NBER WORKING PAPER SERIES

THE WELFARE ECONOMICS OF DEFAULT OPTIONS: A THEORETICAL AND EMPIRICAL ANALYSIS OF 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

We would like to thank seminar participants at the CESifo Venice Summer Institute Conference on Behavioural Welfare Economics and the ECORE Summer School, UCL, Louvain-la-Neuve, 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. We acknowledge financial support from the National Science Foundation through grant number SES-0752854. 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: A Theoretical and Empirical Analysis of 401(k) Plans

B. Douglas Bernheim, Andrey Fradkin, and Igor Popov

NBER Working Paper No. 17587
November 2011
JEL No. D03,D14,D60,D91,J26

ABSTRACT

According to previous research, changing the default contribution rate for a 401(k) pension plan has a powerful effect on the distribution of contributions among relatively new employees. Potential explanations include the following: (1) opting out may entail significant effort and inconvenience; (2) the default rate may serve as a psychological anchor, influencing choices because of its salience or imprimatur; (3) workers may procrastinate, putting off the opt-out decision; (4) workers may be inattentive. We examine the welfare implications of defaults under each of these theories. Because three of them involve non-standard behavioral hypotheses, we adopt and implement the framework for behavioral welfare economics proposed by Bernheim and Rangel (2009). In each case we begin by developing theoretical principles, and then confront the theory with data to reach concrete quantitative conclusions.

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

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

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

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, which receive far more attention. Potential explanations include: (1) opting out entails effort and inconvenience; (2) the default rate serves as a psychological anchor, influencing choices because of its salience or imprimatur;[2] (3) workers procrastinate, putting off the opt-out decision; (4) workers are inattentive.[3] We examine the welfare implications of defaults under choice patterns that can be rationalized by each of those theories. Because three of them involve non-standard behavioral hypotheses, we implement the framework for behavioral welfare economics proposed by Bernheim and Rangel (2009). In each case we develop theoretical principles and then analyze data to reach concrete quantitative conclusions.

Previous theoretical discussions of default effects provide conflicting policy recommendations. Invoking a principle of *ex post* validation, Thaler and Sunstein (2003) propose that companies should set defaults to minimize opt-out frequencies. Yet Carroll et al. (2009) argue that if people are sufficiently time inconsistent, it is optimal (under the "long run" criterion) to force active decisions by setting a highly undesirable default.

Our main theoretical contributions are as follows. First, we show that a generalized Pareto improvement criterion favors a default contribution rate of zero in all cases. Second, we demonstrate that, when default effects arise from small opt-out costs, the surplus-maximizing default rate coincides with either the minimum allowable contribution rate (usu-

[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.

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

[3] The existing evidence on attention is both limited and inconclusive. 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.

ally zero), the maximum allowable contribution rate, or the maximum matchable contribution rate. The considerations that favor those rates are also present (but may not be dominant) when opt-out costs are large. Third, we provide a recipe for creating arbitrarily large divergences between the opt-out-minimizing and surplus-maximizing default rates, thereby establishing that opt-out minimization can be highly undesirable. Fourth, we point out that the apparent desirability of an extreme default rate in settings with acute time inconsistency is artificial, because it presupposes an inability on the part of the employer to penalize inactive decisions or reward active ones. We also provide a new theorem that facilitates the application of the Bernheim-Rangel welfare framework to applied problems.

To conduct quantitative welfare analysis, we calibrate choice mappings consisistent with each of the four theories discussed above using data on distributions of 401(k) contribution rates for three companies. Those data illuminate behavior on a limited domain over which the four characteristic choice patterns are not sharply distinguished. Nevertheless, our results shed light on the reasonableness of each representation. Unrealistically large opt-out costs are required to rationalize observed default effects in the absence of non-standard behavioral explanations. The problem is not resolved by assuming sophsiticated time inconsistency unless one makes other unrealistic assumptions. Neither does naive time inconsistency plausibly explain the observed behavior. In contrast, once the model is generalized to include anchoring effects, the estimated opt-out cost distribution becomes reasonable. Inattentiveness can also account for observed default effects, but we know of no independent evidence that would allow us to evaluate the reasonableness of our parameterization.

We suspend disbelief concerning the reasonableness of any given representation and conduct welfare analysis assuming the full choice mapping (i.e., the mapping defined on its theoretical domain, and not merely on the observed domain) is consistent with a given model, examining the four models one at a time. Assuming the full choice mapping is consistent with a simple model of opt-out costs, we document a strong tendency for the worker-surplus-maximizing default rate to coincide with the maximum matchable contribu-

tion rate. Opt-out minimization is also typically achieved by setting the default equal to the maximum matchable contribution rate, and hence is often worker-optimal. Roughly 30% of the potential economic surplus flowing from a 401(k) plan is lost when the default rate is inefficiently set to zero. With matching provisions removed, the worker-surplus-maximizing default rates vary widely across companies, and diverge from the opt-out minimizing default rates. Deviations from the optima can still dissipate substantial portions of the surplus associated with a 401(k) plan, but opt-out minimization generally entails small welfare losses.

For choice mappings that are characteristic of the three remaining theories, we employ the Bernheim-Rangel framework, wherein conflicts between evaluations made in distinct welfare-relevant choice frames imply normative ambiguity. Assuming the full choice mapping is consistent with a model of anchoring, evaluations are highly frame-dependent; indeed, the worker-surplus-maximizing default rate ranges from the smallest to the largest feasible contribution rate, depending on the evaluation frame. Consequently, when all decision frames are deemed welfare-relevant, the degree of normative ambiguity is substantial. However, when welfare is evaluated in an anchorless (and hence arguably neutral) choice frame, worker surplus varies only slightly with the default rate, and hence the socially optimal default contribution rate (accounting for costs to employers and the government) is zero.

Assuming the full choice mapping is consistent with models of either time inconsistency or inattentiveness, the welfare implications of varying the default rate over the pertinent range are relatively insensitive to the choice frame used for evaluation. Accordingly, even if one treats all choices in all frames as welfare-relevant, the degree of normative ambiguity is surprisingly small. 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 differing degrees makes little difference. As in the basic model, we therefore see a strong tendency for the worker-surplus-maximizing default rate to coincide with the maximum matchable contribution rate. We also examine whether it is desirable (from the

perspective of a forward-looking decision frame) to force active decisions through extreme defaults when companies can also directly penalize inactive choice. Optimal penalties are either zero, in which case the optimal default problem is unchanged, or extremely large, in which case the default matters very little. Finally, we find that if all choice frames are deemed welfare-relevant, the degree of normative ambiguity associated with any given default rate is greater when opt-out choices are made in frames that are more conducive to contributing (e.g., precomittments in the case of time inconsistency). For our calibrated models, the difference is tiny with time inconsistency but large with inattentiveness.

The remainder of the paper is organized as follows. Section 2 sets forth the models of default effects. Section 3 reviews the Bernheim-Rangel framework and discusses its application to the problem at hand. Section 4 develops conceptual insights for each theory of default effects and provides structure for our quantitative analysis. Section 5 explains how we calibrate the models empirically, and Section 6 uses the calibrated models to investigate welfare. Section 7 provides some concluding remarks. Proofs appear in an online appendix.

2 Models of default effects

2.1 The basic model of opt-out costs

Consider an individual who has recently become eligible to participate in his employer's 401(k) plan. His *total contribution rate* $x \in [0, \overline{x}] \equiv X$ (contributions over earnings), reflects his *employee contribution rate*, r, as well as his employer's contributions. The plan has a default employee contribution rate at which the total contribution rate is $d \in [0, \overline{x}]$. We focus on the initial choice (in "period 0") between (a) accepting the default and (b) opting out at some cost e (reflecting inconvenience and effort) by selecting $x \in X \backslash d$.

As of period 0, the worker cares only about 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

[4]The elements of c are potentially indexed by both time and states of nature. All consumption, other

$u(e,\omega)+U(c,\theta)$,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 to opt out of the default. The period 0 choice of x matters because it determines his current 401(k) saving, his default for the next period,[5] 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 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. Where needed, we assume V is continuously differentiable, strictly quasiconcave in (x,z), and strictly increasing in both x and z, with $\lim_{z\to 0} V(x,z,\theta,\pi) = -\infty$ and $\lim_{z\to\infty} V(x,z,\theta,\pi) = +\infty$.

We allow the parameters $\xi \equiv (\gamma,\theta) \in [0,\overline{\gamma}] \times \Theta \equiv \Omega$ to differ across workers,[7] and use H to denote their CDF. 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.

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. We

than e, takes place after period 0.

[5] In principle, these two effects are separable (e.g., upon electing a contribution rate of 3%, the default for the next period could change to 4%), but in practice they always go hand-in-hand (in the 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] We can simplify our notation by treating γ rather than ω as the preference parameter governing opt-out costs because we take the opt-out technology as fixed.

assume τ is strictly increasing, continuous, convex, and piecewise linear, potentially with a finite number of convex kink-points (to allow for declining match rates and match caps); also, $\tau(0) = 0$ and $\tau(\overline{x}) < 1$. Fixing $e = 0$ and maximizing (1) subject to (2), ignoring the costs of opting out, yields an "ideal point" $x^*(\theta)$, which is unique and varies continuously with θ under our assumptions. 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,[8] and that the density is bounded at all other points. For any given worker, utility from the default is $V(d, 1 - \tau(d), \theta)$, while utility from opting out is $V(x^*(\theta), 1 - \tau(x^*(\theta)), \theta) - \gamma$. Hence, the worker opts out of d to $x^*(\theta)$ iff

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

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.[9] That observation allows us 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,[10] and to implement our framework empirically by estimating V rather than U, which we can accomplish with more limited data. 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.[11] With respect to (b), we are skeptical of the prospects for deriving helpful properties of sufficient generality; in any event, our theoretical analysis yields useful insights without additional structure, and our empirical analysis adds appropriate structure by fitting V to data.

[8] 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.

[9] This property hinges on the assumed absence of any relation between π and d, given x.

[10] 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 (e', c') for $c' \in C(x', z', \theta)$ from the observation that (e, x, z) is chosen over (e', x', z').

[11] We explicitly acknowledge a potential exception in Section 6.1.

For similar reasons, one can treat V as a fixed function when deriving comparative statics describing the responses of x and z to *temporary* changes in parameters governing the function τ (match rate, contribution cap, tax rate). However, *permanent* changes would alter π, and consequently modify both the continuation consumption correspondence C and the indirect utility function V.

2.2 Additional behavioral considerations

Next we present as-if models of anchoring, time consistency, and inattentiveness, and identify their characteristic choice patterns (specifically, framing effects).

Anchoring. With anchoring, the default rate d still impacts the opportunity set by changing the effort required to achieve any contribution rate, but it also establishes a psychological frame, $f = d$, that inclines workers toward choosing $x = f$. We could separate those roles through choice experiments in which the default rate and the effort schedule vary independently, e.g., 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 effort-cost effects empirically through additional identifying assumptions (see Section 5).

To incorporate anchoring, we assume the worker acts as if the reduced-form indirect utility function V depends on the default frame $f \in [0, \overline{x}]$.[12] Accordingly, he maximizes

$$W(e, x, z, \omega, \theta, \pi, f) = u(e, \omega) + V(x, z, \theta, \pi, f), \tag{4}$$

where V satisfies the same assumptions conditional on each frame f. We assume that the induced distribution of the worker's as-if ideal point, $x^*(\theta, f)$, has full support on $[0, \overline{x}]$, with atoms at 0, $\overline{x}$, and the kink points of τ (if any), but nowhere else, for all f. For some purposes, we also assume that an increase in f weakly shifts the individual's choices toward

[12] 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.

higher x (*monotonicity*).[13] 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 $f = f'$.[14] On the contrary, we intend (4) merely as an analytic device for recapitulating the dependence of a choice mapping on a decision frame f.

Time inconsistency. A worker can choose x either in a *forward-looking frame* $f = -1$ wherein he commits to a period 0 choice in advance, or a *contemporaneous frame* $f = 0$ wherein he makes that choice "in the moment." Choices made in either frame by itself are observationally equivalent to the basic model. If default effects are attributable to 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.

To incorporate time inconsistency, we assume the worker acts as if he maximizes $u(e, \omega) + \beta_f V(x, z, \theta, \pi)$, with $\beta_0 \in (0, 1)$ in the contemporaneous frame (a period 0 choice), and $\beta_{-1} = 1$ in the forward-looking frame (a "period -1" choice). To allow for the possibility that β_0 varies over the population, we modify our notation by writing $\xi = (\gamma, \theta, \beta_0)$. We define $x^*(\theta)$ as in the basic model. For frame -1, (3) governs the opt-out decision. For frame 0, the condition is

$$\Delta(\theta, d, \pi) \geq \frac{\gamma}{\beta_0}. \tag{5}$$

Inattentiveness. Whether an employee attends to the task of selecting a 401(k) contribution rate depends on the institutional environment created by his employer, which establishes the psychological 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

[13] 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 $f' > f$.

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

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

Because the institution fixes the frame f, we will sometimes suppress f in the notation, writing χ rather than $\chi(f)$. We allow for the possibility that χ varies over the population, and modify our previous notation by writing $\xi = (\gamma, \theta, \chi)$. We assume that the ranking of frames by χ is the same for all workers, and assign labels to frames so that χ is strictly increasing in f.[15] 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 $\overline{f}$. We define $x^*(\theta)$ as in the basic model.

Choices made in any frame by itself are observationally equivalent to the basic model. If default effects are attributable to inattentiveness, opt-out frequencies should be sensitive to interventions that manipulate attention. As mentioned previously, the evidence on that point is both limited and mixed.

On the use of reduced form as-if representations: To justify analyzing choices over (e, x, z) bundles while nevertheless treating the primitive choice objects as (e, c) bundles, we must make the following assumptions. (1) $C^*(x, z, \theta, \pi)$ does not depend on the initial period 0 frame, 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). (2) The worker chooses $(e, C^*(x, z, \theta, \pi, f))$ over $(e', C^*(x', z', \theta, \pi, f))$ in period 0 with full precommitment iff he chooses (e, x, z) over (e', x', z') without precommitment. One can interpret this assumption as requiring that the worker has sophisticated expectations with respect to his future behavior (so that precommitting to what he would choose anyway is irrelevant). As explained in Section 6, we do not see naive time inconsistency as a plausible explanation for 401(k) default effects. These assumptions imply that, for each individual, any given (e, x, z) bundle is associated with a single (e, c) bundle regardless of d or f, both in reality and in the worker's mind.

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

3 Welfare criteria

3.1 The general framework

Assuming workers employ choice mappings characteristic of anchoring effects, time inconsistency, or inattentiveness, we conduct welfare analysis using the framework for behavioral welfare analysis proposed by Bernheim and Rangel (2009). Here we briefly review the mechanics of that framework and offer a new result that facilitates practical applications. In the next subsection we discuss issues related to our application.

The Bernheim-Rangel (henceforth BR) framework generalizes the standard normative paradigm under the interpretation that welfare is defined directly in terms of choice, rather than underlying objectives, which 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. We review each step in turn.

Following BR, let $\mathbb{X}$ denote the set of all possible objects of choice. A *generalized choice situation* (abbreviated GCS), $G = (X, f)$, consists of a constraint set $X \subseteq \mathbb{X}$ and a psychological frame f.[16] 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, $\mathcal{G}^*$, with the property that $C(X, f) \subseteq X$. 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. BR argue that the set of potentially valid reasons for excluding a choice situation from $\mathcal{G}$ is limited. Philosophically, the libertarian principle of respect for choice (upon which both this framework and the standard paradigm are based) does not allow one to overrule or ignore a choice merely because the chosen option seems odd or one disagrees with the chooser about

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

its merits. Nevertheless, it is appropriate to exclude a choice for which the individual's characterization of the opportunity set does not match the analyst's characterization (i.e., cases of *characterization failure*).

To illustrate, suppose someone is presented with a choice between options x and y. He chooses x over y thinking incorrectly that y is z, even though he would choose y over x if he recognized y as y. Here, the choice of x over y is not a suitable guide for a policy maker who must choose between x and y on behalf of the individual. Rather, the policy maker should either construe the choice of x over y as a choice of x over z, or (if the identity of z is unclear) simply disregard that choice.

It is challenging to demonstrate that an individual's characterization of an opportunity set does not match that of the analyst. For example, someone who changes his behavior after being alerted to a potential error may simply be responding to social pressure. Matters are more promising if it can be shown that, in some particular type of frame, the individual generally misunderstands or fails to process information that is essential for identifying the opportunity set.

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 argue that a choice-based welfare criterion should possess five features. First, because welfare analysis is not only about identifying optima but also about comparing alternatives, the criterion should entail a binary relation that identifies improvements. Second, because we seek a choice-based criterion, the relation should depend only on the choice correspondence, defined on the welfare-relevant domain $\mathcal{G}$. Third, the relation should be coherent (i.e., at least acyclic). Fourth, it should respect unambiguous choice: if x is chosen over y in some GCS and y is never chosen over x, then x must be viewed as an improvement over y. Finally, the relation should never overrule a valid choice (one within the welfare-relevant

domain): if x is chosen from X in some GCS within $\mathcal{G}$, then the relation must treat x as unimprovable within X. To do otherwise would be to declare that choice a mistake, even though the choices remaining in $\mathcal{G}$ after step 1 are presumed to have been made with an accurate understanding of the alternatives. Overruling such choices would be contrary to the libertarian principles that the framework seeks to operationalize.

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. P^* satisfies the five requirements listed in the previous paragraph, and it 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).

BR's proposed criterion has several other attractive properties. First, given the breadth of the framework, it is universally applicable. Second, its continuity properties imply that, if one conducts welfare analysis based on a choice correspondence that is approximately correct, the normative conclusions that emerge will also be approximately correct. Thus, to justify the use of any given model in normative analysis, one need only argue that the implied choice correspondence is approximately correct. The reasonableness of the model as a depiction of decision processes is immaterial. Third, the framework is readily applied to specific economic theories and easily adapted to empirical analysis. Because all welfare statements are derived directly from the choice correspondence, *any* empirical representation of a choice correspondence enables welfare analysis. Our current study illustrates that principle. Finally, as we explain next, the 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 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 infinum) increment to income with p that 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 $\mathcal{G}$. 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 coventional 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.[17] The following result 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.*

The task of applying P^* to behavioral models is often simplified by virtue of the following property. Suppose $\mathcal{G} = \mathcal{X} \times \mathcal{F}$, where $\mathcal{X}$ is the set of opportunity sets and $\mathcal{F}$ is the set of welfare-relevant psychological frames, and where the restriction of C to any $f \in \mathcal{F}$ satsifies WARP. Then P^* is equivalent to the multiself Pareto criterion, treating each frame as a different self (BR, Theorem 3). That result does not apply with generality to the familiar quasi-hyperbolic model of time-inconsistency because one cannot write $\mathcal{G}$ as the requisite Cartesian product (decisions made at any given point in time cannot affect past consump-

[17] 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.

tion).[18] BR nevertheless derive an analytic representation of P^* (assuming sophisticated behavior) with no restriction on the welfare-relevant domain. They also examine a possible domain restriction, which we discuss later.

3.2 Context-specific criteria

As a practical matter, when applying the BR framework one must restrict attention to choice mappings that allow for some limited set of behavioral patterns, rather than all conceivable anomalies. In our analysis, we consider choice mappings that give rise to anomalies in the context of initial 401(k) enrollment decisions, but not otherwise.[19] We acknowledge that psychological framing may also influence choices after "period 0," and that consideration of such effects might lead to greater normative ambiguity when comparing consumption trajectories.[20] One could in principle consider a broader class of choice mappings that subsume the ambiguities arising from a wider variety of time-inconsistent behaviors, but implementation would then require much richer data than those currently available.

We have articulated each theory in Section 2.2 by explicitly defining psychological frames and providing a model of frame-dependent choice. Thus, once we define the welfare-relevant domain $\mathcal{G}$, application of the BR framework is straightforward. Significantly, we can proceed without taking any of the as-if theories literally as models of cognition; we are free to think of them merely as analytically convenient representations of choice correspondences. Even so, insights concerning cognition remain relevant because they can provide grounds for restricting $\mathcal{G}$. Here we consider potential theory-specific justifications for such restrictions.

Anchoring effects. A potential strategy for restricting $\mathcal{G}$ is to admit choices only if they

[18] Technically, for the reduced-form model of time inconsistency described in Section 2.2, $\mathcal{G}^*$ *can* be written as the requisite Cartesian product, and thus the result *does* apply, because nothing is consumed in period -1.

[19] Equivalently, one could say that we refine the welfare-relevant domain to exclude choice situations giving rise to anomalies other than those associated with initial 401(k) enrollment.

[20] For example, if a tendency to procrastinate when facing a 401(k) enrollment decision reflects a general (rather than context-specific) tendency to make present-biased choices, then future choices affecting the worker's consumption trajectory may also differ according to whether they are made in forward-looking or contemporaneous frames.

are made in an arguably neutral frame where choices are free from the influence of anchors. One possible candidate is a frame in which an active 401(k) election is a precondition of employment. Decisions in that frame are free from contribution-rate anchors, but their normative superiority to anchored choices is not obvious absent additional evidence. What evidence would suffice? If it is possible to show that the presence of an anchor causes the worker to ignore information he himself characterizes as pertinent (regardless of frame), and that no such distraction occurs in the neutral frame, then arguably the worker correctly characterizes his alternatives only in the neutral frame. Though we know of no such evidence, we nevertheless conduct some welfare exercises under the maintained hypothesis that a particular frame (defined below) is neutral.

Time inconsistency. For the $\beta\delta$ model, BR show (Theorem 11) that restricting $\mathcal{G}$ to choices that resolve all options at single points in time validates the "long-run" criterion (δ discounting, ignoring β). That restriction may be justified on the grounds that other types of decisions bring the potentially inconsistent objectives pursued at different points of time into conflict, producing outcomes of dubious normative significance. Here, the long-run criterion amounts to assessing welfare from the perspective of the forward-looking frame, $f = -1$.

Inattentiveness. To the extent we take the as-if representation literally, $\mathcal{G}$ should be limited to choices made in a frame f with $\chi(f) = 0$ so that the social planner does not emulate neglectful decision makers.[21] However, an empirically compelling justification for that restriction would require more than a showing that choices respond to interventions *intended* to manipulate attentiveness, as such interventions may also influence choices through other mechanisms, for example by browbeating or embarassing the decision maker. The domain restriction requires an explicit and credible demonstration that the workers who stick with the default in some cases simply neglect their decisions (e.g., forget, or give the matter no thought).

[21] 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.

4 Welfare analytics

We now analyze the welfare implications of setting 401(k) default rates, where a heterogeneous group of workers face a single default. Our focus here is on worker welfare, but Section 6 also provides empirical results on costs to employers (matching contributions) and the government (foregone tax revenue). When analyzing aggregate economic surplus, we abstract from distributional concerns by assuming that both earnings and the marginal social value of a dollar are the same for all workers. One can reinterpret our analysis as pertaining to cases in which preferences are homothetic, the cost of effort is proportional to earnings (e.g., because of the value of time), and the preference parameters (γ, θ) are distributed independently of earnings. Changing d then has the same effect within all income strata, and hence one can take any given stratum as representative. Because π plays no role in this analysis, we suppress it in our notation.

4.1 Welfare in the basic model

We begin with the Parteo criterion. Because $x^*(\theta)$ is assumed to have full support on $[0, \overline{x}]$, every feasible d is Pareto optimal. However, the following simple observation provides a possible justification for favoring $d = 0$:

Theorem 2: *Offering a 401(k) plan in the current period weakly Pareto improves upon not offering such a plan if and only if* $d = 0$.[22]

To assess worker surplus, we compute the equivalent variation (EV) 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).

[22] For the purpose of this result, we take the features of any future 401(k) offering as fixed.

For workers who do not opt out of the default, the EV is the value of $m^0(d,\theta)$ satisfying $V(0, 1 + m^0(d,\theta), \theta) = V(d, 1 - \tau(d), \theta)$. For those who do opt out, the EV is the value of $m^1(\theta,\gamma)$ satisfying $V(0, 1 + m^1(\theta,\gamma), \theta) = V(x^*(\theta), 1 - \tau(x^*(\theta)), \theta) - \gamma$.[23] Because the worker elects the default iff $m^0(d,\theta) \geq m^1(\theta,\gamma)$, aggregate worker surplus is:

$$\int_\Omega m^1(\xi)dH(\xi) + \int_{D(d)} \left[m^0(d,\theta) - m^1(\xi)\right] dH(\xi),$$

where $D(d)$ denotes the values of ξ for which the worker does not opt out. Only the second term, which measures the incremental benefit received by workers who elect the default, varies with d. Thus the worker-surplus maximization problem is:

$$\max_d \int_{D(d)} \left[m^0(d,\theta) - m^1(\xi)\right] dH(\xi) \tag{7}$$

If m^0 is differentiable, the first order condition is simply $\int_{D(d)} m^0_d(d,\theta)dH(\xi) = 0$.

In contrast, opt-out minimization requires us to solve $\max_d \int_{D(d)} dH(\xi)$, which (7) resembles, except that the density $dH(\xi)$ is weighted by $m^0(d,\theta) - m^1(\xi)$. That difference can give rise to large discrepancies. To see why, assume the cost of opt-out is zero for some otherwise representative portion of workers, $\kappa \in (0,1)$.[24] Also assume there are no atoms in the distribution of $x^*(\theta)$ on the interior of $[0,\overline{x}]$ (which could appear if τ were kinked).

Theorem 3: *The worker-surplus-maximizing default rate is independent of κ, the fraction of workers for whom the cost of opt-out is zero. In contrast, if that fraction is sufficiently large, the opt-out minimizing default rate is either $d = 0$ or $d = \overline{x}$.*

Theorem 3 implies that the surplus-maximizing and opt-out minimizing default rates may differ substantially. That divergence occurs because $m^0(d,\theta) - m^1(\xi)$ is much smaller on average for those who stick with the default when it equals 0 or $\overline{x}$ (because a positive fraction of those individuals have zero switching costs) than for those who stick with any other default (because a negligible fraction of those individuals have zero switching costs). As we will see in the Section 6.1, Theorem 3 is empirically relevant.

[23] Given our assumptions on V, solutions to both equations exist and are unique.

[24] The same conclusion would follow if the costs for that group were simply very small.

Our next result shows that 0, $\overline{x}$, and the kink points of τ (e.g., a match cap) are particularly attractive from the perspective of surplus maximization. Define $\mathcal{A} \subset [0, \overline{x}]$ to contain those values, and 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 $[0, \overline{\gamma}]$.

Theorem 4: *Consider a sequence of CDFs H^γ_k with $\overline{\gamma}_k \to 0$ and mean γ_k such that $\gamma_k/\overline{\gamma}_k > e^*$ for all k and some $e^* > 0$.*[25] *The surplus-maximizing default rates, d^*_k, converge to a point in $\mathcal{A}$.*

This result is notable given the historical prevalence of non-enrollment defaults ($d = 0$), but it also potentially argues for setting defaults equal to match caps. It exploits the fact that, as the distribution of costs converges toward zero, both the maximand and the measure of $D(d)$ converge to zero, except at points in $\mathcal{A}$, where the measure of $D(d)$ remains bounded away from zero. The larger the atom at a given point, the more attractive it becomes. As we will see in the Section 6, elements of $\mathcal{A}$ frequently emerge as surplus-maximizing defaults for the reasons highlighted by this theorem (even when opt-out costs are relatively large).

4.2 Welfare with additional behavioral considerations

For the three behavioral theories, generalized Pareto optimality is not a discerning criterion.[26] However, Theorem 2 generalizes: the (weak generalized) Pareto improvement criterion implies that the plan should have a default of zero and, in the cases of as-if time inconsistency and inattentiveness, that the frame in which workers make the opt-out decision (f_D) should be 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.

[25] 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.

[26] Plainly, if there is some frame f and worker for whom $x^*(\omega, f) = d$, then d is a generalized Pareto optimal default rate.

Theorem 5: *Regardless of whether the welfare-relevant domain is unrestricted or restricted to any subset of frames, offering a 401(k) plan in the current period creates a weak generalized Pareto improvement over not offering a plan in the current period if and only if* $d = 0$ *and, for the cases of time inconsistency and inattentiveness,* $f_D \geq f_M$*. Moreover, for the cases of as-if time inconsistency and inattentiveness,* $(d, f) = (0, f_M)$ *creates a weak generalized Pareto improvement over* $(d, f) = (0, f)$ *for any* $f > f_M$*.*

Next we discuss aggregate surplus, taking the three theories one at a time.

Anchoring. For the special case where workers act as if default effects arise solely from anchoring rather than opt-out costs, we obtain a negative result: every default rate maximizes worker surplus (and hence is worker-optimal) from the perspective of some frame. Consequently, we cannot say that any default is unambiguously better than another unless we restrict the welfare-relevant domain.

Theorem 6: *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* $d = 0$*.*

Despite its negative message, Theorem 6 does not imply that welfare analysis is uninformative in this special case. The range $[EV_A, EV_B]$ still bounds the degree of ambiguity concerning the worker benefits, which is of interest, e.g., if the object is to compare the benefits of a 401(k) plan with its social costs. Also, as-if opt-out costs may be non-trivial, and there may be sound justifications for domain restrictions.

Next we derive expressions for EV_A and EV_B allowing for positive as-if opt-out costs. For those who do not opt out of the default, the EV evaluated from the perspective of frame f is the value of $m_A^0(d, \theta, f)$ (where the subscript A indicates anchoring) satisfying $V(0, 1 + m_A^0(d, \theta, f), \theta, f) = V(d, 1 - \tau(d), \theta, f)$. For those who opt out, the EV evaluated from the perspective of frame f is the value of $m_A^1(d, \xi, f)$ satisfying $V(0, 1 + m_A^1(d, \xi, f), \theta, f) =$

$V(x^*(\theta,d), 1-\tau\left(x^*(\theta,d)\right),\theta,f)-\gamma$. In both cases, m_A^0 and m_A^1 are monotonic in f (given our monotonicity assumption), so we obtain EV_A and EV_B by evaluating EV in frames $f=\overline{x}$ and $f=0$, respectively.

The aggregate EV given a default rate of d from the perspective of default frame f is:

$$\int_{\Omega} m_A^1(d,\xi,f)dH(\xi)+\int_{D_A(d)}\left[m_A^0(d,\theta,f)-m_A^1(d,\xi,f)\right]dH(\xi)\ ,$$

where $D_A(d)$ is the subset of ξ for which the worker does not opt out ($\xi\in D_A(d)$ iff $m_A^0(d,\theta,d)\geq m_A^1(d,\xi,d)$). If opt-out were costless, only the first term would remain. By Theorem 6, that term is maximized at $d=f$, the frame used for evaluation. Positive opt-out costs introduce the second term, which resembles the maximand in (7), except that the limits of integration depend on decisions made within the frame of the prevailing default rate, rather than the frame used for evaluation. The integrand is therefore generally non-zero at those limits – positive at the opt-*up* boundary and negative at the opt-*down* boundary for $f=\overline{x}$, and conversely for $f=0$. Because an increase in d reduces the opt-up set and expands the opt-down set, the welfare effect of increasing d flowing through the limits of integration are therefore positive for EV_A and negative for EV_B, which reinforces effects flowing through the first term. In contrast, the derivative of the second term's integrand can be positive or negative regardless of the frame, and may favor intermediate default rates, just as in the basic model.

Time inconsistency: For those who do not opt out of the default, the EV evaluated from the perspective of either frame is the value of $m_T^0(d,\theta)$ satisfying $V(0,1+m_T^0(d,\theta),\theta)=V(d,1-\tau(d),\theta)$. For those who opt out, the EV evaluated from the perspective of frame f is the value of $m_T^1(\xi,f)$ satisfying $V(0,1+m_T^1(\xi,f),\theta)=V(x^*(\theta),\left(1-\tau(x^*(\theta))\right),\theta)-\frac{\gamma}{\beta_f}$. When the opt-out choice is made in frame h, the aggregate EV given a default rate d from the perspective of default frame f is:

$$\int_{\Omega} m_T^1(\xi,f)dH(\xi)+\int_{D_T(d,h)}\left[m_T^0(d,\theta)-m_T^1(\xi,f)\right]dH(\xi)\ , \tag{8}$$

where $D_T(d,h)$ is the subset of ξ for which the worker does not opt out in frame h ($\xi \in D_T(d,h)$ iff $m_T^0(d,\theta) \geq m_T^1(\xi,h)$). When the choice frame and the evaluation frame coincide ($h = f$), our analysis is the same as for the basic model (except that γ is inflated by the factor $\frac{1}{\beta_0}$ for $h = f = 0$). When those frames differ, the choice frame h governs the opt-out decision and hence the limits of integration for the second term, whereas the EVs for each individual and hence the integrands reflect evaluation frame f. For the special case where $h = 0$ and $f = -1$, our analysis resembles that of Della Vigna and Malmendier (2004). To evaluate worker welfare without restricting $\mathcal{G}$, we compute EV_A in frame -1 and EV_B in frame 0 (because $m_T^1(\xi,-1) > m_T^1(\xi,0)$).

Here we can also treat the choice frame as a policy instrument. For example, a company could offer workers opportunities to commit to making 401(k) elections by self-imposed deadlines, with penalties for non-performance. Even without a restriction on $\mathcal{G}$, one choice frame may be preferrable to another (e.g., if choices in frame f are less damaging from the perspective of frame f' than are choices in frame f' from the perspective of frame f).

With $\mathcal{G}$ restricted to choices made in the forward-looking frame, the best policy allows for opt-out precommitment and sets the default rate exactly as in the basic model. The absence of such arrangements may raise doubts about the applicability of the $\beta\delta$ model (with sophistication), the welfare-relevance of the forward-looking frame, or both. Possibly state-invariant precommitments entail high costs due to uncertainty concerning the future opportunity cost of time, in which case it is appropriate to focus on policies that offer opt-out choices only in the contemporaneous frame ($h = 0$).

Consider then the problem of maximizing (8) over d, fixing $h = 0$ and evaluating welfare in the forward-looking frame ($f = -1$). Suppose for the moment that opt-out costs are universally small, but that many workers stick with the default because $\beta_0 \ll 1$. Then an extreme default rate achieves near-universal opt-out, which is nearly first-best. In constrast, with a moderate default rate, some workers adhere to the default despite substantial efficiency losses. Accordingly, as Carrol et. al. (2009) observed, the optimum is to compel active

decisions by setting extreme defaults.

We view the aforementioned policy prescription as artificial and unattractive for two reasons. First, those who fail to opt-out either due to extreme present bias or (more likely) inattentiveness must actually endure highly inefficient consequences. Second, the default rate is a blunt and potentially inappropriate instrument for encouraging opt-out. Simply penalizing workers for failing to make active decisions accomplishes the same end without distorting the default option.[27] Note that we can incorporate non-monentary penalties into our analysis (e.g., pressure, disapproval, etc.) by interpreting γ as a differential utility cost and treating it as a policy instrument.

Inattentiveness: The EV formulas for our model of as-if inattentiveness are the same as those for time inconsistency, except $\chi(f) + \gamma$ replaces $\frac{\gamma}{\beta_f}$. If we remain agnostic about the cognitive interpretation of χ and treat choices frames as welfare-relevant regardless of whether they are thought to encourage attentiveness, then we assess EV_A in the fully attentive frame and EV_B in the least attentive frame. To evaluate welfare from the perspective of a fully attentive frame (f_a such that $\chi(f_a) = 0$), we treat the welfare cost of opt-out as γ, even though the worker acts as if it is $\chi(f_p) + \gamma$ (where f_p is the prevailing frame). Because we have no basis for assessing the largest value of $\chi(f)$ achieved in any pertinent frame, we use the prevailing frame in place of the least attentive frame. Welfare analysis from the perspective of f_p treats $\chi(f)$ as a cost, but is otherwise the same as for the basic model.

Just as with time-inconsistency, one can treat the choice frame as a policy instrument. With $\mathcal{G}$ restricted to fully attentive choices, the ideal policy is to implement f_a and set the default exactly as in the basic model. We will assume that such measures either are not cost-effective, or are so invasive that they entail significant psychological costs.

[27] An even more efficient alternative would be to subsidize those who make active decisions and tax those who do not, subject to a balanced-budget constraint.

5 Empirical calibration

5.1 Parameterization

The basic model. We assume the indirect utility function has the following form:

$$V(x, z, \alpha, \rho) = \rho \ln(x + \alpha) + \ln(z). \tag{9}$$

Thus, the vector θ consists of the pair (α, ρ). This specification has the attractive implication that the monetary value of the effort required to opt out of the default does not vary with the size of the worker's desired 401(k) contribution. With $\alpha = 0$, V is a Cobb-Douglas function in x and z, expenditure shares are fixed, the employee's contribution rate, r, is unresponsive to a temporary change in the uncapped employer match, and the marginal utility of 401(k) contributions is (implausibly) zero for workers who prefer to contribute nothing. In contrast, with $\alpha > 0$, a temporary increase in an uncapped employer match rate increases the optimal employee contribution rate, and the marginal utility of contributions can be positive even when the worker's ideal is to contribute nothing.[28]

Because α governs the sensitivity of the employee contribution rate to the concurrent employer matching rate and tax rate, it can be identified from any data that reveals that sensitivity. Match rates do not vary within firm in our data; however, we can identify the relevant elasticity from the degree of bunching at kink points in the opportunity set (as in Saez, 2009), which occur at the maximum matchable contribution rates. Larger values of α imply larger elasticities, and hence more pronounced bunching.

We assume $z = 1 - \frac{tx}{1+m}$ for $x \leq x_M$ and $z = \overline{Z} - tx$ for $x \geq x_M$, where $\overline{Z} = 1 + tx_M\left(1 - \frac{1}{1+m}\right)$, x_M is the total contribution rate when the worker reaches the cap on matchable contributions,[29] t is the tax adjustment parameter, and m is the matching rate. One can interpret $\overline{Z} - 1$ as the "virtual income" implicit in the kinked budget constraint

[28] 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.

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

when $x \geq x_M$. Throughout, we assume $t = 0.8$ because most workers fell into the 15% or 25% marginal tax brackets during the relevant time period.

We assume $\rho = \max\{\widetilde{\rho}, 0\}$, where the CDF for the random variable $\widetilde{\rho}$, denoted F, is normal with mean μ and variance σ^2. We assume the CDF for γ, denoted Φ, is a mixture between an exponential distribution and a probability atom at zero:

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

The parameter λ_1 represents the fraction of workers who act as if opt-out costs are negligible. We take the distributions of $\widetilde{\rho}$ and λ to be independent. We treat α as a fixed parameter, common to all workers.

Because some groups of employees may be more motivated savers than others, we allow μ to differ across firms, writing μ_i for firm i. We use ψ to denote the values of the other underlying parameters (α, σ, λ_1, and λ_2), which we take to be the same across all firms.

Adding anchoring. We assume anchoring effects draw the distribution of preferences toward the anchor from both directions. Formally, for any given values of α and d, let ρ^* denote the value of ρ for which $x^*(\alpha, \rho^*) = d$.[30] 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 is then the 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.

Both anchoring effects and switching costs 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 have formulated it) tends to

[30] 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.

shift each half of the distribution of $\widetilde{\rho}$ toward the default, thereby creating a spike without a neighboring trough. Thus, given our maintained hypotheses, we can separately identify ζ and (λ_1, λ_2), and thereby evaluate the relative importance of the two explanations.

Adding time inconsistency and inattentiveness. We cannot identify the degree of as-if time inconsistency or inattentiveness from the available data. However, interpreted through the lens of the model with time inconsistency, estimates of the basic model yield the distribution of $\frac{\gamma}{\beta_0}$. We infer the distribution of γ under the identifying assumption that $\beta_0 = 0.8$ (which is roughly in line with the pertinent literature). Similarly, interpreted through the lens of the model with inattentiveness, estimates of the basic model yield the distribution of $\gamma + \chi(f_p)$. Lacking any independent evidence concerning the magnitude of $\chi(f_p)$, we fix it at a value that implies a reasonable distribution of γ (as detailed below).

5.2 Calibration method

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 picks 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 $x^i_k = 0.01\left[(k-1) + m^i \min\{k-1, 100 r^i_M\}\right]$, and $K = 100\overline{r} + 1$.[31]

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 will 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

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

[31] 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 $X^i = \{0, 0.015, ..., 0.075, 0.09, 0.1, ..., 0.17, 0.18\}$.

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

$$\Pr{}_i(x_k^i \mid \psi, \mu_i, d) = \int_{B_k^i(\alpha)} \Phi\left(\Gamma_k^i(\alpha, \max\{0, \widetilde{\rho}\}, d)\right) dF(\widetilde{\rho}) \tag{10}$$

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.[32] 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 (10) with

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

5.3 Data

Our data consist of distributions of 401(k) contribution rates for workers at three companies, which compose a subset of those studied in Beshears et al. (2008), Choi et al. (2006), Madrian and Shea (2001), and several other papers by combinations of those authors.[33] We focused on three particular companies for two reasons. First, all of them switched between regimes with strictly positive default rates. We insisted on this feature because we were

[32] 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.

[33] The data 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.

concerned that switching from a default of zero to a positive rate (which requires a change to automatic enrollment) might be qualitatively different than switching the default between two positive rates. In practice, our model performed equally well in fitting distributions for zero and strictly positive default rates.[34] Second, with few exceptions, features of the 401(k) plans other than default rates remained stable across default regimes for all three firms.

For each company and default regime, the data indicate the fraction of recently eligible employees who elected each allowable contribution rate.[35] 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.

5.4 Estimates

Estimates of the basic model appear in Table 2. All parameters are estimated precisely. The value of α is positive, as it must be for the model to generate spikes in the distributions of contributions at the maximum matchable contribution rates. The mean utility weight for each company accords with average contributions, and 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 $V(0, 1 - v(\gamma), \theta) = V(0, 1, \theta) - \gamma$. For specification (9), $v(0.0847) = 0.0812$ 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 \$50,000, the monetary equivalent of γ is more

[34] Company 3 also operated under a regime with a 0% default rate. We discarded that data because, when the company implemented automatic enrollment, it applied that policy retroactively to workers hired under the 0% default regime.

[35] 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 more than a year.

than $4,000 at the mean of the distribution and more than $2,800 at the median. Yet it is difficult to believe that more than a handful of employees would actually turn down a payment of several hundred dollars, let alone several thousand, to avoid making an active 401(k) election.

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.

Sophisticated time inconsistency does not resolve the puzzle. 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 provide an adequate explanation: if the deadlines for changing 401(k) elections are frequent (e.g., biweekly), workers would presumably learn from numerous failures to follow through on intentions over the course of more than a year; if they are infrequent (e.g., quarterly), the original puzzle remains. Given the paucity of evidence on attention, it is more difficult to evaluate the plausibility of the assumption that workers are often inattentive to opportunities worth a few thousand dollars.

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 $50,000 per year, the monetary equivalent of γ is therefore $67 at the mean and $46 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. Our analysis therefore suggests that bunching at the default option is primarily attributable to anchoring rather than to opt-out costs.

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.

6 Welfare implications

6.1 The basic model

Using the our estimates of the basic model, we simulate workers' choices for various default rates and conduct welfare analysis (suspending any disbelief concerning as-if opt-out cost magnitudes). Figure 2 graphs average EV, as well as two opt-out frequencies: "overall opt-out frequency," the ratio of all opt-outs to all workers, and "zero-cost opt-out frequency," the ratio of opt-outs among those with zero opt-out costs to all workers. The average EV is maximized and the opt-out frequencies minimized for a default rate equal to the maximum matchable contribution rate (6%) at all three companies. Indeed, the EV-maximizing default rate remains 6% for all three companies even when we vary the opt-out cost distribution parameter over a wide range. This finding reflects the importance of the forces that give rise to Theorem 4.

With a default of 0%, the average EV is 7.07% of earnings for company 1, 1.97% for company 2, and 2.75% for company 3. The figure is higher for company 1 because μ_1 substantially exceeds μ_2 and μ_3, and because company 1 has a more generous matching rate. By way of comparison, the average simulated employee and employer contributions are, respectively, 6.90% and 4.35% for company 1, 3.45% and 1.22% for company 2, and 4.60% and 1.46% for company 3. Thus, in each case, the average EV roughly equals the

employer contribution plus 20% to 40% of the employee contribution, which seems plausible. With a default rate of 6%, the average EV rises to 9.86% for company 1 (an increase of 2.89 percentage points), 2.71% for company 2 (an increase of 0.74 percentage points), and 3.93% for company 3 (an increase of 1.18 percentage point). 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. Those magnitudes, though very large, are not surprising in light of the opt-out costs required to rationalize observed default effects.

Because the identity of the EV-maximizing default rate in Figure 2 is driven by matching provisions, we next examine optimal defaults without a match. As mentioned in Section 2.1, we can use our model to simulate outcomes with the employer match for the current period removed. We note, however, 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.

Figure 3 resembles Figure 2, except that we have removed the effects of employer matching provisions. Naturally, the EVs fall dramatically. The EV-maximizing default now differs considerably across the companies, reflecting the differences in μ_i: it is 13% for company 1 (with average EV equal to 4.60% of earnings), 2% for company 2 (with average EV equal to 1.08% of earnings), and 6% for company 3 (with average EV equal to 1.86% of earnings). The efficiency losses associated with setting a default of zero are also smaller: 1.3% of earnings at company 1, 0.03% at company 2, and 0.26% at company 3. Opt-out minimization now leads to the wrong default rate for all three companies – 15% for company 1, and 0% for companies 2 and 3 – illustrating the relevance of Theorem 3. However, the welfare losses

from opt-out minimization are small for companies 1 and 2 (0.04% and 0.03% of earnings, respectively) and modest at company 3 (0.26%).

Figure 2 raises the possibility that setting the default equal to the match cap may be a good rule of thumb for companies offering matching contributions. Figure 4 explores the robustness of that finding. We simulate behavior for a range of match caps, and plot the EV-maximizing default rate against the match cap. While the two rates coincide for intermediate values, they differ for low and high match caps, in some cases dramatically.

Maximization of average EV is only one possible objective. A policy maker who simply wished to stimulate saving might choose the default rate that maximizes average employee contributions. Contributions need not (and in some cases do not) increase monotonically with the default rate. Nevertheless, in our base-case simulations, they are maximized when the default equals the contribution cap; the resulting contribution rates are 11.2% for company 1, 8.3% for company 2, and 10.1% for company 3 (which has a higher cap). A similar result holds with matching provisions removed.

The benefits depicted in Figures 2 and 3 come at a cost. We graph the default rate versus costs to the employer (matching contributions) and to the government (tax revenues) in Figures 5(a) (the base case) and 5(b) (without matching provisions). We cannot perform a full social cost-benefit analysis because these costs pertain only to the current period.

6.2 Anchoring

Using the estimated model with as-if achoring effects, we simulate workers' choices for various default rates. Figure 6 graphs average EV_A, EV_B, EV assessed from the perspective of an arguably neutral frame, and the overall opt-out frequency against the default rate. We assume the worker would act as if $\zeta = 0$ in the neutral frame. That assumption appears reasonable under a literal interpretation of the as-if utility function, but Bernheim (2009) warns against such literalism. We acknowledge that more extensive data on framing effects would be required to justify properly the selection of a neutral frame.

Because our estimates imply low opt-out costs, Theorem 6 applies as an approximation.

Thus, EV_A is maximized at the highest allowable contribution rate, and EV_B is maximized at zero. Those peaks are unrelated to opt-out minimization or match caps. The substantial gaps between EV_A and EV_B are direct reflections of the large anchoring effects which create wide regions of ambiguity. The gap is smallest for $d = 0$, but are large even in that case: 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.

Figure 6 appears to suggest that a policy maker who wishes to "play it safe" should set $d = 0$ (to maximize EV_B), in which case the average benefit employees receive from 401(k) eligibility is unambiguously no less than 8.5% of earnings. However, that finding depends on our choice of the initial regime used to evaluate EV. If the initial regime required a 401(k) contribution equal to the overall cap, 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. We therefore turn to EV evaluated from the perspective of the putative neutral frame (henceforth EV-N). Strikingly, all the EV-N curves in Figure 6 are virtually flat. Regardless of the default rate, EV-N is roughly 14.5% of earnings for company 1, 5% for company 2, and 6.5% for company 3. 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. Because costs to the employer and the government rise with the default rate,[36] $d = 0$ emerges as socially optimal in the neutral frame: it saves on costs while achieving nearly all the employee benefits.

Figure 7 presents a second set of simulations with the companies' matching provisions removed. We observe similar patterns for EV_A and EV_B, although both measures of surplus are much lower without the match. The default rate continues to make relatively little difference in the neutral frame, as the EV-N curves are once again nearly flat. Though

[36] Graphs of employer and government costs versus the default rate appear in the online appendix; they are very similar to those shown in Figure 5, which pertains to the model without anchoring effects.

the stakes are small, EV-N is maximized at the contribution limit for company 1, 0% for company 2, and 8% for company 3.

6.3 Time-inconsistency and inattentiveness

Finally, we examine welfare with as-if time inconsistency and inattentiveness. To start, we assume that opt-out decisions are made in the *prevailing frames* (f_0 in the case of time inconsistency, f_p in the case of inattentiveness). The simulated distributions of workers' choices are then the same as in Section 6.1. Figure 8 graphs average EV_A and EV_B against the default rate for each of the three companies. For EV_A, any incurred opt-out costs are simply discounted by the assumed value of β. We calibrate our models using $\beta = 0.8$ in the case of time inconsistency and $\beta = 0.015$ in the case of inattentiveness. We select the latter value because it generates a reasonable distribution of opt-out costs (the mean is \$63 and the median \$44 for a worker earning \$50,000 per year). In both cases, EV_B is the same as for the basic model; we calculate it by applying no discount to measured opt-out costs, which amounts to using $\beta = 1$.

Surprisingly, varying the value of β used to evaluate welfare from 0.015 to unity has no impact on the EV-maximizing default rate (the maximum matchable contribution rate of 6% in all cases), and only a small effect on the level of EV for any given default. Thus, when one interprets the data through the lens of as-if time inconsistency or inattentiveness, the scope of ambiguity concerning welfare, $[EV_B, EV_A]$, is rather small, despite the presence of large default effects. The reason is simple: 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 opt-out costs are enormous from the perspective of the prevailing frame, discounting them heavily makes little difference.

Figure 9 presents a second set of simulations with the companies' matching provisions removed. Here, the model and frame of evaluation matter a bit more than in Figure 8, but not dramatically. Even with inattentiveness (for which we discount measured opt-out costs to a much larger degree), the differences between EV_A and EV_B remain modest, and both

measures of consumer surplus are maximized at similar values (14% versus 13% for company 1, 0% versus 2% for company 2, and 6% in both cases for company 3).

Surprisingly, we therefore find that the degree to which time inconsistency or inattention inflates opt-out costs is not terribly consequential either for the optimal default rate or for the magnitude of the economic benefits workers derive from 401(k) plans, assuming workers make the opt-out choice in the prevailing frame. What then of the finding in Carroll et al. (2009) that an extreme default is optimal from the forward-looking perspective when β is sufficiently low? The result still holds, but only for very small β (e.g., 0.015, not 0.8), and for defaults substantially outside the range considered, where the evaluation frame matters to a much greater degree. For each company, the EV curve with $\beta = 0.015$ 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).

As an alternative to an extreme default rate, employers could penalize workers for failing to make active 401(k) elections. We therefore perform simulations in which we optimize over penalties and default rates simultaneously. From the perspective of the forward-looking frame ($\beta = 0.8$), the optimal penalty is always zero and the optimal default problem is as before. From the perspective of the attentive frame ($\beta = 0.015$), the optimal penalty is enormous (roughly 40% of earnings) and the default is of practically 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, there is a sharp transition (e.g., at around $\eta = 0.007$ for company 1 with a match) to a regime in which the optimal penalty is zero and the optimal default problem is as before. Figure 10 shows why we never obtain a small optimal penalty. 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 zero to 1%. 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

virtually irrelevant.

So far we have focused on the selection of a default rate assuming decisions are made in the prevailing frame. As mentioned in Section 4.2, we can also treat the decision frame as a policy variable. Figures 11 (with matching provisions) and 12 (without matching provisions) display simulated opt-out frequencies for decisions made not only in the prevailing institutional frame (as above), but also in the forward-looking frame for the case of time inconsistency (assuming $\beta = 0.8$) and the fully attentive frame for the case of inattentiveness (assuming $\beta = 0.015$). The decision frame plainly has a large effect on aggregate behavior within the calibrated inattentiveness model, but a much smaller effect within the calibrated time inconsistency model due to the disperseness of the estimated opt-out cost distribution.

Figure 13 graphs average EV_A and EV_B against the default rate when the opt-out decision is made in the forward-looking and fully attentive frames (for the cases of time inconsistency and inattentiveness respectively). For our calibrated model of time inconsistency, the curves differ only slightly from their counterparts in Figure 8, which pertain to decisions made in the contemporaneous frame; there is a tiny increase in normative ambiguity. The surprising finding that the decision frame has practically no effect on welfare is explained by its small impact on opt-out frequencies (Figure 11). In contrast, for our calibrated model of inattentiveness, the curves in Figure 13 differ dramatically from their counterparts in Figure 8. Switching choices from the partially attentive to fully attentive frames substantially increases EV_A and decreases EV_B. As a result, ambiguity concerning welfare, measured by the gap between EV_A and EV_B, increases dramatically. Accordingly, unless one has adequate objective grounds for excluding the partially attentive frame from the welfare-relevant domain, shifting decisions to the fully attentive frame has the disadvantage of introducing substantially normative ambiguity. All of these conclusions also hold in simulations where the companies' matching provisions are removed (see Figure 14).

7 Concluding remarks

In this paper, we have attempted to make two distinct types of contributions. From a substantive perspective, we have offered new conceptual observations and quantitative results concerning a policy issue of considerable practical importance. From a methodological perspective, we have demonstrated the practicality of the framework for behavioral welfare economics developed in Bernheim and Rangel (2009).

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

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] 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.

[13] 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.

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

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

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

[17] 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

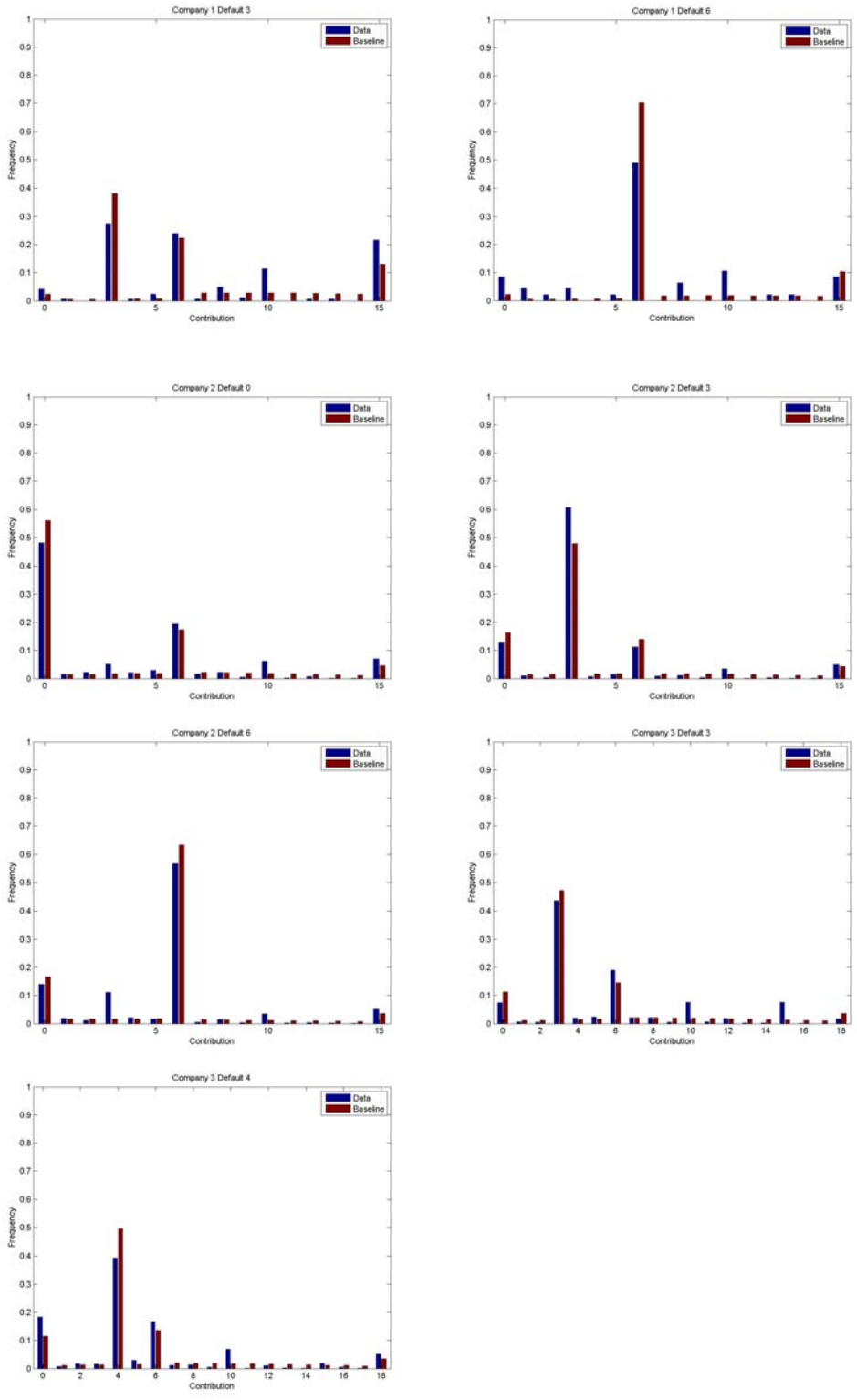

Figure 1: Fitted versus actual distributions

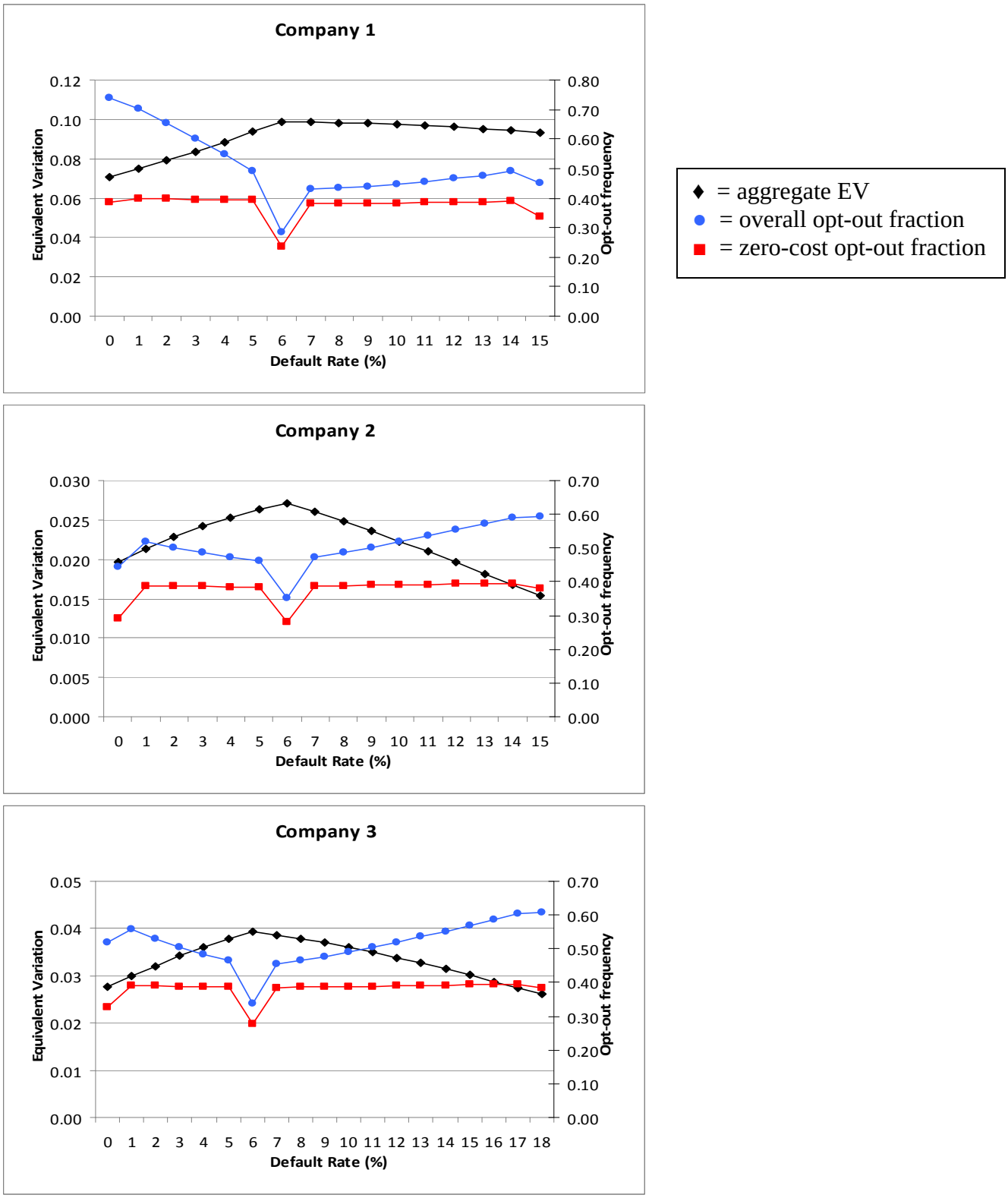

Figure 2: Average equivalent variation and opt-out frequencies, with an employer match

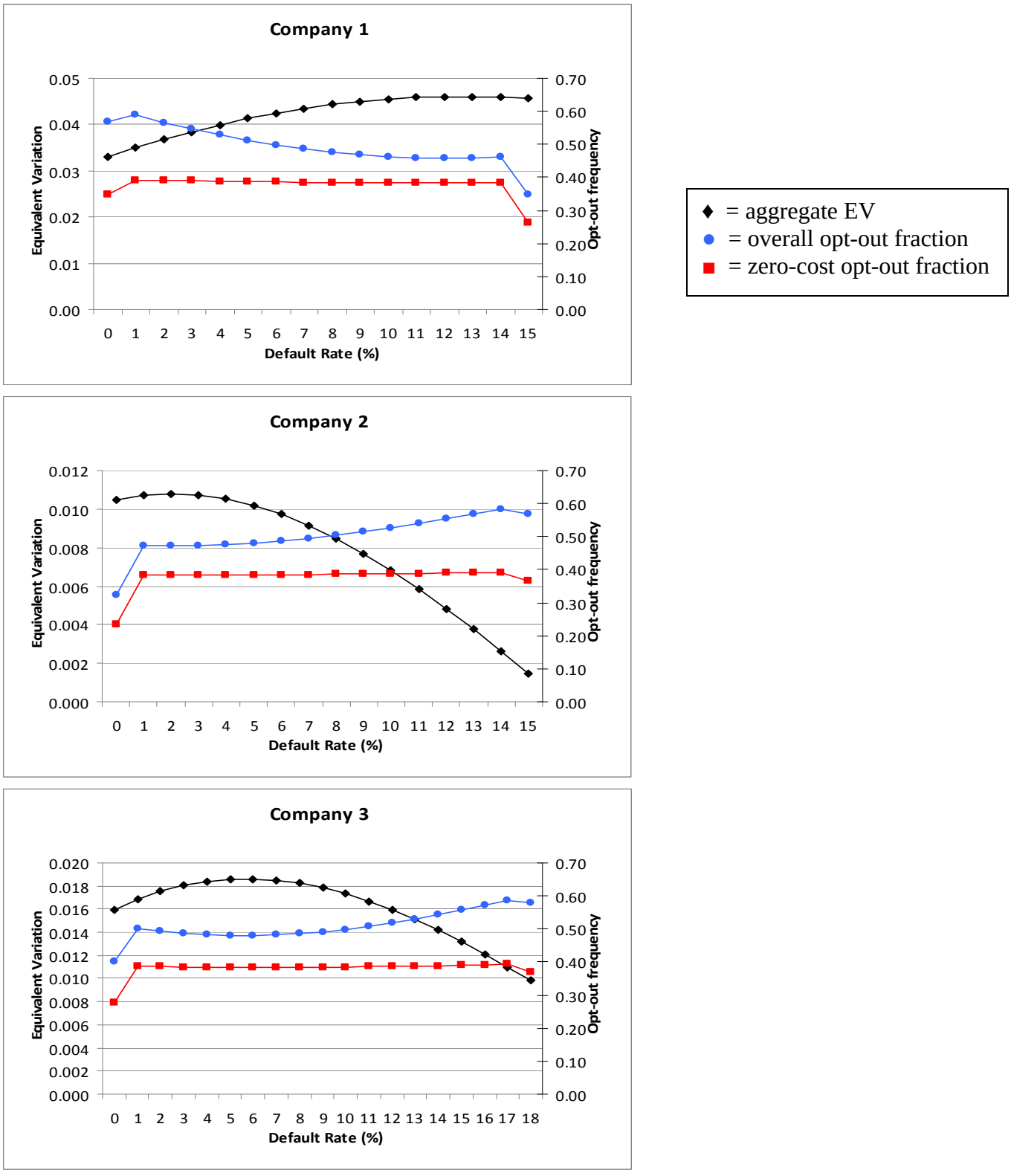

Figure 3: Average equivalent variation and opt-out frequencies, without an employer match

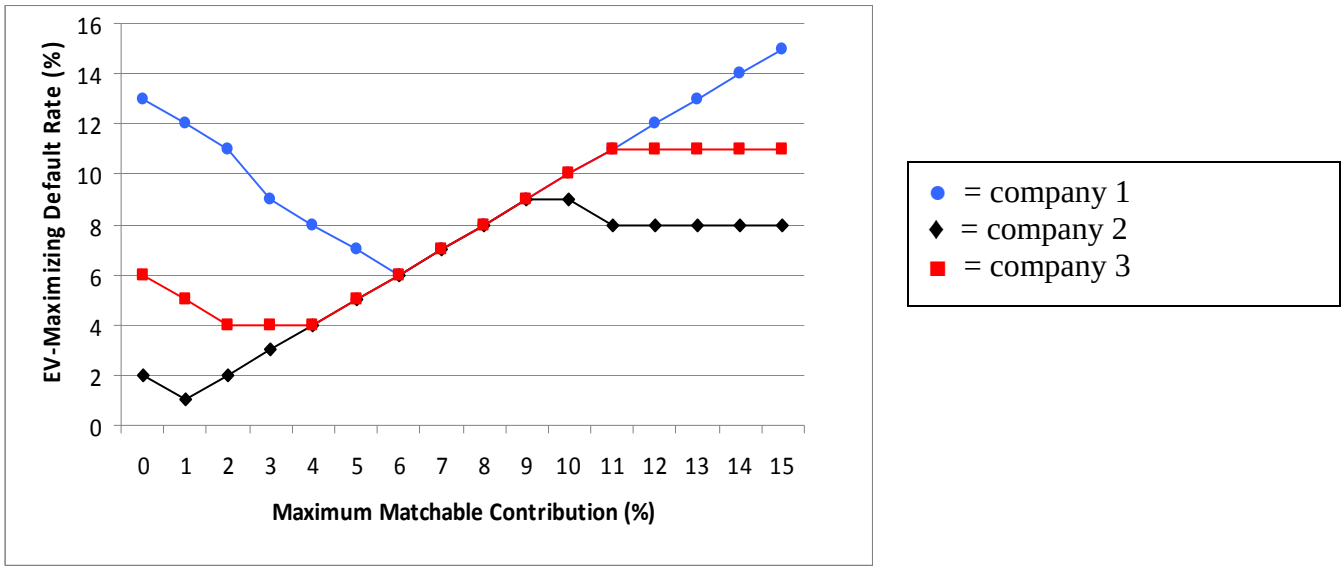

Figure 4: EV-maximizing default rate versus maximum matchable employee contribution

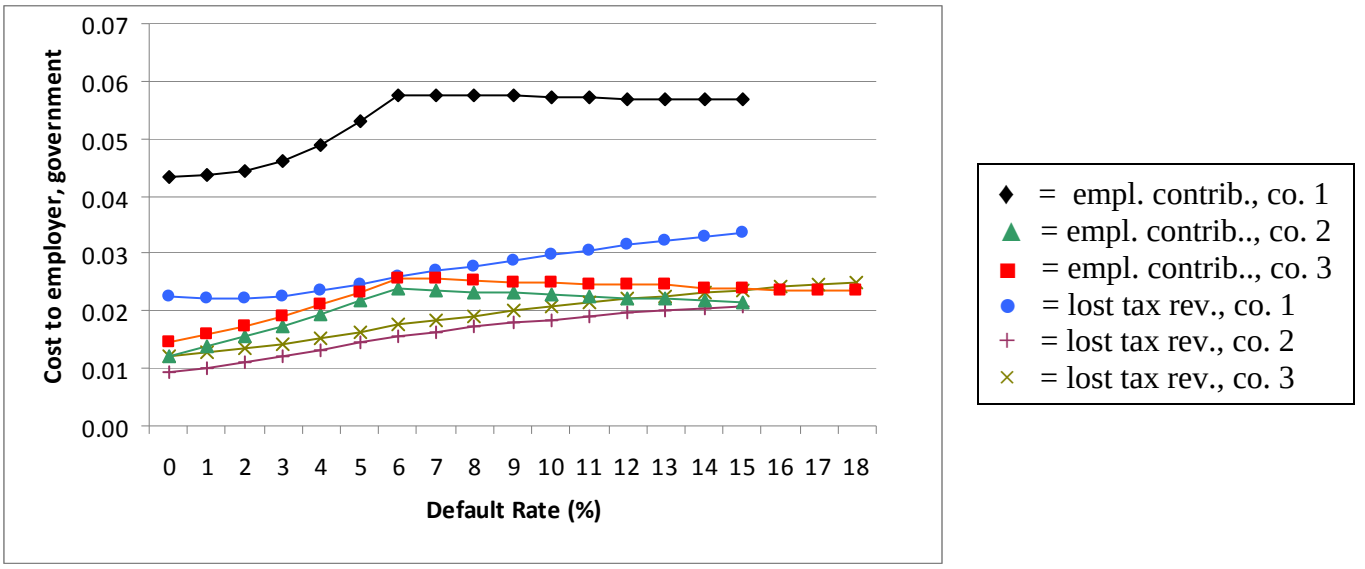

Figure 5(a): Employer contributions and lost government revenue versus default rate, base case

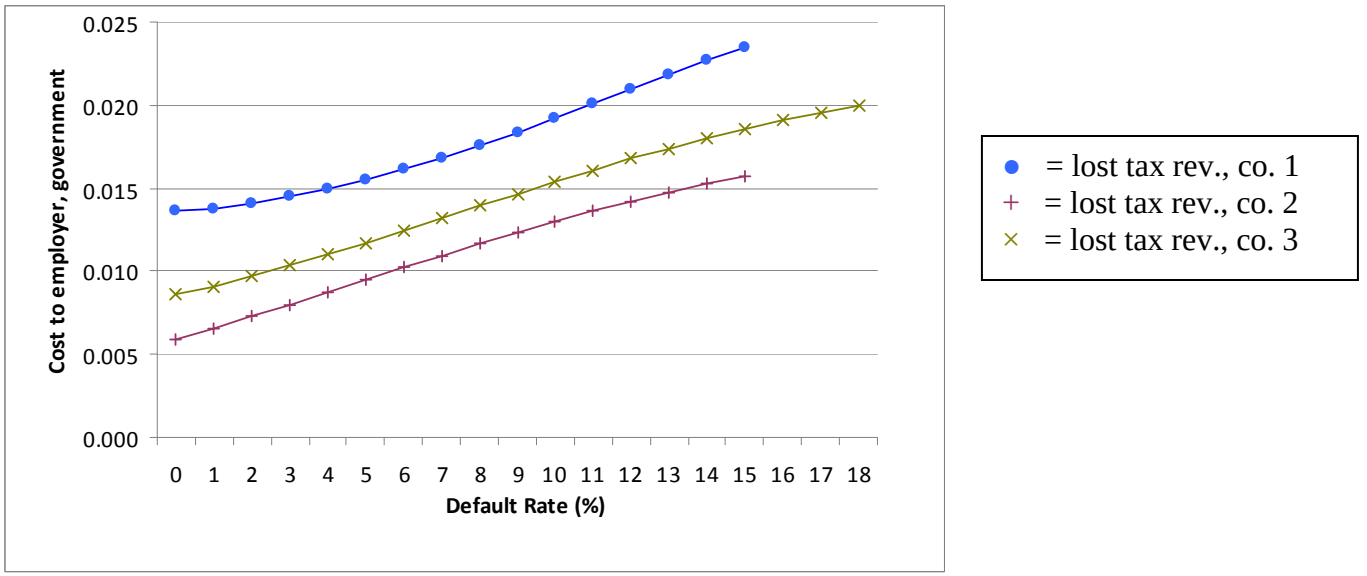

Figure 5(b): Lost government revenue versus default rate, no employer match

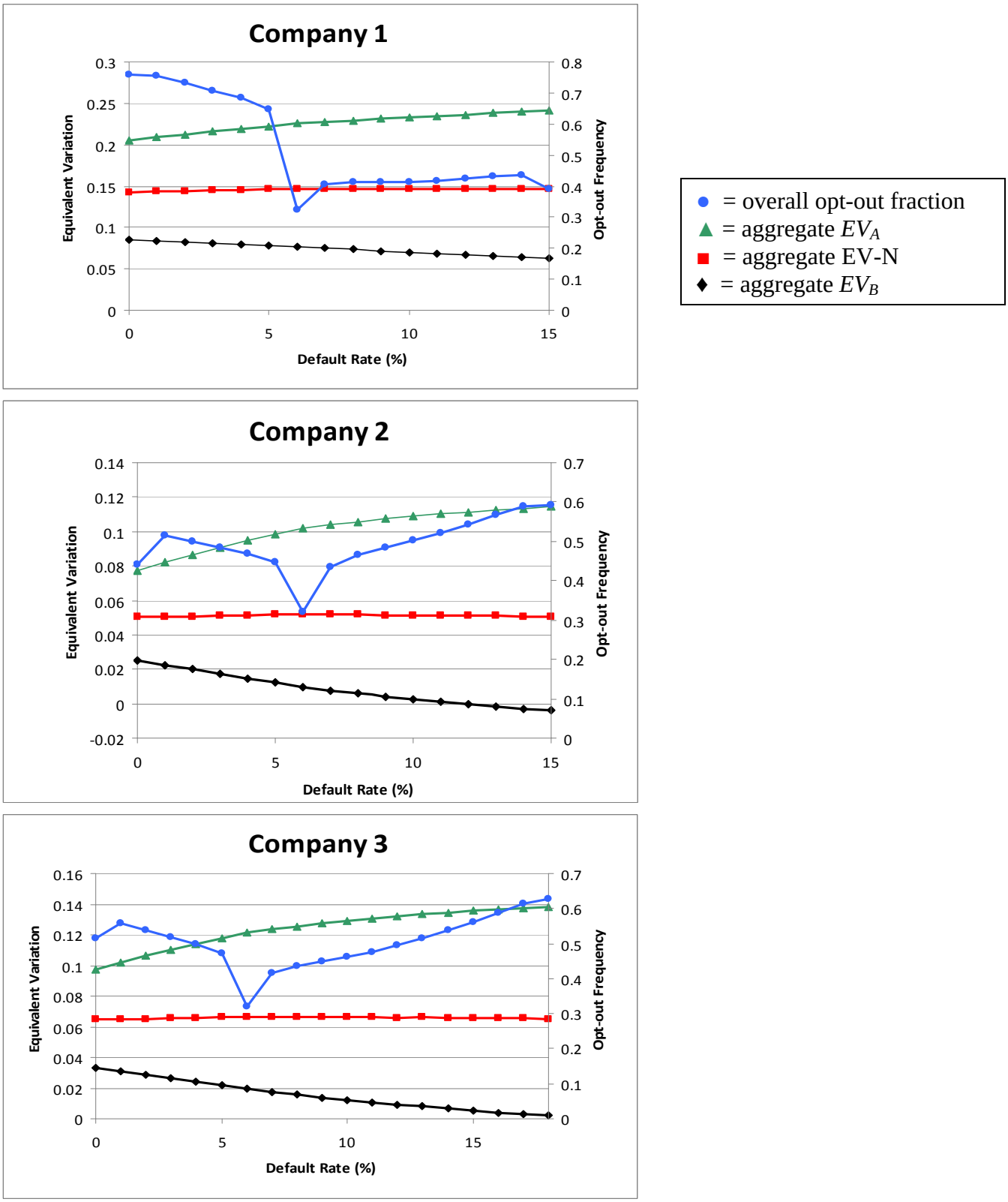

Figure 6: Average equivalent variation and opt-out frequency, with anchoring and an employer match

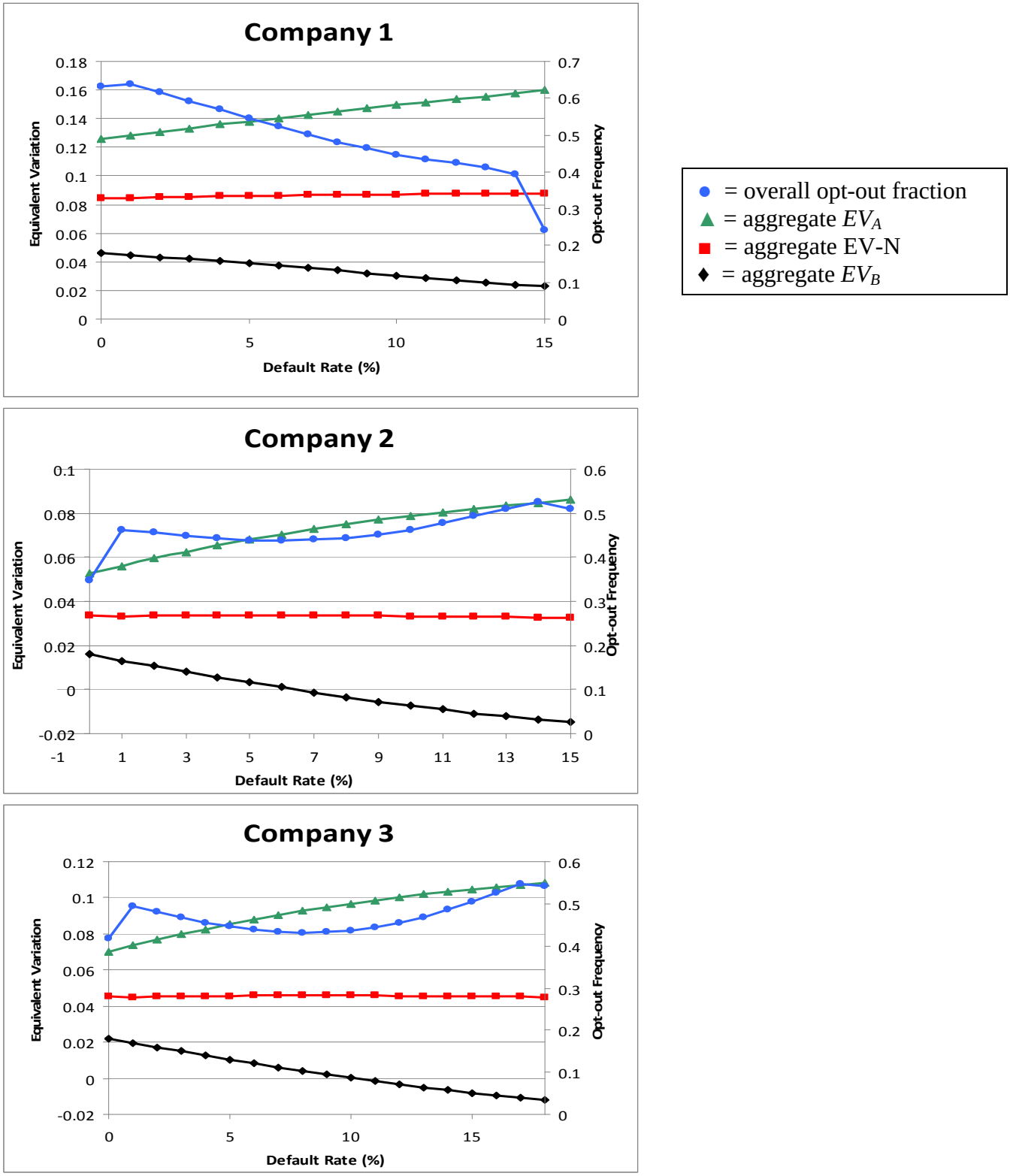

Figure 7: Average equivalent variation and opt-out frequency, with anchoring and no employer match

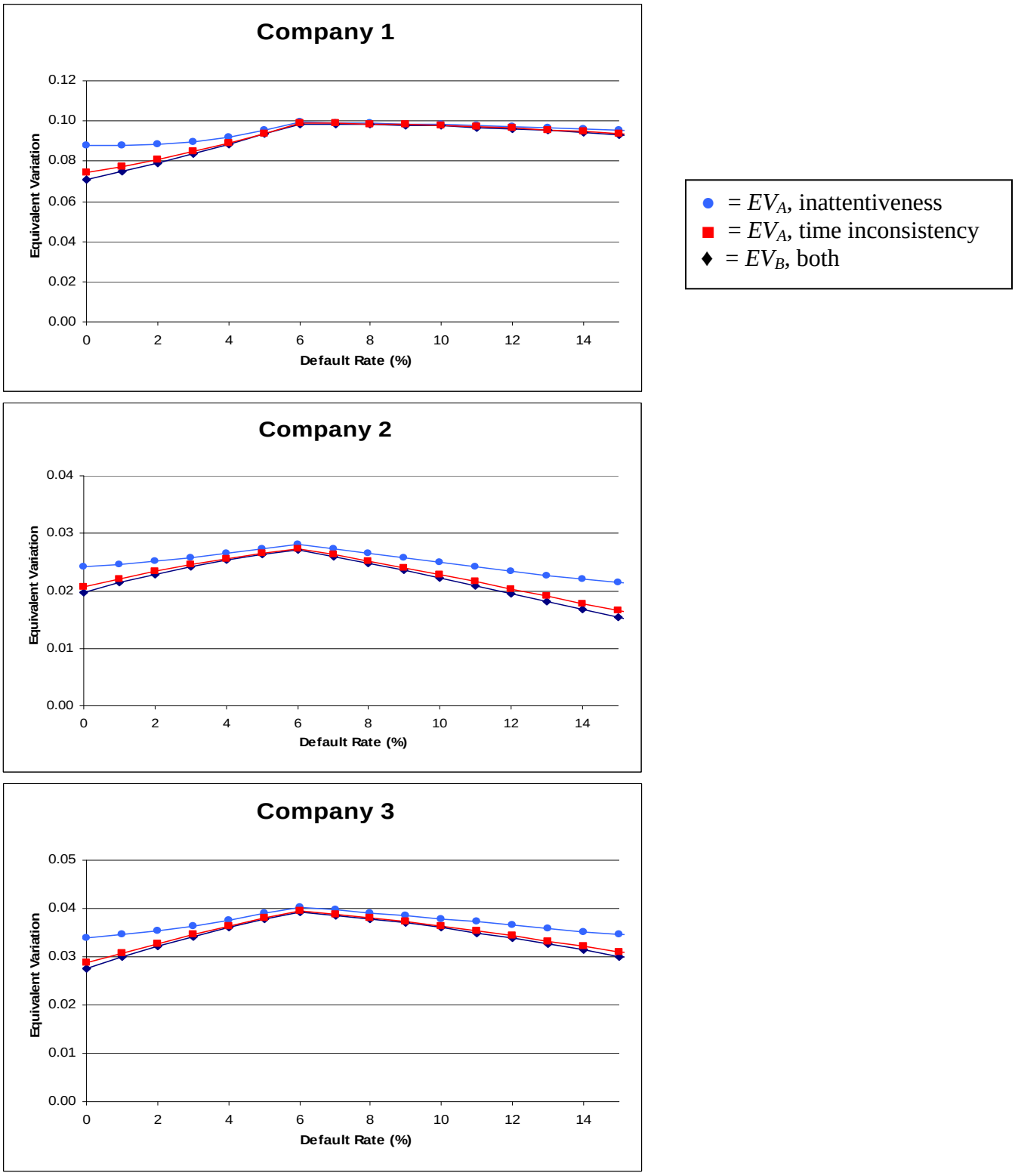

Figure 8: Average equivalent variation with time inconsistency or inattentiveness and an employer match

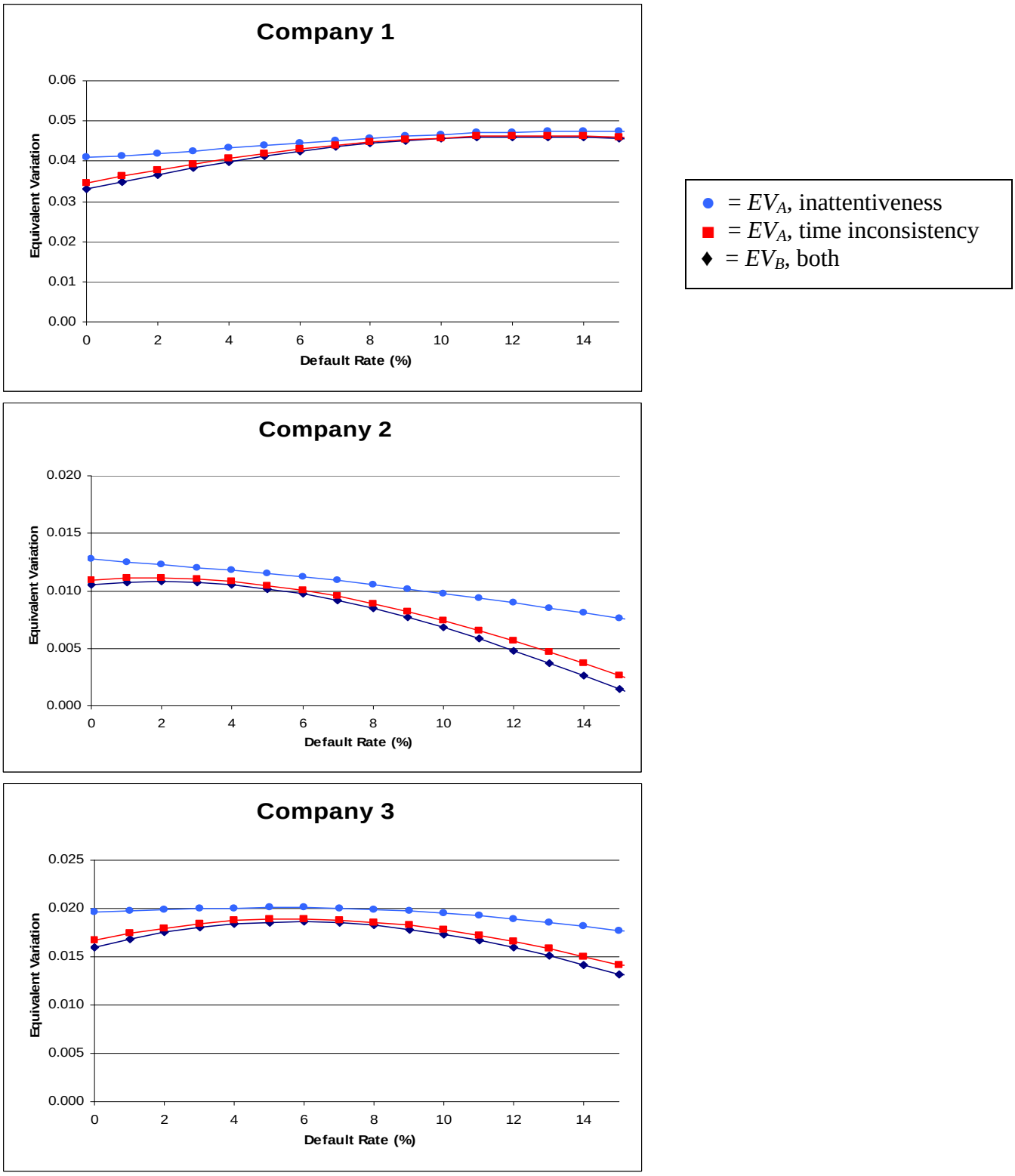

Figure 9: Average equivalent variation with time inconsistency or inattentiveness and no employer match

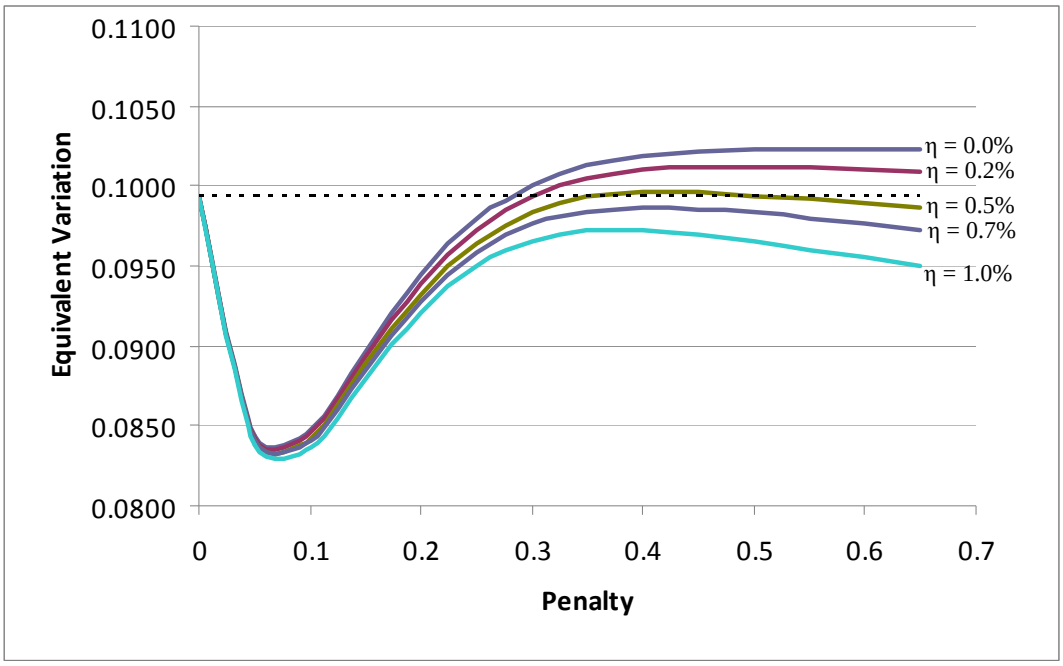

Figure 10: Average equivalent variation as a function of the penalty for inactive choice, for company 1 with a default of 6% and an employer match, various values of η (the fraction of workers who never make an active decision)

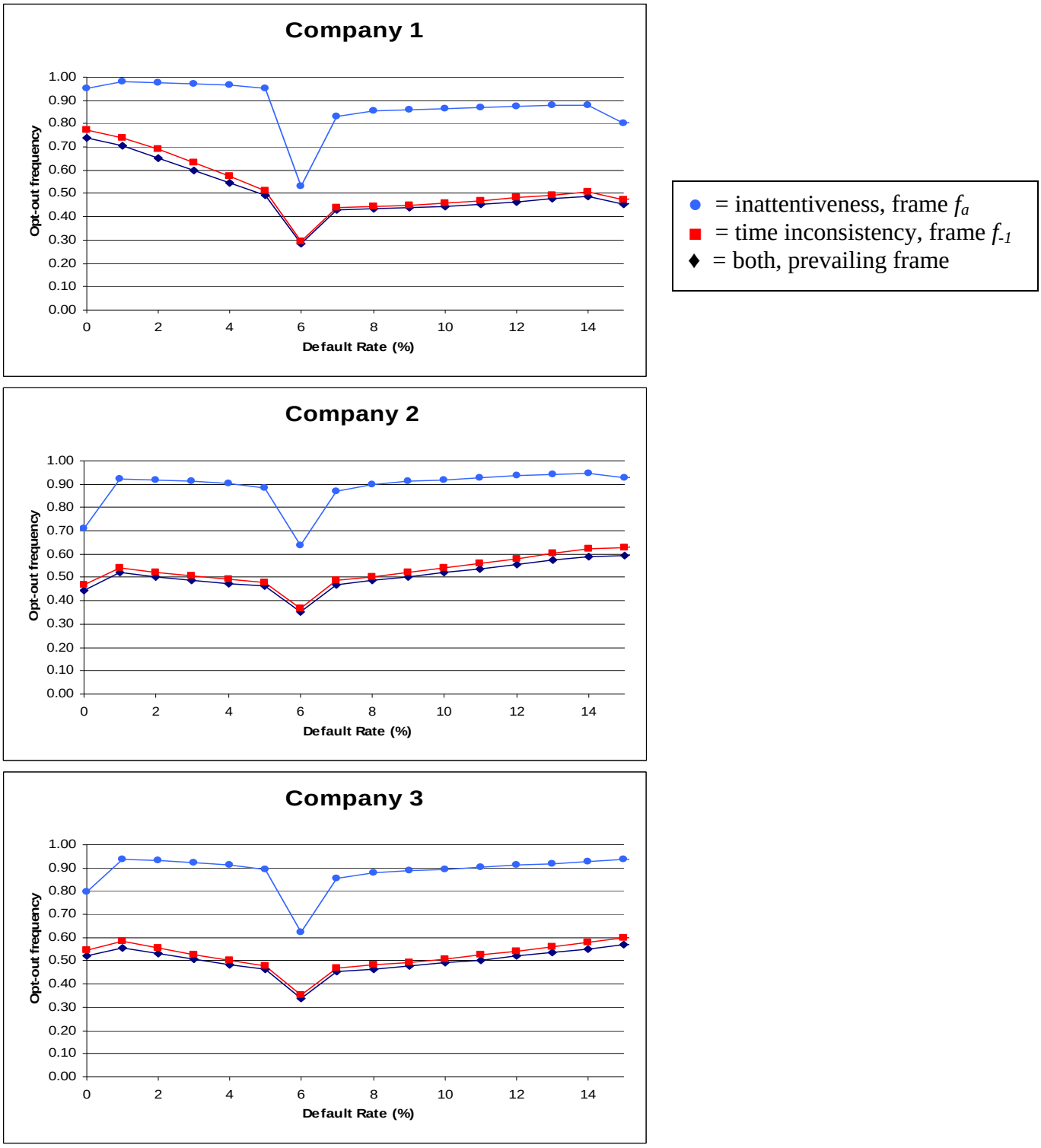

Figure 11: Opt-out frequencies for various decision frames, with an employer match

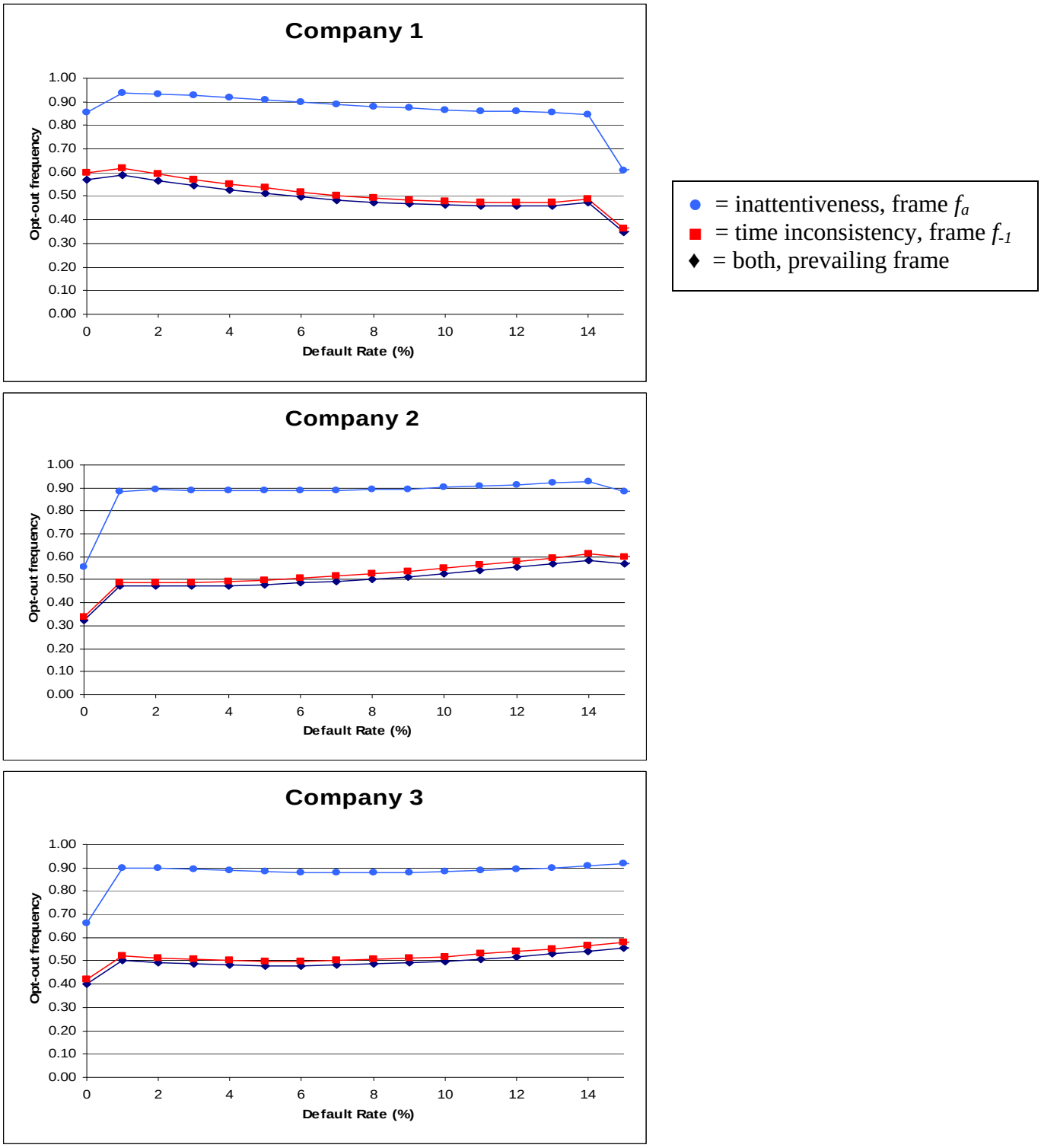

Figure 12: Opt-out frequencies for various decision frames, without an employer match

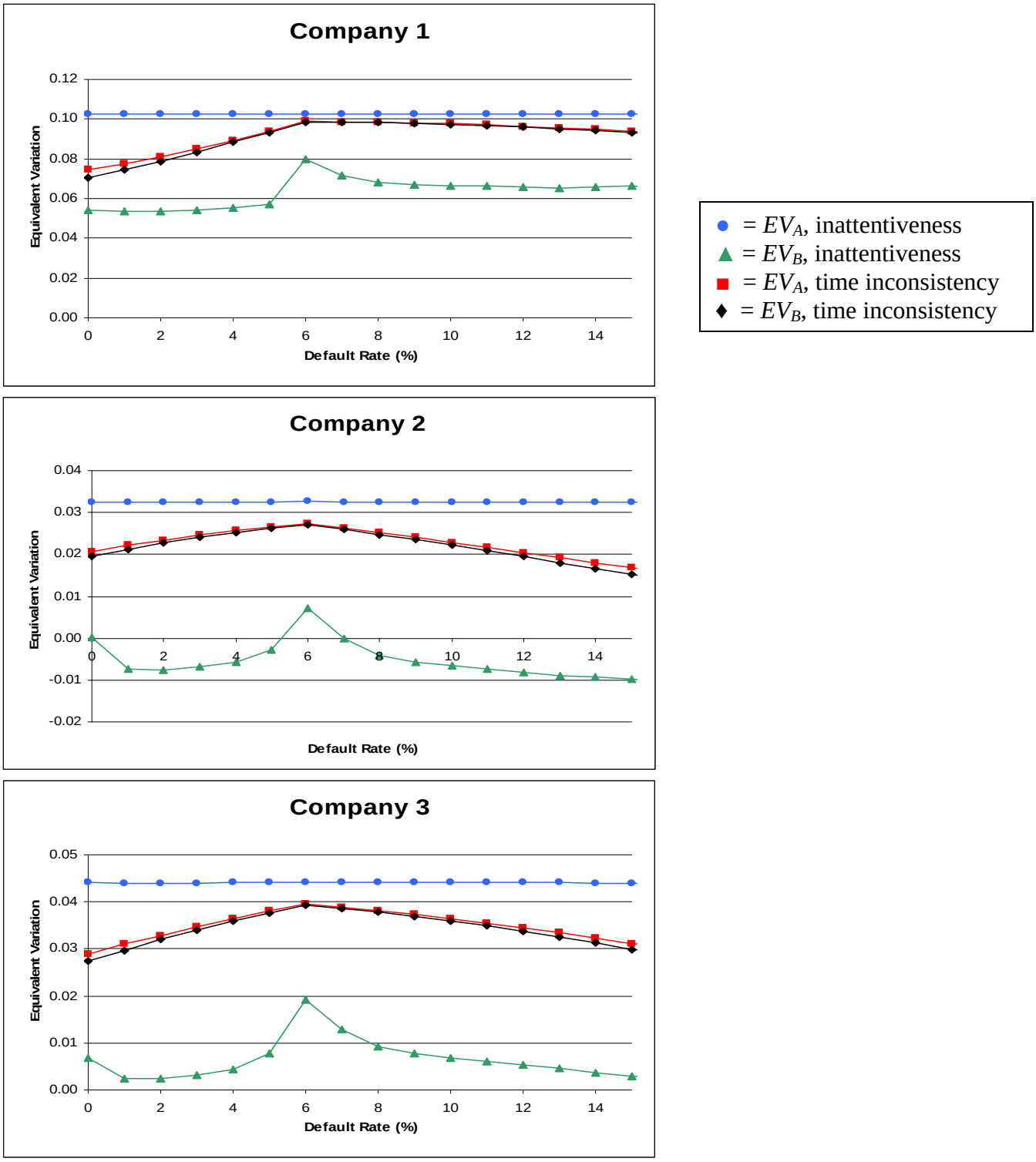

Figure 13: Average equivalent variation, decisions made in the forward-looking frame for time inconsistency and in the fully attentive frame for inattentiveness, with an employer match

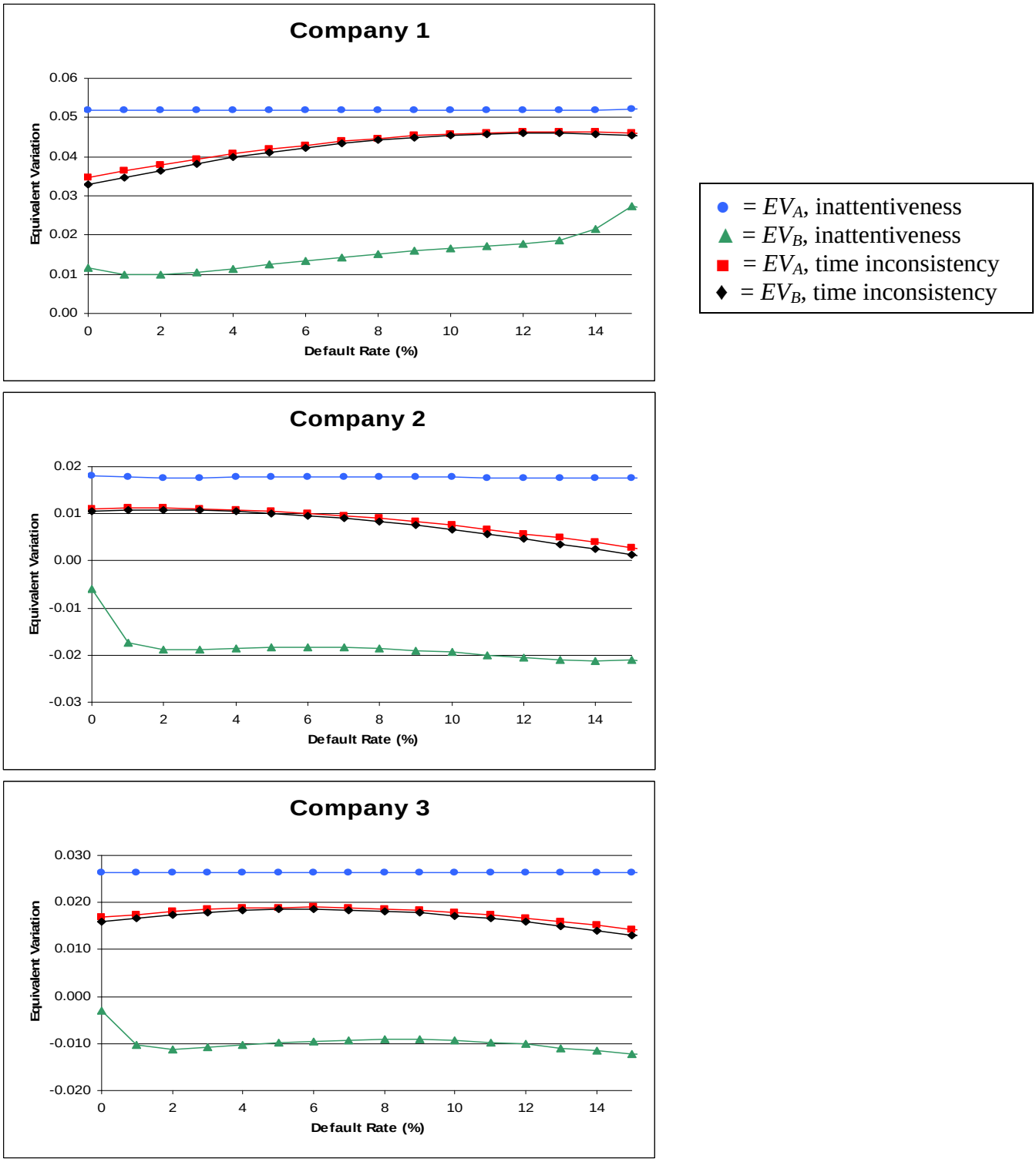

Figure 14: Average equivalent variation, decisions made in the forward-looking frame for time inconsistency and in the fully attentive frame for inattentiveness, with no employer match