediately if and only if

$$D(f)\Delta(\theta, d, \pi) \geq \widehat{\delta}(f)\gamma. \tag{5}$$

Notice that we can rewrite condition (5) in the form of condition (4) simply by taking $\delta(f) = \frac{\widehat{\delta}(f)}{D(f)}$.

Beliefs are never actually observed directly; they are inferred from choices.[13] Thus, when we say that a decision maker is naively time-inconistent, we are asserting that his choices manifest a second form of frame-dependence (in addition to the form discussed in the context of sophisticated time-inconsistency). Specifically, for a naive individual, choices will differ according to whether future consequences are *explicit* or *inferred*. When future consequences are explicit, the worker is told that each given choice in period 0 will be followed by a specific consumption trajectory; when they are inferred, he is left to choose for himself, and must anticipate his own future actions. A naive decision maker manifests frame dependence because he makes a different choice in period 0 depending on whether the *same* future consequences are explicit or inferred. Here that frame dependence naturally takes a particular form: if the date of switching from d to $x^*(\theta)$ is explicit in frame f_1 and inferred in frame f_2, then $D(f_1) > D(f_2)$ (in other words, a naive worker will not anticipate future procrastination resulting from present-bias).

Thus, for a naive time-inconsistent decision maker, there are four relevant decision frames, corresponding to whether the opt-out choice is contemporaneous or forward-looking, and whether future consequences are explicit or inferred. For existing institutions, opt-out choices are made in the contemporanous frame with inferred future consequences. Notice that $\delta(f) = \frac{\widehat{\delta}(f)}{D(f)}$ is larger for that frame than for any of the other three frames.

If default effects are attributable to naive time inconsistency, then the frequency with which workers opt out should differ among the four frames mentioned above. We know of

[13] Obviously, one can ask someone to report their beliefs. If the reports are not incentivized, it is not clear what they represent. If they are incentivized, then we interpret consequential choices as implying beliefs, but we still do not observe the beliefs.

no direct evidence on that point.

Inattentiveness. Whether an employee attends to the task of selecting a 401(k) contribution rate may depend on the psychological decision frame, f. We assume the worker behaves as if he attends if and only if the choice is sufficiently consequential, in the sense that the stakes exceed some threshold, $\chi(f)$. Thus, the worker attends and opts out if and only if

$$\Delta(\theta, d, \pi) \geq \chi(f) + \gamma, \tag{6}$$

Notice that we can rewrite condition (6) in the form of condition (4) simply by taking $\delta(f) = \frac{\chi(f)}{\gamma} + 1$. Here, the naturally occuring frame is defined by the institutional environment (including benefits communications programs) maintained by the worker's employer. Our utility normalization, $\delta(f^*) = 1$, is equivalent to $\chi(f^*) = 0$.

If default effects are attributable to inattentiveness, opt-out frequencies should be sensitive to interventions that manipulate attention. The evidence on that point is both limited and mixed.[14]

2.3 A model with anchoring

Defaults may also influence decisions through the power of suggestion; i.e., they themselves may frame the way workers think about their choices. For example, a default may provide a salient starting point (or "anchor") for a worker's thinking,[15] or workers may see it as a "stamp of approval" by an authority on retirement planning. We can model these effects as forms of frame dependence, but natural formulations do not involve frame-dependent weighting. Here, the default rate d not only impacts the worker's opportunity set by changing the effort required to achieve any contribution rate, but also establishes a psychological frame,

[14] 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.

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

$f = d$, that inclines workers toward choosing $x = f$. The opportunity-set effect and the framing effect are empirically distinguishable given appropriate data. For example, one could separate them through choice experiments in which the default rate and the effort schedule are varied independently, possibly by adding red tape to make some alternatives (including the default) more or less time consuming than others, so as to reveal the choices workers would make with a default frame f when the effort schedule favors some other $d \neq f$. Because such variation does not exist in practice, we must separate the framing and opportunity-set effects empirically through additional identifying assumptions (see Section 3).

To incorporate anchoring, we assume the worker acts as if the reduced-form indirect utility function V depends on the default frame $f \in [0, \overline{x}]$.[16] Accordingly, he maximizes

$$W(e, x, z, \omega, \theta, \pi, f) = u(e, \omega) + V(x, z, \theta, \pi, f). \tag{7}$$

We use $x^*(\theta, f)$ to denote the worker's as-if ideal point, which now depends on the frame. The opt-out decision is still governed by (3), except that f appears as an additional argument of V (and hence Δ).

Our formulation is not meant to suggest that the default directly affects "true well-being;" indeed, comparisons of $V(x, z, \theta, \pi, f)$ and $V(x', z', \theta, \pi, f')$ are meaningful only if $f = f'$.[17] On the contrary, we intend (7) merely as an analytic device for recapitulating the dependence of a choice mapping on a decision frame f.

2.4 Additional technical assumptions

While the focus of this paper is mainly quantitative, we also supplement our calculations with some theoretical results that provide either additional insight or some assurance of generality. Those results require some additional technical assumptions, which we collect

[16] 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.

[17] Like θ, f parameterizes ordinal preferences over (e, x, z) bundles.

in this section. Readers uninterested in technical theoretical details can safely skip to the next section.

We assume that V is strictly quasiconcave in (x, z), strictly increasing in both x and z, with $\lim_{z\to 0} V(x, z, \theta, \pi) = -\infty$ and $\lim_{z\to\infty} V(x, z, \theta, \pi) = +\infty$, and continuously differentiable (except at $z = 0$).[18] We allow the preference parameters $\xi \equiv (\gamma, \theta) \in [0, \overline{\gamma}] \times \Theta \equiv \Omega$ to differ across workers and use H to denote their CDF.[19] Except where stated otherwise, we assume H has full support on Ω and $\overline{\gamma}$ is very large, so that the fraction of individuals opting out of any default lies strictly between 0 and unity. We take Θ (and hence Ω) to be compact. We assume τ is strictly increasing, piecewise linear, continuous, and convex. Under our assumptions, the ideal point $x^*(\theta)$ is unique and varies continuously with θ. We assume that the (induced) distribution of $x^*(\theta)$ has full support on $[0, \overline{x}]$, with atoms at 0, $\overline{x}$, and the kink points of τ (if any), but nowhere else,[20] and that the density is bounded at all other points.

For all models with frame-dependent weighting, we assume that the ranking of frames by δ is the same for all workers, and assign labels to frames so that δ is strictly increasing in f.[21] For the model of attention, we posit the existence of some frame $\overline{f}$ least conducive to attention, and assume that, for any default d, the set of workers opting out has positive measure even with $\overline{f}$.

For the model with anchoring, we assume for some purposes that an increase in f weakly shifts the individual's choices toward higher x (*monotonicity*).[22]

[18] When extending the model to anchoring, we make the same assumptions conditional on each frame f.

[19] Notice that we treat γ rather than ω as the preference parameter governing opt-out costs; this is valid as long as we take the opt-out technology as fixed.

[20] This reasonable property can be derived from more primitive assumptions about the distribution of θ and the properties of V, but the associated technical issues do not illuminate the problem of interest. For the anchoring model, we make the same assumption about $x^*(\theta, f)$ for each f.

[21] Implicitly, we treat any set of frames yielding the same value of δ as a single frame.

[22] Formaly, if $W(e, x, z, \omega, \theta, f) \geq W(e', x', z', \omega, \theta, f)$, where $x > x'$ and $z < z'$, then $W(e, x, z, \omega, \theta, f') > W(e', x', z', \omega, \theta, f')$ for $f' > f$.

3 Calibration

As explained in the introduction, our main goal is to make quantitative statements concerning the welfare effects of 401(k) default options for plausibly calibrated behavioral models. In this section, we discuss the calibrated models and describe their derivation.

3.1 Data

Any calibration strategy is of course constrained by the nature of the available data. In this case, those data describe distributions of 401(k) contribution rates for workers at various companies before and after changes in default contribution rates. Such data have been studied previously by Madrian and Shea (2001), Choi et al. (2006), Beshears et al. (2008), and several other papers by combinations of those authors.

To avoid the possibility of confounding the effects of a change in the default rate with some ancillary consequence of establishing automatic enrollment and/or adopting other policy changes, we restricted attention to companies that (i) switched between regimes with strictly positive default rates, and (ii) did not change their 401(k) plans in other important ways. Three of the companies examined in the aforementioned references satisfied these criteria.[23]

For each company and default regime, the data indicate the fraction of *recently eligible* employees who elected each allowable contribution rate.[24] The various papers cited above provide details concerning each of the three firms and their retirement plans. To conserve space, we summarize the salient details in Table 1.[25]

[23] 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). We thank Brigitte Madrian for her generous help in providing these distributions.

[24] According to the previously cited papers, the data for all three companies cover employees with similar tenure. It appears that included employees were generally eligible for several months to a little more than a year.

[25] As shown in Table 1, Company 2 *initially* switched from a default of zero to a positive rate, and then *subsequently* switched to a different positive rate. Unexpectedly, our model performed equally well in fitting distributions for zero and strictly positive default rates. Although not indicated in the table, Company 3 also initially operated under a regime with a 0% default rate. We discarded those data because, when the company implemented automatic enrollment, it applied the policy retroactively to workers hired under the 0% default regime.

3.2 Identification strategy and specification

3.2.1 The basic model with costly opt-out

To conduct welfare analysis for the basic model with costly opt-out, one must extract two types of information from the data: (1) the value workers derive from 401(k) contributions as a function of the amount contributed, and (2) the level of "as-if" opt-out costs (i.e., the level of costs that rationalizes observed choices).

Ignoring for the moment the presence of opt-out costs, the first type of information would be recoverable in the usual way from estimates of the demand curve for 401(k) contributions. With exogenous variation in matching rates, one could in principle estimate the demand elasticity directly. However, because we lack such variation in our data, another identification strategy is required. Our strategy is adopted from Saez (2009), who showed that it is possible to recover the elasicity 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, and hence a higher elasticity. In the current setting, an analogous kink in the worker's opportunity set appears at the contribution rate that exhausts the employer's matching contributions. A higher degree of bunching at that kink point implies that the demand for 401(k) contributions responds more elastically to the variation 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 as a function of the amount purchased, one can extract the second type of information (the level of as-if opt-out costs) directly from the degree of bunching at the default contribution rate: greater bunching implies that workers are willing to forego greater value to avoid the costs of opt-out. Of course, greater opt-out costs also dampen the elasticity of demand for 401(k) contributions without altering the marginal benefits of those contributions; hence, the identification of the two types of information must be simultaneous, rather than sequential (contrary to our

intuitive explanation).

Formally, then, our approach is to fit the data on distributions of 401(k) contributions to a model that is parsimonious on the one hand, while on the other hand sufficiently flexible to accomodate a wide range of possibilities with respect to the elasticity of demand for 401(k) contributions and the distribution of as-if opt-out costs (i.e., the key inputs into welfare calculations). For the indirect utility function, we employ the following functional form:

$$V(x, z, \alpha, \rho) = \rho \ln(x + \alpha) + \ln(z) \tag{8}$$

(so that the vector θ consists of the pair (α, ρ)). The parameter ρ governs the overall division of resources between 401(k) contributions and other uses, while the parameter α governs the sensitivity of the employee contribution rate to the concurrent employer matching rate. With $\alpha = 0$, V is a Cobb-Douglas function in x and z, expenditure shares are fixed, and the employee's contribution rate, r, is unresponsive to a temporary change in an uncapped employer match. In contrast, a temporary increase in an uncapped employer match rate increases the optimal employee contribution rate if $\alpha > 0$, and reduces it if $\alpha < 0$.[26]

As mentioned previously, the existence of a limit on the employer's matching contributions creates a kink-point in the worker's opportunity set. Specifically, $z = 1 - \frac{(1-t)x}{1+m}$ 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 when the worker reaches the cap on matchable contributions,[27] t is the marginal personal tax rate, m is the matching rate, and $\overline{Z} = 1 + (1-t)x_M \left(1 - \frac{1}{1+m}\right)$. One can interpret $\overline{Z} - 1$ as the "virtual income" implicit in the kinked budget constraint when $x \geq x_M$. Throughout, we assume $t = 0.2$ because most workers fell into the 15% or 25% marginal tax brackets during the relevant time period.

Intuitively, ρ is identified by the overall level of workers' contributions, while α is identified from the degree of bunching in the distribution of contributions at the maximum matchable

[26] We could also allow for responsiveness of the employee contribution rate to changes in an uncapped employer match rate 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.

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

contribution rate. We treat α, the parameter governing the elasticity of demand for 401(k) contributions, as common to all workers. To allow for heterogeneity in tastes for 401(k) contributions among workers, we assume $\rho = \max\{\tilde{\rho}, 0\}$, and that the CDF for the random variable $\tilde{\rho}$, denoted F, is normal with mean μ_i, where i denotes the firm, and variance σ^2. Notice that this specification allows average contributions to differ systematically across firms, in recognition of the possibility that some groups of employees are more motivated savers than others.

We also allow for heterogeneity with respect to as-if opt-out costs. Specifically, we assume the CDF for γ, denoted Φ, is a mixture between an exponential distribution and a probability atom at zero:

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

Thus, the parameter λ_1 represents the fraction of workers who act as if opt-out costs is essentially costless. We take the distributions of $\tilde{\rho}$ and λ to be independent.

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

3.2.2 Models with frame-dependent weighting

For models with frame-dependent weighting, decisions in the naturally occurring frame (f^*) are observationally equivalent to those implied by the basic model with costly opt-out, and hence we calibrate those models for f^* in the same way as the basic model. To complete their calibration, we need only specify the range of weights, $\delta(f)$, prevailing in other frames. Notice that it is not possible to deduce these weights from the available data, which pertain *only* to the naturally occurring frame.

For the purpose of our calculations, we will assume that the naturally occurring frame, f^*, maximizes the weight placed on opt-out costs, $\delta(f)$. For the reasons discussed in Section 2.2, that assumption is certainly appropriate in the context of either sophisticated or naive time inconsistency, considering that the contemporaneous frame (with inferred future

consequences in the case of naivete) is naturally occurring. In the context of inattentiveness, the assumption presupposes that employers have taken no special steps to make opt-out decisions salient, and that such steps would have the effect of focusing greater attention on the consequences of opt-out. In light of the enormous opt-out costs implied by the calibrated model (see Section 3.4, below), these assumptions also strikes us as reasonable.

Our welfare analysis also requires us to specify the *minimum* weight attached to opt-out costs in any decision frame. Lacking choice-based evidence, we take that minimum to be one percent of the weight applied in the naturally occurring frame. In other words, we assume there is some frame in which the worker would act as if opt-out were nearly costless. As explained below, this weight implies an objectively plausible dollar-equivalent for the cost of opt-out. In addition, being close to zero, it effectively bounds the range of possibilities.

3.2.3 The model with anchoring

To introduce anchoring, we assume that any given default frame shifts $\widetilde{\rho}$, the parameter governing the worker's ideal division of resources between 401(k) contributions and other uses, toward the value that would rationalize the default contribution rate as an optimal choice. Like switching costs, anchoring effects of this type can produce bunching of choices at the default option. However, switching costs tend to sweep out density near the default, creating a trough in the distribution of choices, whereas anchoring (as we formulate it) tends to shift each half of the distribution of $\widetilde{\rho}$ toward the default, thereby creating a spike without a neighboring trough. Thus, given our assumptions, it is 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.[28] 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 the anchoring parameter. Thus, the anchor shifts a worker's as-if utility weight by the amount ζ toward the weight that rationalizes the default, but not beyond. The default then becomes the as-if ideal point for all individuals with $\widetilde{\rho} \in \{\rho^* - \zeta, \rho^* + \zeta\}$, which implies a spike in the distribution of choices at the default.

3.3 Calibration method

As mentioned above, we fit the model to distributions of employee contribution rates for a sample of firms that changed their default contribution rates without altering other important features of their 401(k) plans, such as match rates. Workers at firm i pick r from a *discrete* set $R^i \equiv \{0, 0.01, 0.02, ..., \overline{r}\}$, and the employer matches contributions at the rate m^i up to r^i_M, so $x^i_M = (1 + m^i)r^i_M$, and $x = r + m^i \min\{r, r^i_M\}$. Thus, $X^i = \{x^i_1, x^i_2, ..., x^i_K\}$ where $x^i_k = 0.01\left[(k-1) + m^i \min\{k - 1, 100r^i_M\}\right]$, and $K = 100\overline{r} + 1$.[29]

For any fixed α and firm i, we can partition the range of ρ into intervals, $B^i_1(\alpha) = [0, \rho^i_1(\alpha)]$, $B^i_2(\alpha) = [\rho^i_1(\alpha), \rho^i_2(\alpha)]$,..., $B^i_K(\alpha) = [\rho^i_{K-1}(\alpha), \infty]$, such that an individual with utility weight ρ and no opt-out costs is willing to choose $x^i_k \in X^i$ iff $\rho \in B^i_k(\alpha)$. With opt-out cost γ and default d, a worker with $\rho \in B^i_k$ at firm i rejects the default iff

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

The probability that a worker at firm i chooses $x^i_k \neq d$ is then

$$\Pr{}_i(x^i_k \mid \psi, \mu_i, d) = \int_{B^i_k(\alpha)} \Phi\left(\Gamma^i_k(\alpha, \max\{0, \widetilde{\rho}\}, d)\right) dF(\widetilde{\rho}) \tag{9}$$

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

[29] So, for example, if $\overline{r} = 0.15$, $r^i_M = 0.06$, and $m^i = 0.5$, then $x^i_M = 0.09$ and $X^i = \{0, 0.015, ..., 0.075, 0.09, 0.1, ..., 0.17, 0.18\}$.

For $x_k^i = d$, we calculate the analogous probability as a residual:

$$\Pr{}_i(d \mid \psi, \mu_i, d) = 1 - \sum_{k \text{ s.t. } x_k \neq d} \int_{B_k^i(\alpha)} \Phi\left(\Gamma_k^i(\alpha, \max\{0, \widetilde{\rho}\}, d)\right) dF(\widetilde{\rho})$$

We label the firms $i = 1, ..., I$. Firm i has S_i default regimes with default d_i^s in regime s. For each firm and default regime s, N_{ik}^s is the number of individuals choosing r_k at firm i in regime k. We do not have information on workers' characteristics; any influence of such factors on tastes enter through the distribution of ρ.[30] The total log-likelihood is:

$$\sum_{i=1}^{I} \sum_{s=1}^{S_i} \sum_{k=1}^{K} N_{ik}^s \log\left[\Pr{}_i(x_k \mid \alpha, \lambda_i, d_i^s)\right].$$

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

For the anchoring model, we simply replace (9) with

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

3.4 Estimates, interpretation, and fit

Estimates of the basic model appear in Table 2. All parameters are estimated precisely. The mean utility weight for each company accords with average 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%, 3%, and 5%. The associated standard deviation 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. The mean of γ (among the 60% of workers with positive opt-out costs) 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, evaluated in a setting without 401(k) eligibility, is given by $v(\gamma)$, the solution to $V(0, 1 - v(\gamma), \theta) =$

[30] Data on worker characteristics would allow us to compute the welfare effects of defaults for separate subgroups, but it would not alter aggregate welfare effects or the determination of the default rate that maximizes total economic surplus.

$V(0,1,\theta) - \gamma$. For specification (8), $v(0.0847) = 0.0812$ and $v(0.0587) = 0.0567$. If, as an approximation, we construe the data as representing decisions taken over the first year of eligibility during which the worker earns \$40,000, the monetary equivalent of γ is more than \$3,200 at the mean of the distribution and more than \$2,200 at the median. Yet it is difficult to believe that more than a handful of employees would actually turn down a payment of, say, a hundred dollars, let alone several thousand, to avoid making an active 401(k) election. Considering the effort involved, we would place a plausible dollar-equivalent for the cost of opt-out at roughly one percent of the estimated as-if cost (i.e., \$25 to \$30); that is why our calibration assumes that the minimum weight attached to opt-out costs in any decision frame is one percent of the weight applied in the naturally occurring frame.

Why does the basic model require enormous opt-out costs to rationalize observed behavior? For those who would save for retirement even without a 401(k), the EV associated with 401(k) eligibility must be very large due to matching provisions and tax deductibility. To explain why many such individuals stop contributing when the default rate falls from 3% to 0%, one must assume that opt-out costs are extremely high. Notably, DellaVigna (2009) reached a similar conclusion based on a back-of-the-envelope calculation concerning the economic value of the employer matching contributions, which he placed at \$1,200 (for a worker earning \$40,000). The rough similarity between that figure and our estimated as-if opt-out costs is reassuring.

Plainly, the enormous size of as-if opt-out costs calls for a behavioral explanation. Sophisticated time inconsistency seems a poor candidate: only a value of β_0 much smaller than documented in the literature would render the implied distribution of γ plausible. Nor does naive time inconsistency strike us as a likely explanation: naivete would only help to rationalize large values of γ if the deadlines for changing 401(k) elections were frequent (e.g., biweekly); but then, workers would presumably learn from numerous failures to follow through on intentions over the course of more than a year. Still, we do not exclude either explanation.

Consider next the model that allows for anchoring effects (also Table 2). The estimates of α, μ_1, μ_2, μ_3, and σ change relatively little. The estimate of ζ reflects a large and statistically significant as-if anchoring effect: anchoring can shift the utility weight (ρ) by up to roughly two-thirds of its standard deviation (σ). Significantly, the estimated as-if opt-out cost distribution changes dramatically. Only an estimated 10.9% of workers act as if opt-out cost are negligible. However, 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 – on the order of one-tenth of a percent of earnings in both cases. For an employee earning $40,000 per year, the monetary equivalent of γ is therefore $54 at the mean and $37 at the median. Those magnitudes strike us as reasonable estimates of the amount a typical worker would be willing to accept in exchange for taking the time to fill out a few forms. Consideration of our calibrated models therefore suggests that bunching at the default option is primarily attributable to anchoring rather than to opt-out costs (but we do not make that assumption in what follows).

Figure 1 illustrates, for the basic model, the fitted and actual distributions of employee contribution rates under each default regime for each of the three companies. The model generally performs well, reproducing the spikes in the distributions at 0%, the default option, the maximum matchable contribution rate, and the overall cap (though predictably missing some smaller spikes at 10%). For the anchoring model, the fit (not shown) is slightly better.

4 Welfare criteria

We conduct welfare analysis using the framework proposed by Bernheim and Rangel (2009), henceforth BR, which generalizes the standard normative paradigm to non-standard settings under the interpretation that welfare is defined directly in terms of choice, rather than underlying objectives (on the grounds that the latter may not be recoverable; see Bernheim, 2009). Its use involves three steps: first, specify the set of "welfare-relevant" choices; second, construct the welfare criterion; third, apply it to the problem of interest.

To encompass non-standard choice patterns, BR define a *generalized choice situation* (abbreviated GCS), $G = (X, f)$ as a constraint set X and a psychological frame f.[31] A psychological frame is a condition under which a decision is made, rather than a condition of experience, that affects choice. Possible examples include (but are not limited to) the point time at which a choice is made or the way information is presented. Either a theory or data provide us with a choice correspondence C defined on some domain of GCSs, $\mathcal{G}^*$. Choices may exhibit anomalies such as frame-dependence, intransitivities, and choice reversals.

The first step is to specify a welfare-relevant domain, $\mathcal{G} \subseteq \mathcal{G}^*$. In some contexts we may accept all GCSs as welfare relevant ($\mathcal{G} = \mathcal{G}^*$), but in others we may refine that domain. For example, BR argue for excluding a choice from the welfare-relevant domain when there is evidence that the decision maker incorrectly understood his or her opportunity set (i.e., cases of *characterization failure*).

The second step in the BR framework is to construct the welfare criterion. If the choice correspondence satisfies WARP on $\mathcal{G}$, it can be represented by a standard preference relation, and one can proceed as in the standard normative paradigm. However, if C violates WARP on $\mathcal{G}$, one must proceed differently. BR define the unambiguous choice relation, P^*, as follows: xP^*y iff y is chosen in no GCS where x is available. P^* generalizes the standard (strict) revealed preference relation P, in the sense that the two coincide when the choice correspondence satisfies WARP on the welfare-relevant domain. BR show that P^* satisfies a collection of desirable properties for a choice-based welfare criterion, and in fact is the only welfare criterion that does so. 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).

Among other desirable features, the BR framework yields generalizations of the standard tools of applied welfare economics, including equivalent and compensating variation, consumer surplus, and Pareto optimality. A generalized notion of equivalent or compensating

[31] Bernheim and Rangel (2009) used the term "ancillary condition" rather than psychological frame.

variation must accomodate any ambiguity in the welfare criterion. Accordingly, for a change from policy p to policy p', BR define EV_A as the smallest (in the sense of infimum) increment to income with p such that the bundle obtained with p is unambiguously chosen over (P^*) the bundle obtained with p'. Similarly, EV_B is the largest (in the sense of supremum) increment to income with p that such that the bundle obtained with p' is unambiguously chosen over (P^*) the bundle obtained with p. It is always the case that $EV_A \geq EV_B$, and the two coincide with the standard measure of equivalent variation when C satisfies WARP on $\mathcal{G}$. Thus, one can say that the policy change is unambiguously worth at least EV_B, and no more than EV_A. BR generalize compensating variation similarly.

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. One can find conventional Pareto optima by maximizing either the weighted sum of utilities or, because equivalent variation is a monotonic transformation of utility, the weighted sum of EVs.[32] The following result, upon which we rely heavily in the next section, generalizes this property when P^* is transitive (which holds for many behavioral models, including those considered here):

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.*

We have articulated each behavioral theory in Section 2 by explicitly defining psychological frames and providing a model of frame-dependent choice. Thus, once we specify the welfare-relevant domain $\mathcal{G}$, application of the BR framework is straightforward.[33] Indeed,

[32] 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.

[33] When applying the BR framework to a particular model, we limit consideration to $\mathcal{G}^*$, the choice domain encompassed by the model. Stepping outside the domain of the model, behavior may exhibit other non-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.

for the particular models considered here, it turns out that P^* is equivalent to the multiself Pareto criterion, treating each frame as a different self (BR, Theorem 3).[34]

There are, of course, potential theory-specific justifications for choosing a welfare-relevant domain, $\mathcal{G}$, that excludes portions of a model's choice domain, $\mathcal{G}^*$. Most obviously, in a model of inattentiveness, it may seem natural to restrict $\mathcal{G}$ to choices made in a frame f with $\chi(f) = 0$ so that the social planner does not emulate neglectful decision makers.[35] Yet caution is warranted, inasmuch as the available empirical evidence may sustain only an as-if interpretation of the model, rather than a literal one. For instance, interventions *intended* to manipulate attentiveness may influence choices through other mechanisms, e.g., by browbeating or embarassing the decision maker. Similarly, in the context of time inconsistency, it is sometimes argued that welfare should be assessed in the forward-looking frame (i.e., according to the "long-run" criterion; see, e.g., BR, Theorem 11). Yet it is also arguable that people tend to overintellectualize their choices when making decisions at arms length, and that they properly appreciate experiences only "in the moment." In light of these possibilities, one may wish to avoid taking a stand on which choice frame is the "correct" one for the purpose of welfare evaluation. The BR framework permits one either to take a stand or to remain agnostic.

Before proceding to our welfare analysis, a final word concerning the BR framework is in order. Specifically, BR demonstrate that their framework has an attractive continuity property: if one conducts welfare analysis based on a choice correspondence that is approximately correct, the normative conclusions that emerge will also be approximately correct. Accordingly, for our current purposes, what matters is not whether our calibrated models accurately depict workers' decision *processes* (i.e., the ways in which decisions are reached),

[34] That result does not apply with generality to the familiar quasi-hyperbolic model of time-inconsistency because it requires that one can write $\mathcal{G}$ as the Cartesian product of a set of frames and a set of opportunity sets (which is not possible for the model in question because decisions made at any given point in time cannot affect past consumption). However, for the reduced-form model of time inconsistency described in Section 2.2, $\mathcal{G}^*$ *can* be written as the requisite Cartesian product because nothing is consumed in period -1, and thus the result *does* apply.

[35] If no such frame exists in practice, one could in principle impute the associated choices by observing the decisions people make when they are attentive.

but rather simply whether they capture the mapping from choice settings to decisions with a reasonable degree of accuracy.

5 Welfare analysis

5.1 Welfare analysis in models with frame-dependent weighting

We now examine the welfare effects of default options for our models with frame-dependent weighting. As our measure of welfare, we compute the equivalent variation associated with switching from some initial regime to one in which the worker becomes eligible for the 401(k) in period 0 with an initial default rate of d. Recognizing that the reduced form utility function V implicitly presupposes the availability of a 401(k) plan with a default rate of x from period 1 onward, we compute EV based on an initial regime in which the worker cannot contribute to a 401(k) in the current period, but can do so in future periods (with an initial default of zero). Our analysis focuses on five issues: the degree of ambiguity in welfare evaluation; the identity of the optimal default; the size of the stakes; the desirability of using penalties to encourage active decision making; and the desirability of allowing workers to precommit to making benefit elections.

1. The degree of ambiguity. A potential concern in applying the BR framework is that welfare analysis can prove to be highly ambiguous, and hence of little value, when evaluations vary sharply over decision frames within the welfare-relevant choice domain. In that case, to perform a discerning evaluation, one would need to adopt (and justify) some strong refinement of that domain. In effect, that is the approach adopted by CCLMM (who examine the welfare effects of default options from the perspective of the forward-looking frame in a model with time inconsistency).

The need for a refinement may seem apparent from our calibrated model with frame-dependent weighting: EV_B reflects the naturally occurring frame, in which opt-out is treated as having an enormous cost (averaging thousands of dollars for the typical worker), while EV_A reflects a frame in which only 1 percent of that cost is treated as "real." Surprisingly,

however, our first main finding is that the decision frame used for welfare evaluation makes only a modest difference over the pertinent range for default options. As a result, the degree of ambiguity concerning welfare is relatively small, even if one accepts all decision frames as welfare-relevant, and one can reach useful conclusions concerning the welfare effects of default options for this class of models without taking a potentially controversial stand on the "correct" welfare perspective.

Specifically, using our estimates of the basic model (which represents choices in the naturally occurring frame for our models of time inconsistency and inattentiveness), we simulate workers' choices for various default rates and conduct welfare analysis. Figure 2 graphs, as functions of the default rate, aggregate EV_B, equivalent variation evaluated in the naturally occurring frame, and EV_A, equivalent variation evaluated in a decision frame for which the weight on opt-out costs is reduced by 99 percent, both expressed as fractions of the typical worker's income, for each of the three firms. (It also graphs two opt-out frequencies, which we discuss below.) The most striking feature of the figure is that the scope of ambiguity concerning welfare, $[EV_B, EV_A]$, 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 (which equals 6%, the maximum matchable contribution rate, in all cases).

The explanation for this surprising finding is straightforward: for the range of default rates considered, the populations of opt-outs are dominated by workers whose opt-out costs are zero or relatively small. Therefore, even though average population-wide opt-out costs are enormous from the perspective of the naturally occurring frame, heavily discounting *incurred* opt-out costs makes little difference. Matters change, of course, when a broader range of default options is considered. With an extremely high default option (say, 70% of earnings), virtually all workers will opt out, including those with very high opt-out costs; hence the difference between EV_A and EV_B becomes quite large (see the online appendix). But as a practical matter such extreme default rates are not relevant, because they exceed

statutory limits on 401(k) contributions.

2. The optimal default. As mentioned above, aggregate EV_A and EV_B are both maximized for a default rate equal to the maximum matchable contribution rate (6%) at all three companies.[36] Figure 2 also shows two opt-out frequencies, the ratio of all opt-outs to all workers ("overall opt-out frequency") and the ratio of opt-outs among those with zero opt-out costs to all workers ("zero-cost opt-out frequency"). Note that in all cases the opt-out frequencies are minimized at the EV-maximizing contribution rate. Accordingly, the Thaler-Sunstein rule of thumb is a good policy guide in this context.

A natural question is whether the coincidence of the EV-maximizing and opt-out-minimizing default rates reflects a general principle, or is instead an artifact of their separate coincidence with the maximum matchable contribution rate. Certainly, achieving low opt-out is advantageous from a welfare perspective because it avoids the real economic costs associated with forcing workers to make adjustments. Moreover, 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 (0) and maximum ($\overline{x}$) contribution rates (due to truncation of the distribution of ideal points), and any contribution rates corresponding to convex kink points in workers' opportunity sets (i.e., the match cap). Thus, as a general matter, one would expect to see some tendency for the EV-maximizing default rate to coincide with one of those values.

With small opt-out costs, the preceding conjecture holds with considerable generality. 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 recall that the support of H^γ is $[0, \overline{\gamma}]$.

[36] What then of the finding in Carroll et al. (2009) that an extreme default is optimal from the forward-looking perspective when time inconsistency is sufficiently severe, as we assume it is in our calibrated 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, the EV_A reaches a minimum at a default rate near 30%, and then increases monotonically, achieving a plateau and a global maximum for default rates above 90% (see the online appendix).

Theorem 2: *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 $e^* > 0$.*[37] *The EV-maximizing default rates, d_k^*, converge to a point in $\mathcal{A}$.*

Interestingly, Theorem 2 provides a potential justification for setting an extreme default rate (0, which most plan sponsors used historically, or $\overline{x}$). In that sense, it seems reminiscent of the result due to Carroll et al. (2009). However, the logic of the result is entirely different. In Carroll et al. (2009), an extreme default is used to maximize active decision making (i.e., opt-out); here, it is optimal for precisely the opposite reason.

Though Theorem 2 identifies a connection between EV-maximization and opt-out minimization, it does not imply that the two are generally the same. In the first place, opt-out costs may not be small; in the second place, there is no guarantee that EV is maximized at the point of *greatest* accumulation within $\mathcal{A}$. Indeed, as a matter of theory, the divergence between the EV-maximizing and opt-out minimizing default rates can be arbitrarily large. To understand why, observe that the EV-maximizing default rate *only* depends on the preferences of workers with positive as-if opt-out costs. Thus, the "zero-cost opt-out frequency" curves in Figures 2 and 3 have no bearing on the optimal defaul rate; only the opt-out frequency among those with positive opt-out costs matters. But if the fraction of workers with zero as-if opt-out costs is large, the opt-out-minimizing default rate will depend only on the preferences of workers with zero opt-out costs. More generally, the variation in the overall opt-out frequency over default rates is primarily driven by those with low opt-out costs (because their opt-out decisions are more sensitive to the default rate – see, e.g., Figure 2). But those are precisely the workers for whom the default rate is least important.

We explore these issues quantitatively by simulating workers' choices and evaluating welfare effects with alternative employer matching provisions. First, we remove the matching provisions for the current period.[38] To simulate choices, we simply change the worker's

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

[38] Our reduced-form approach does not allow us to simulate the effects of changes in future matching provisions on V.

opportunity constraint to $z = 1 - tx$.[39] Figure 3 displays the results. Naturally, the EVs fall dramatically. Once again, the frame of evaluation matters, but (as discussed above) to a much smaller extent than we had originally anticipated. The EV_A-maximizing default now differs considerably across the companies, reflecting the differences in μ_i: it is 13% for company 1, 2% for company 2, and 6% for company 3 (mirroring the median ideal contribution rates of 11%, 3%, and 5%, as reported in Section 3.4). The EV_B-maximizing default rates are similar: 14% for company 1, 0% for company 2, and 6% for company 3. Notice that, except in one case, these rates do not coincide with the minimum or maximum contribution rates. Nor do they generally coincide with the opt-out-minimizing default rates, which are 15% for company 1 and 0% for companies 2 and 3. For company 3 in particular, opt-out minimization yields a default rate that differs significantly from the optimum.

Second, we simulate choices and evaluate welfare effects with match caps other than 6%. Figure 4(a) plots the EV_A-maximizing default rate as a function of the match cap for each of the three companies, while Figure 4(b) plots the EV_B-maximizing default rate. In each case, the EV-maximizing default rate 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). Thus, setting the default rate equal to the match cap is a good rule of thumb in some instances, but not in others. Note also the similarity between Figures 4(a) and 4(b): generally, the frame of evaluation does not have much bearing on the welfare-optimal policy.

3. The stakes. With a default rate of 0% and a match cap of 6%, and assessing welfare from the perspective of the naturally occurring choice frame (i.e., EV_B), the equivalent variation associated with the opportunity to make 401(k) contributions in the current period

[39] 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 r_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 r_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.

(shown in Figure 2) roughly equals total employer contributions plus 20% to 40% of employee contributions (depending on the firm), a range which strikes us as plausible. Setting the default contribution rate optimally (i.e., at 6%) rather than at 0% raises EV_B by 2.89% of earnings (from 7.07% to 9.86%) for company 1, by 0.74% of earnings (from 1.97% to 2.71%) for company 2, and by 1.18% of earnings (from 2.75% to 3.93%) for company 3. In each case, 27% to 30% of the potential economic surplus flowing from the 401(k) is lost when the default rate is inefficiently set to zero. For EV_A, the implied gains are still substantial, but smaller: the portion of the potential economic surplus lost with a default rate of zero ranges between 12% and 18%.

When we simulate the effect of eliminating the match (as in Figure 3), the welfare cost of setting a default rate of zero rather than the optimal rate shrinks considerably. For EV_B, it equals 1.43% of earnings or 40% of the potential economic surplus flowing from the 401(k) at company 1, 0.03% of earnings or 3% of the potential economic surplus at company 2, and 0.26% of earnings or 16% of the potential economic surplus at company 3. For EV_A, the welfare costs of setting a default rate of zero are even smaller: 0.65% of earnings or 15% of the potential economic surplus flowing from the 401(k) at company 1, zero at company 2 (because a default of zero is optimal), and 0.05% of earnings or 2.6% of the potential economic surplus at company 3. Thus, setting a positive default rate is normatively consequential primarily because of employer matching provisions.

As we have noted, opt-out minimization typically would not be optimal for these companies in the absence of matching provisions. However, the welfare costs of following the Shefrin-Sustein rule of thumb are relatively small: for companies 1, 2, and 3, respectively, the losses are 0.17%, 0.03%, and 0.26% of earnings according EV_B, and 0%, 0%, and 0.05% of earnings according to EV_A.

4. Penalties for passive choice. CCLMM raise the possibility that, with a sufficient degree of present bias, 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 contribution rate. Plainly, it is possible to accomplish the same objective by setting a large penalty for passive choice; indeed, CCLMM mention that one can think of an extreme default as standing in for that alternative. 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 instead of an extreme default (or equivalently an extreme penalty that compels active choice).

To address that issue, we perform simulations for our model of time-inconsistent choice, in which we optimize over penalties and default rates simultaneously, evaluating welfare from the perspective of the forward-looking decision frame. We find that the optimal penalty is enormous (roughly 65% of earnings) and the default is of essentially no consequence. We can overturn the latter result by assuming that some small fraction of the population, η, never makes an active decision. But as we increase η from zero, the optimum changes sharply (e.g., at around $\eta = 0.034$ for company 1 with a match) from a policy that involves an extreme penalty (so that the default is virtually irrelevant) to one with no penalty and attractive default (so that active decisions are not compelled).

Figure 5 shows why the optimum never involves a strictly positive penalty small enough so that the default contribution rate is meaningfully consequential. Fixing a default rate of 6%, the figure graphs average EV_A for company 1 against the size of the penalty (measured as a fraction of earnings) with η ranging from 1% to 5%.[40] 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. Welfare analysis with precommitment opportunities. If default effects result from *sophisticated* time inconsistency, and if one evaluates welfare from the perspective of the forward-looking decision frame, a policy that compels active decision making is plainly inferior to one that provides workers with appropriate precommitment opportunities. To put

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

the matter another way, it is better to allow each worker to decide, in advance, whether he or she will be subject to a large penalty for passive choice, rather than to impose that penalty indiscriminately. Indeed, the potential benefit of precommitment opportunities is an important theme in the literature on time inconsistency.[41] However, as we show next, such opportunities can have a previously unrecognized down-side: to the extent one regards the forward-looking choice frame as a controversial welfare standard, offering precommitment opportunities creates *substantial ambiguity* concerning the welfare effects of alternative default policies. In other words, while welfare evaluations are similar from the perspective of the contemporaneous and forward-looking frames when the decision to opt out is actually made in the contemporaneous frame, they are very different when that decision is actually made in the forward-looking frame. A similar point holds in the context of our attentiveness model: to the extent one regards the "fully attentive" frame as a controversial welfare standard (e.g., because one is not convinced that the actual cognitive mechanism behind the as-if model involves attention), adopting a policy that places workers in that frame when they make their opt-out decisions also creates substantial ambiguity concerning the welfare effects of alternative default policies.

Figure 6 (which presupposes a 50% match on contributions up to 6% of earnings) and Figure 7 (which presupposes no match) are identical to Figures 2 and 3, except that (a) the welfare evaluations presume that opt-out decisions are made in the forward-looking (or fully attentive) frame, rather than in the naturally occurring frame, and (b) we have not plotted the zero-cost opt-out frequency. Notice first that, in all cases, the gap between EV_A and EV_B is enormous. Indeed, in some cases, there is actually ambiguity as to whether the 401(k) plan creates or *destroys* value. In addition, if one evaluates welfare from the perspective of the forward-looking (or fully attentive) frame, the choice of default rate is of little consequence. However, if one evaluates welfare from the perspective of the contemporaneous frame, the opt-out minimizing default rate is in all cases strongly preferred. Thus, reducing welfare

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

ambiguity emerges as a potential reason to prefer policies that force workers to make opt-out decisions in the contemporaneous (or inattentive) frame.

The explanation for this surprising increase in welfare ambiguity is straightforward. When opt-out decisions are actually made in the forward-looking (or fully attentive) frame, the vast majority of workers opt out (as shown in Figures 6 and 7) – even those who have extremely high as-if opt-out costs in the naturally occurring frame. Consequently, whether one evaluates these outcomes according to the naturally occurring frame or the alternative frame is enormously consequential. In contrast, welfare ambiguity is relatively low when opt-out decisions are made in the naturally occurring frame because, in that case, those with high as-if opt-out costs do not actually incur those costs, because they do not opt out.

5.2 Welfare analysis with anchoring

Because our calibrated model with anchoring involves low opt-out costs, one can gain an understanding of the welfare calculations reported below by considering the special case where those costs are zero:

Theorem 3: *Assuming $\gamma = 0$, every default rate d maximizes EV for every worker evaluated from the perspective of the frame $f = d$. With $\mathcal{G} = \mathcal{G}^*$, EV_A is non-decreasing in d on $[0, \overline{x}]$ and maximized at $d = \overline{x}$, while EV_B is non-increasing on $[0, \overline{x}]$ and maximized at $d = 0$.*

From this theorem, it follows immediately that, unless one adopts a refinement of the welfare-relevant domain, one cannot say that any default rate is unambiguously better than any other.

Figure 8 (which presupposes a 50% match on contributions up to 6% of earnings) and Figure 9 (which presupposes no match) graph, as functions of the default rate, aggregate EV_A, equivalent variation evaluated in the decision frame $f = \overline{x}$, and EV_B, equivalent variation evaluated in a decision frame $f = 0$, for each of the three firms. Just as Theorem 6 suggests, EV_A is increasing in d and maximized at $d = \overline{x}$, while EV_B is decreasing in d

and maximized at $d = 0$. As shown in the figure, the EV-maximizing default rates are unrelated to opt-out minimization or match caps. The substantial gaps between EV_A and EV_B, indicating substantial welfare ambiguity, are direct reflections of the large anchoring effects. The gap is smallest for $d = 0$, but is large even in that case: in Figure 8, 20.6% vs. 8.5% of earnings for company 1, 7.7% vs. 2.5% for company 2, and 9.7% vs. 3.4% for company 3.

Figures 8 and 9 appear to suggest that a policy maker who wishes to "play it safe" should set $d = 0$, so that EV_B, the lower bound on the benefit that the average worker receives, is maximized. However, that finding depends on the identity of the "status quo" policy used in the EV calculations. If instead of a status quo policy involving no current 401(k) eligibility we had used one involving a required 401(k) contribution equal to the overall contribution cap, then EV_A would entail evaluation in the decision frame $f = 0$ while EV_B would entail evaluation in the decision frame $f = \overline{x}$; the EV_B-maximizing (and apparently safest) choice would then be to set the default equal to the cap.

In this setting, the large differences between EV_A and EV_B limit our ability to make precise welfare statements unless we adopt a domain restriction. A potential strategy for restricting $\mathcal{G}$ is to admit choices only if they are made in an arguably neutral frame where choices are free from the influence of anchors. Formally, we define the neutral frame as one in which $\zeta = 0$. That definition appears reasonable under a literal interpretation of the as-if utility function, but a more complete understanding of anchoring effects would be required to justify it.[42] In Figure 8 and 9, equivalent variation evaluated in the putative neutral frame is labeled EV_N.

Strikingly, all the EV_N curves in Figures 8 and 9 are virtually flat. With a match (Figure 8) EV_N is roughly 14.5% of earnings for company 1, 5% for company 2, and 6.5%

[42] For example, that definition 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.

for company 3, regardless of the default rate. Technically, EV_N is maximized at 9% for company 1, and at the match cap (6%) for companies 2 and 3, but the welfare loss from inefficiently setting a default of zero is only 0.35% of income for company 1 and 0.15% for companies 2 and 3. Results without a match (Figure 9) are generally similar.[43]

Up to this point, we have limited our discussion to employee welfare and have steered clear of broader statements concerning social welfare (which also encompasses effects on employers and government revenue), primarily because we are unable to measure the impact of default effects on the PDV of revenues accurately with the available data. However, it is obvious that, by increasing contributions, a higher default imposes costs on both employers (through matching provisions) and the government. Consequently, if worker welfare is largely unaffected by the default contribution rate (as is the case for EV_N in Figures 8 and 9), $d = 0$ emerges as the social optimum.

5.3 An observation concerning the Pareto criterion

So far, we have treated the task of selecting a 401(k) default contribution rate as an exercise in optimal *de novo* policy design. From the perspective of an employer operating a going concern, it is more properly considered an exercise in optimal policy *reform*. As noted by Feldstein (1976), the problem of reform differs from that of *de novo* design in that it depends on the starting point.

In the current context, when contemplating the introduction of opportunities to save through 401(k) accounts, an employer may prefer, if possible, to design a plan so that its existence makes no worker worse off (the "Pareto improvement criterion"). In this section, we make the point that a given plan meets this criterion if and only if $d = 0$.[44] This observation provides a potential justification for setting the default contribution rate equal to zero.

[43] The optimal default rates differ, however: EV_N is maximized at the contribution limit for company 1, 0% for company 2, and 8% for company 3.

[44] 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.

It is trivial to verify the preceding claim in the context of a standard model with opt-out costs, absent "behavioral" considerations such as those discussed in previous sections.[45] However, with frame-dependent weighting or anchoring, some additional considerations arise, and the principle is not completely general. Still, under some additional technical assumptions concerning monotonicity (see Section 2.4), the (weak generalized) 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 be the welfare-relevant frame *least* conducive to contributing (f_M). Here we assume that the welfare-relevant domain can be written as $\mathcal{G} = \mathcal{X} \times \mathcal{F}$; f_M is then defined as the largest element of $\mathcal{F}$ for the cases of time inconsistency and inattentiveness.

Theorem 4: *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,* $f_D \geq f_M$. *Moreover, for those same cases,* $(d, f) = (0, f_M)$ *creates a weak generalized Pareto improvement over* $(d, f) = (0, f)$ *for any* $f > f_M$.

6 Concluding remarks

This paper has investigated the welfare effects of default contribution rates in 401(k) pension plans under various behavioral assumptions concerning the sources of default effects, using the welfare framework proposed by Bernheim and Rangel (2009). We summarized our main conclusions in the introductory section and will avoid repeating them here.

Naturally, the paper leaves many important questions unanswered. More research is required to distinguish empirically between the choice patterns associated with the various

[45] The explanation is straightforward: with $d = 0$, no worker can be worse off because each has the option not to contribute without incurring 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.

theories of default effects, and to justify the restrictions on the welfare-relevant domain that are in some cases required to obtain usefully discerning conclusions. There may also be other explanations for default effects that we have not yet explored. For example, the opt-out costs captured by our models are properly interpreted as the costs of implementing a decision, rather than the costs of reaching the decision. 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; whether that problem is worth solving requires him to solve yet another problem; 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, and Antonio Rangel (2007), "Toward Choice-Theoretic Foundations for Behavioral Welfare Economics," *American Economic Review Papers and Proceedings* 97(2), 464-470.

[4] 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.

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.

[13] 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.

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

[15] 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.

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

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

Table 1: Description of the companies

Parameter	Company 1	Company 2	Company 3
Default regimes	3%, 6%	0%, 3%, 6%	3%, 4%
Matching rate	100%	50%	50%
Maximum matchable contribution	6%	6%	6%
Contribution limit	15%	15%	Up to 25%, censored at 18%
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 2: Estimated Models

Parameter	Description of parameter	Basic Model	Basic Model with Anchoring
α	Retirement saving shift parameter	0.1340 (0.0023)	0.1027 (0.0680)
μ_1	Mean utility weight, company 1	0.2150 (0.0079)	0.2155 (0.0263)
μ_2	Mean utility weight, company 2	0.1313 (0.0016)	0.1260 (0.0419)
μ_3	Mean utility weight, company 3	0.1570 (0.0023)	0.1487 (0.0214)
σ	Standard deviation of utility weight	0.0910 (0.0005)	0.1222 (0.0369)
λ_1	Fraction of employees with zero opt-out costs	0.4011 (0.0021)	0.1094 (0.0422)
λ_2	Opt-out cost distribution parameter	11.81 (0.16)	747.2 (199.4)
ζ	Anchoring parameter		0.0785 0.0209
Log Likelihood		-2.825×10^5	-2.805×10^5

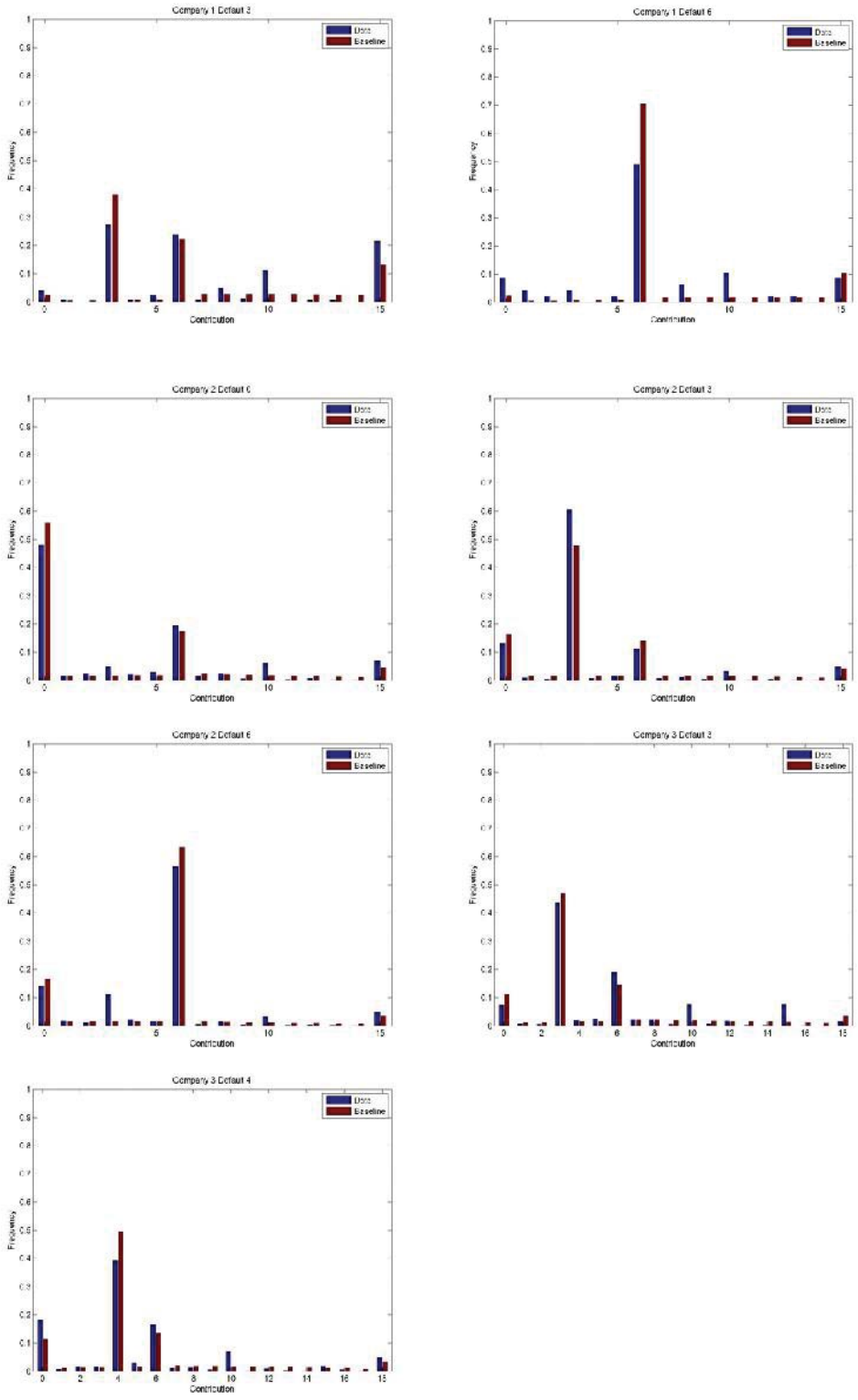

Figure 1: Fitted versus actual distributions

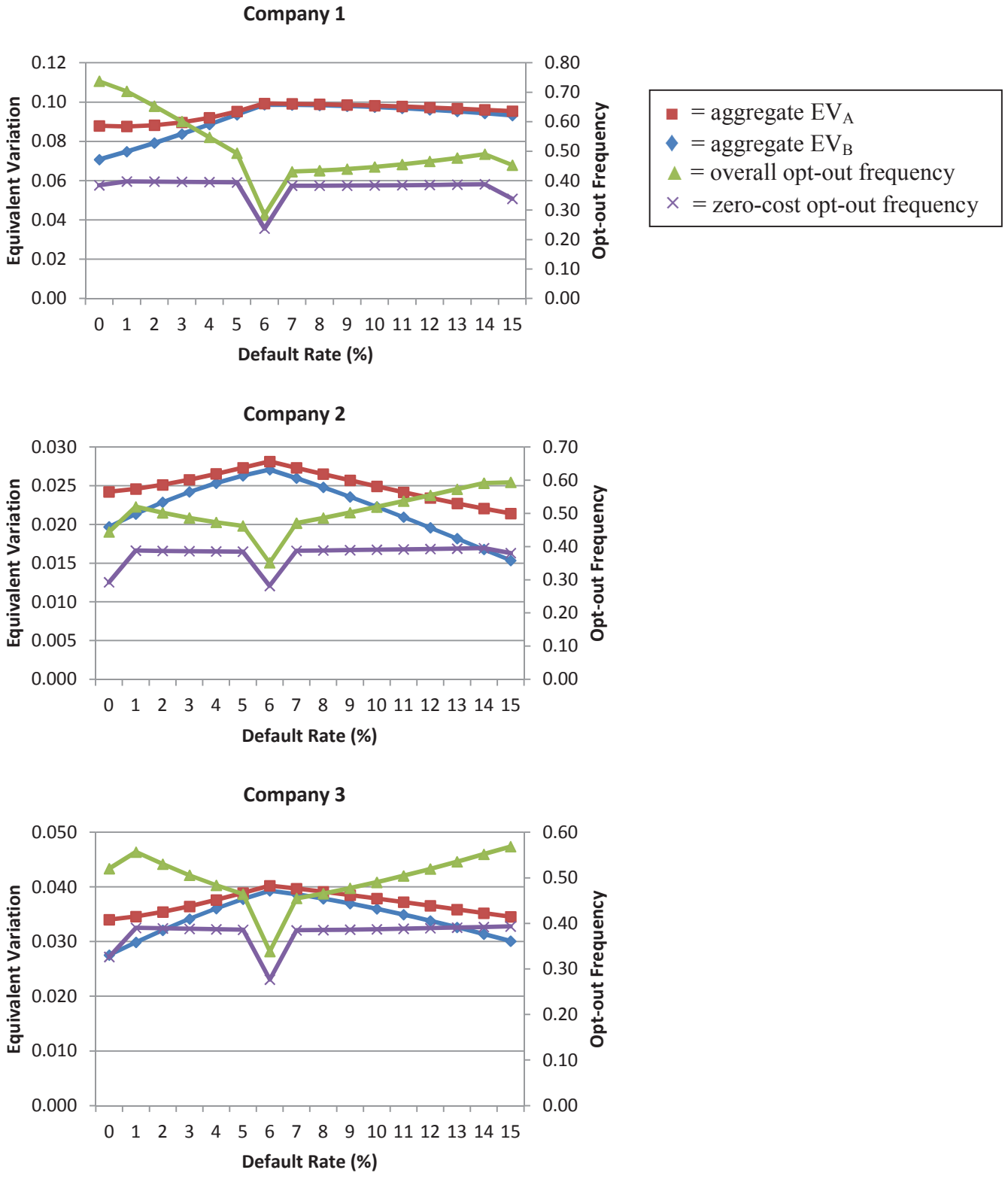

Figure 2: Average equivalent variations and opt-out frequencies, with employer match

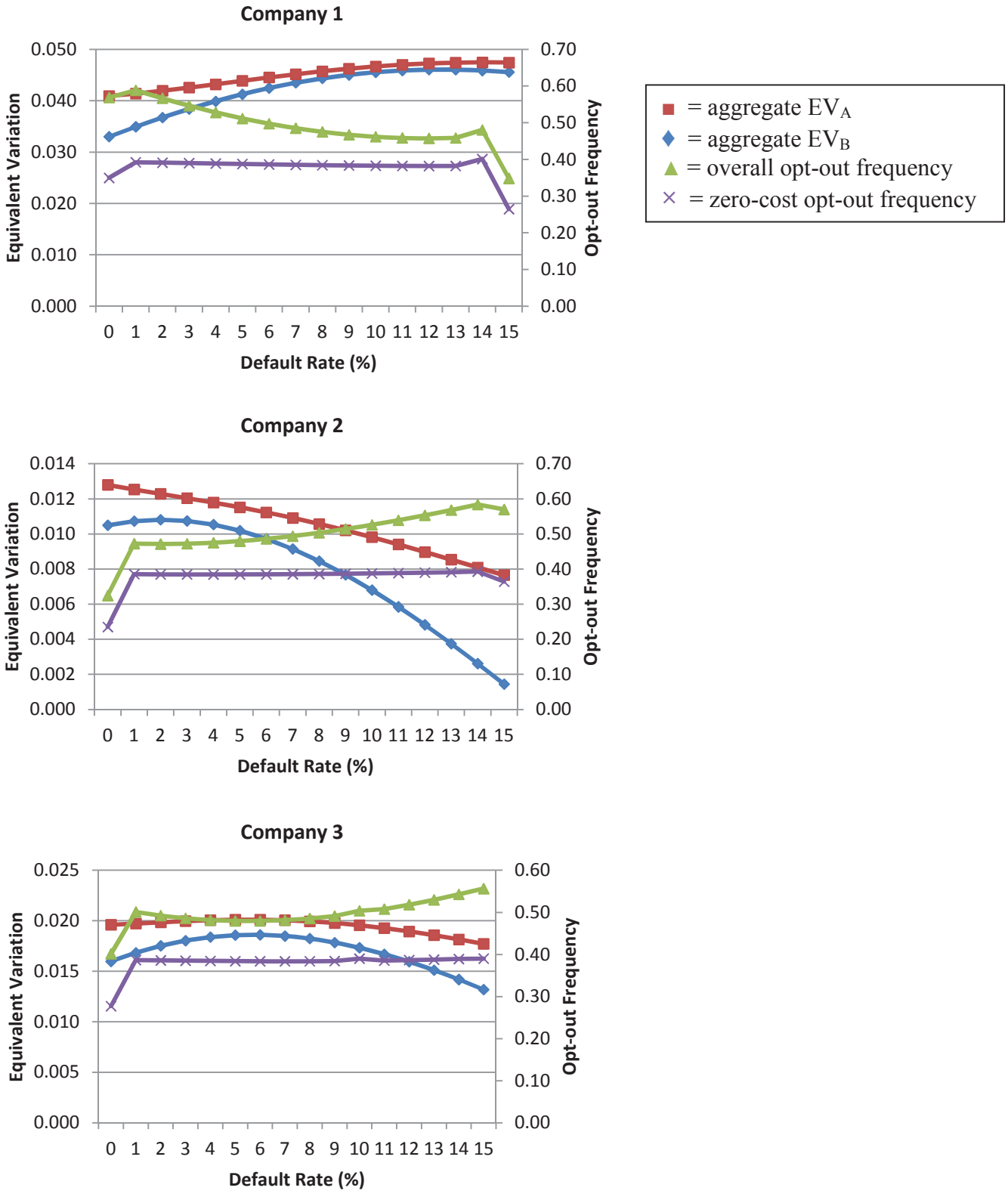

Figure 3: Average equivalent variations and opt-out frequencies, without employer match

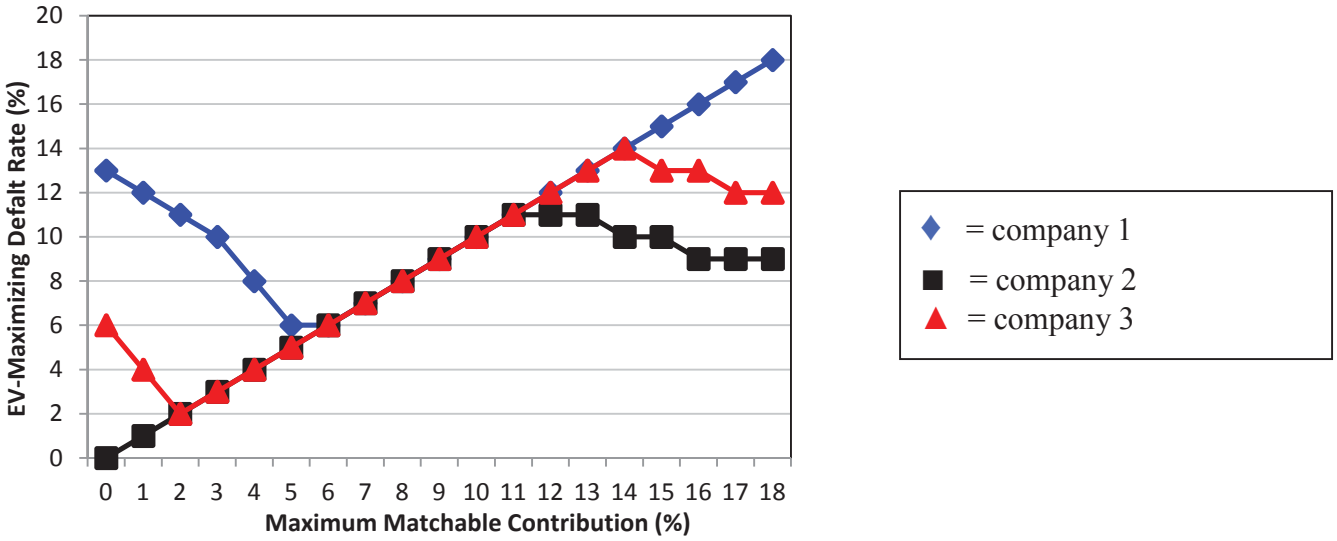

Figure 4(a): EV_A-maximizing default rate versus maximum matchable employee contribution.

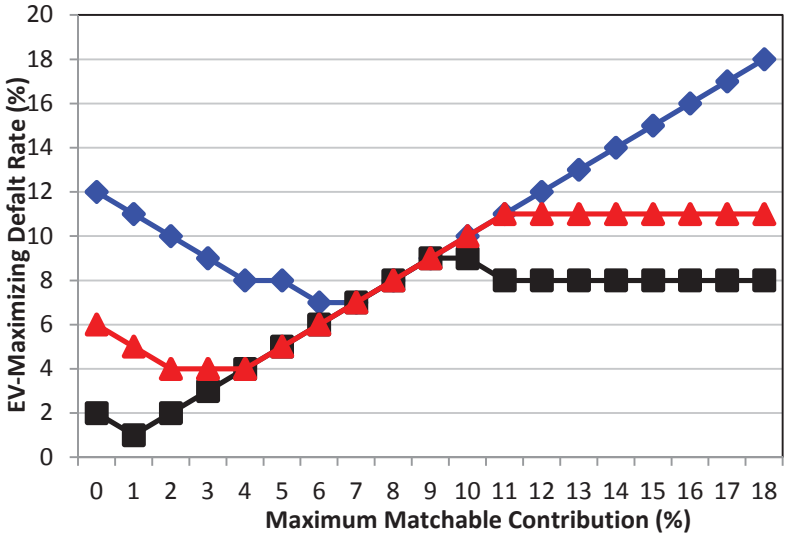

Figure 4(b): EV_B-maximizing default rate versus maximum matchable employee contribution.

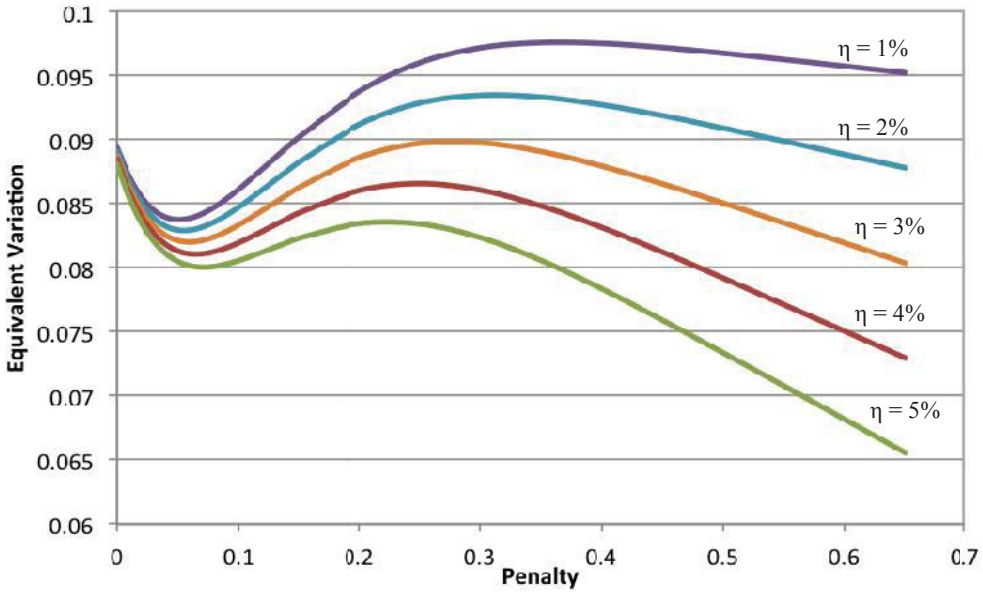

Figure 5: Average equivalent variation as a function of the penalty for inactive choice, with the default rate fixed (company 1, with employer match)

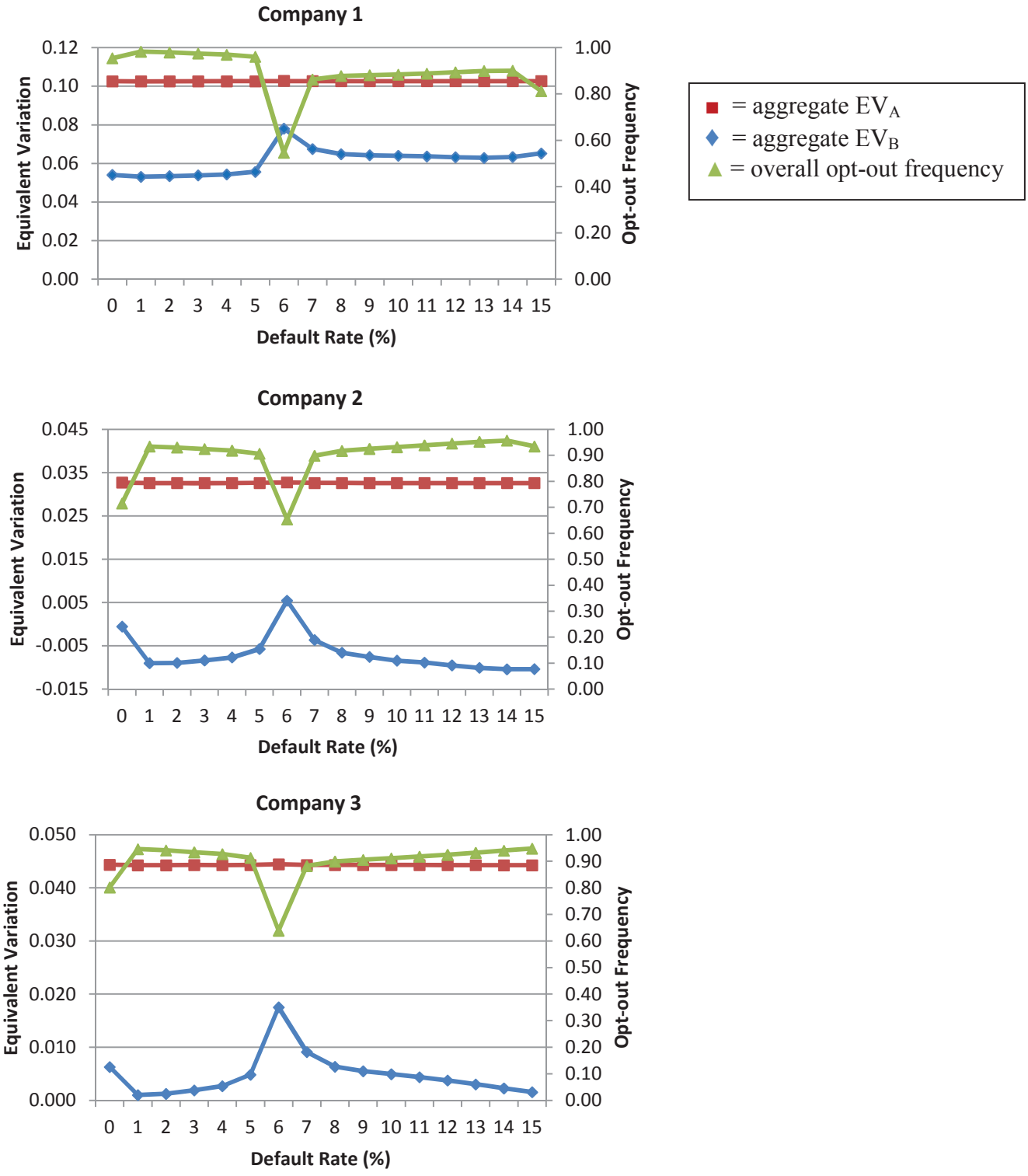

Figure 6: Average equivalent variation, decisions made in the alternative frame, with an employer match.

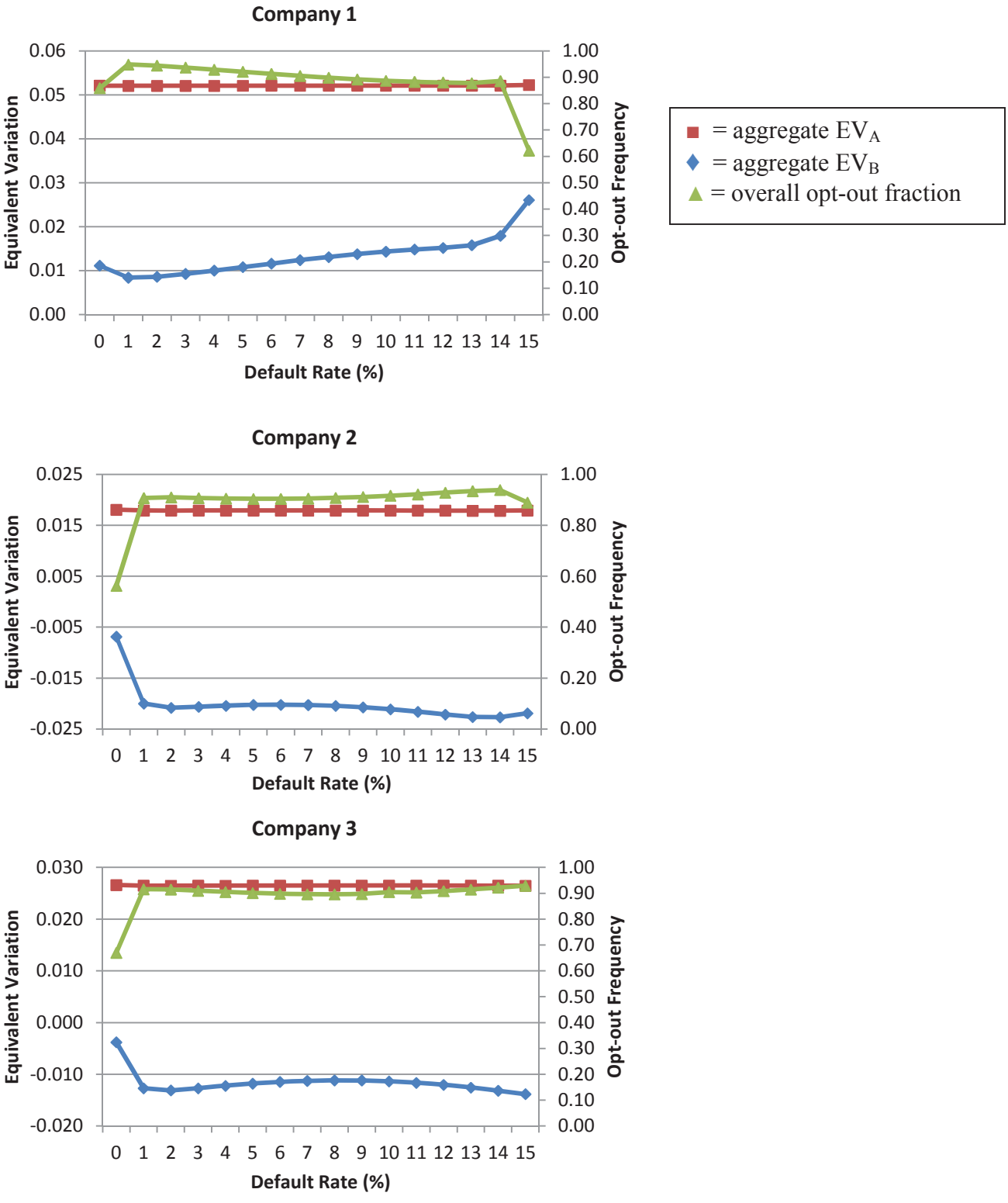

Figure 7: Average equivalent variation, decisions made in the alternative frame, without an employer match

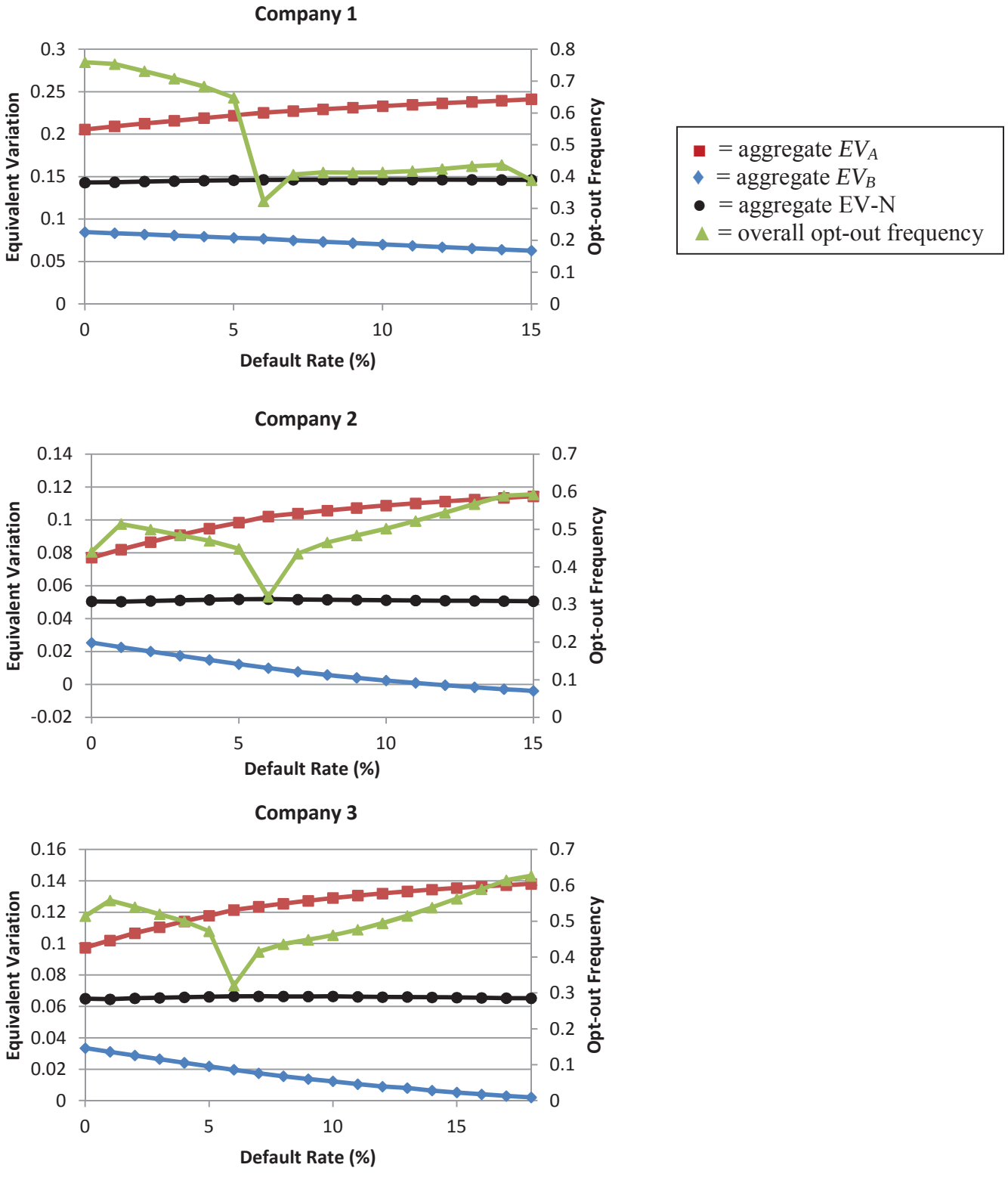

Figure 8: Average equivalent variation and opt-out frequency, with anchoring and an employer match

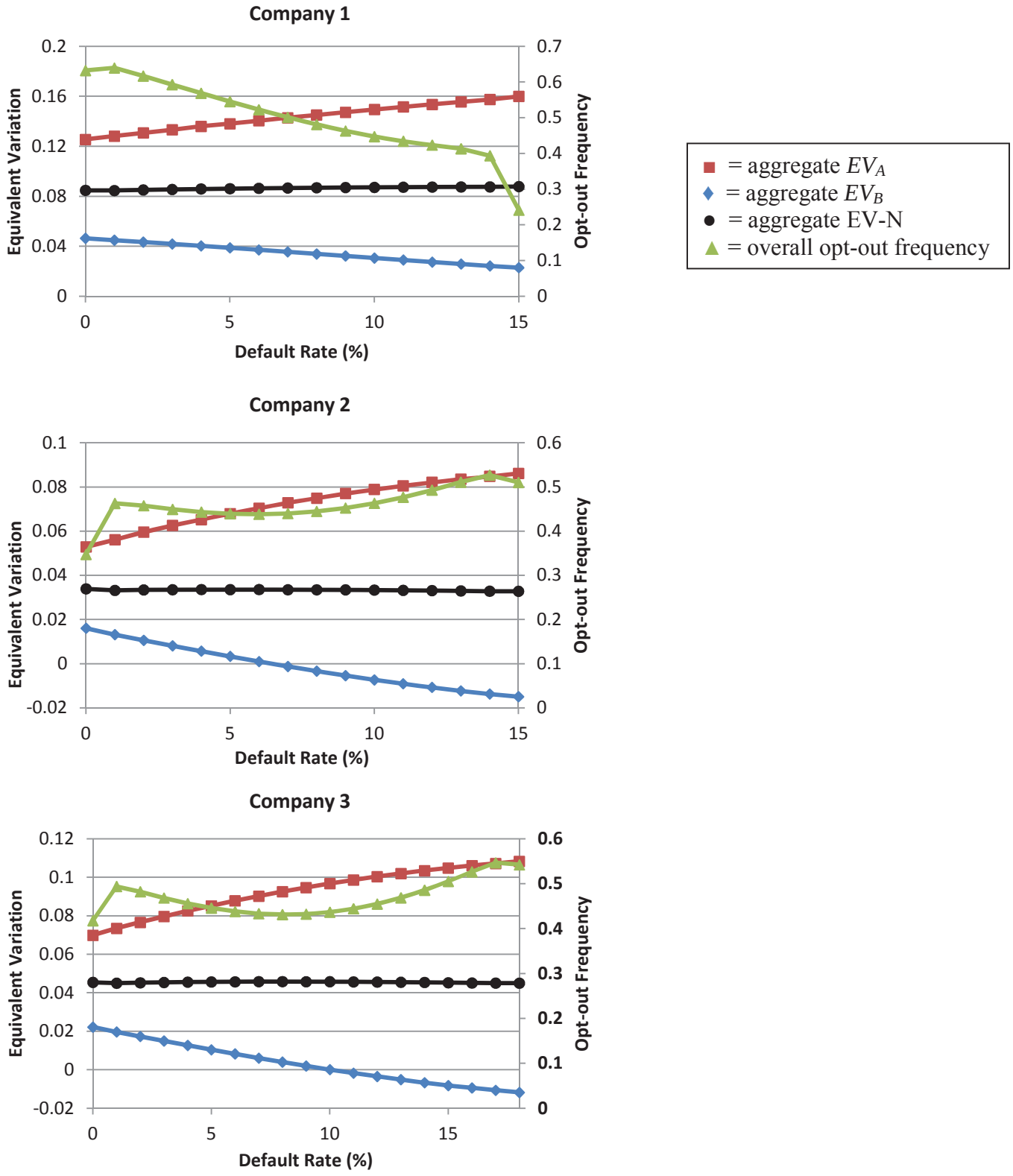

Figure 9: Average equivalent variation and opt-out frequency, with anchoring and no employer match