AN EQUILIBRIUM THEORY OF RETIREMENT PLAN DESIGN

RYAN BUBB* AND PATRICK L. WARREN**

ABSTRACT. We develop an equilibrium theory of employer-sponsored retirement plan design using a behavioral contract theory approach. The operation of the labor market results in retirement plans that generally cater to, rather than correct, workers' mistakes. Our theory provides novel explanations for a range of facts about retirement plan design, including the use of employer matching contributions that result in cross-subsidization of rational workers by myopic workers and the use of default employee contribution rates in automatic enrollment plans that lower, rather than raise, workers' savings. These equilibrium outcomes call into question the practice of depending on employers to design plans to counteract the mistakes of workers.

1. INTRODUCTION

Employer-sponsored retirement savings plans are the predominant vehicle for private retirement savings in the United States. A growing literature shows that the design of these plans affects savings behavior in ways inconsistent with rational optimization (e.g., Madrian and Shea, 2001; Thaler and Benartzi, 2004). These empirical findings have informed normative claims by behavioral economists about how employers *should* design their plans. In a survey article, Benartzi and Thaler (2007) ask, "What can employers do so that more plan participants enroll in retirement plans, contribute an amount that will build a reasonable retirement nest-egg, and allocate the funds among assets in an appropriately diversified way?" They proceed to suggest to employers a range of plan design options to improve their workers' retirement savings outcomes. Employers should paternalistically harness the stickiness of default rules, for example, to counteract myopic workers'

temptation to save too little (Thaler and Benartzi, 2004; Carroll, Choi, Laibson, Madrian, and Metrick, 2009). These papers take a "public finance" approach to retirement plan design, modeling the employer as if it acts as a social planner, designing its retirement plan to maximize social welfare.

In response, Congress has enacted legislation that attempts to harness employers to correct their workers' mistakes. At the urging of behavioral economists, Congress removed regulatory barriers to employers automatically enrolling their workers in their retirement plan. The hope behind this approach was that employers would then adopt default contribution rates that would increase retirement savings by pointing workers in a pro-saving direction when they fail to make active decisions on their own (Orszag, Iwry, and Gale, 2006).

The existing literature, however, has not considered whether such paternalism is consistent with employers' incentives, and in particular with the incentives produced by the operation of the labor market. We develop an equilibrium theory of employer-sponsored retirement plan design using a behavioral contract theory approach. Retirement plans are an important feature of compensation contracts, designed by employers to attract workers. The approach we take is essentially neoclassical: in our model firms maximize profits and workers maximize their utility. The rational benchmark entails a simple wage contract. Retirement plans serve no useful purpose for rational, time-consistent exponential discounters (tax benefits aside). However, following the behavioral literature on retirement savings, we allow workers' decision utility at the time of contracting to deviate from their experienced utility in canonical ways and characterize the equilibrium retirement plan designs that result.

First, we consider present-biased or myopic workers with varying degrees of sophistication. Perfectly sophisticated myopic workers, who understand that they suffer from a time-inconsistency problem, value retirement plans with employer contributions as a form of commitment, and in equilibrium receive a plan that acts as a first-best commitment device. Imperfectly sophisticated myopic workers, in contrast, overestimate their future savings. As a result, naive myopic workers overvalue firms' offers to match their retirement savings and hence receive such retirement plans in equilibrium. While matching contributions can help offset naive workers' present bias, their

level in equilibrium is not finely calibrated to workers' need for commitment. Moreover matching results in cross-subsidization of rational workers by naive myopic workers.

Second, we incorporate forms of passive savings behavior that have been well-documented in the literature on defaults and extend the model to analyze equilibrium default employee contribution rates. When some workers can be influenced by the default, employers have incentives to choose the default that *minimizes*, rather than raises, worker savings given the other terms of the contract. Doing so reduces the level of matching contributions they must make, relaxing their zero-profit constraint, and thus allows them to offer better terms on the salient dimensions of compensation.

Whether a positive default contribution rate (i.e., automatic enrollment) results in lower savings in the equilibrium contract than a zero default (i.e., opt-in) depends on the parameters. Employers will use automatic enrollment in equilibrium if and only if the resulting reduction in savings under the contract by those who follow defaults as implicit advice is larger than the increase in savings by procrastinators who would not have opted in on their own. If the employer does automatically enroll employees, it will set the default contribution amount below the contract's cap on employee contributions that the employer matches.

Our theory provides novel explanations of many key facts about employer retirement plan design, showing the power of applying standard models of market equilibrium to understanding these plans. Most defined contribution plans offer matching contributions, and a substantial fraction of workers in such plans fail to contribute enough to receive the full match. Moreover, most employers that have adopted automatic enrollment have chosen the minimum default initial contribution rate allowed under the regulatory safe-harbor Congress created for such plans. In the vast majority of automatic enrollment plans, the default is set below the amount needed to receive the full employer match. Existing evidence suggests that employer adoption of automatic enrollment has failed to raise, and may even have lowered, overall retirement savings (Bubb and Pildes, 2014).

The approach we take to analyzing employer-sponsored retirement plans builds on an existing literature in behavioral contract theory that so far has focused on product markets, such as consumer credit (DellaVigna and Malmendier, 2004; Heidhues and Kőszegi, 2010; Bar-Gill, 2012),

cell-phone service (Grubb, 2009), add-on goods (Gabaix and Laibson, 2006), and gym memberships (DellaVigna and Malmendier, 2006). One justification for not taking a similar approach to understanding employer-sponsored retirement plans is the view that markets do not provide important incentives for employers with respect to retirement plan design. For example, Barr, Mullainathan, and Shafir (2013) argue that attempts to boost participation in retirement plans face "at worst indifferent and at best positively inclined employers and financial firms." They contrast this with other markets, like consumer credit, in which firms have strong incentives to exploit consumer mistakes. But as we show in this paper, a standard equilibrium model in which firms maximize profits and workers maximize their decision utility produces a rich positive theory that matches many key stylized facts about employer-sponsored retirement plan design. There are, of course, motivations for employers in designing their retirement plans that we ignore in our model, ranging from reputational concerns to tax incentives, but the essentially neoclassical model we develop here is the natural place to begin.

Recent attempts by behavioral economists to reform employer-sponsored retirement plans by simply showing employers what plan designs would improve worker savings outcomes and removing regulatory barriers to offering them (see, e.g., Thaler and Benartzi, 2004; Orszag, Iwry, and Gale, 2006) are unlikely to be effective. In our model, even if employers would like to act paternalistically, due to intrinsic preferences or otherwise, competition in the labor market leaves no room for such paternalistic motivations to be expressed. Meaningful employer paternalism in retirement plan design requires a concurrence of two factors that is unlikely to be widespread: employers must *both* be paternalistically motivated and have significant market power. If the motivation for retirement savings policy generally, and the preferential tax treatment of employer plans specifically, is to correct mistakes workers make in planning and saving for retirement (Kotlikoff, 1987), then our analysis shows that the delegation of plan design to employers will result in perverse outcomes for the myopic and inertial workers that retirement savings policy aims to help.

2. Baseline Model with Homogenous Types

Consider a perfectly competitive labor market. Labor contracts specify a wage $w \geq 0$ and a retirement plan that is composed of a non-elective employer contribution to the plan, $r \geq 0$, plus employer matching contributions of $m \geq 0$ dollars for every dollar the worker saves for retirement. Total employer retirement plan contributions are thus $r + sm$, where $s \geq 0$ is the amount the worker voluntarily chooses to save for retirement. We model retirement plans in this way to match the basic structure of employer-sponsored retirement plans we observe in the real world. Profits from an employed worker are given by $\pi = \gamma - w - sm - r$, where γ is the value the worker produces.

Workers have access to a savings technology through their employer's retirement savings plan with a rate of return normalized to zero, but they cannot borrow. For simplicity we have assumed away any motivation to save outside of the employer's retirement plan. There is thus no need to incorporate taxes to reflect the favorable tax treatment of employer-sponsored retirement plans relative to other forms of savings.

There are three periods in which the sequence of decisions is as follows.

- Period 0: Firms make contract offers (w, r, m) and workers choose among offers.
- Period 1: Workers receive wage w and decide how much of the wage to save, s, consuming the remainder, $w - s$.
- Period 2: Retired workers consume their savings and retirement plan benefits, $r + (1+m)s$.

A worker's period-0 self ("self 0") has utility $u(c_1) + u(c_2)$, where c_i is anticipated consumption in period i, $u(\cdot)$ is increasing and concave, and the discount factor is normalized to one.

Self 1, by contrast, chooses savings to maximize the utility function $u(c_1) + \beta u(c_2)$, where $\beta \in (0, 1]$ is the worker's time-inconsistent present-bias factor. Thus, facing a contract (w, r, m), self 1 solves,

$$(1) \qquad \max_{s \geq 0} u(w - s) + \beta u(r + (1 + m)s).$$

If $\beta < 1$, we refer to the worker as myopic. If $\beta = 1$, we refer to the worker as rational.

Self 0 chooses a contract to maximize her utility taking into account her anticipated future savings behavior. Importantly, however, we assume that self 0 *believes* that self 1 will choose savings by applying a present bias factor $\hat{\beta} \in [\beta, 1]$, following O'Donoghue and Rabin (2001)'s approach to modeling partial naivete. We refer to myopic workers with $\hat{\beta} = \beta$ as sophisticated, and to those with $\hat{\beta} > \beta$ as naive.

We begin by assuming homogenous workers of a single type $(\beta, \hat{\beta})$. This can also be thought of as the case in which firms observe workers' types so that each type gets its own contract.

Firms are willing to offer any contract that would result in nonnegative profits, given workers' actual savings behavior, but perfect competition implies that firms must break even in equilibrium. Equilibrium labor contracts are the zero-profit contracts that maximize self 0's utility, given her beliefs about self 1's savings behavior. They are thus the solution to,

(2)
$$\max_{w,r,m} u(w - s(w, r, m|\hat{\beta})) + u(r + (1 + m)s(w, r, m|\hat{\beta})),$$

subject to,

(3)
$$w + r + ms(w, r, m|\beta) = \gamma,$$

(4)
$$s(w, r, m|\hat{\beta}) = \arg\max_{s \geq 0} u(w - s) + \hat{\beta}u(r + (1 + m)s),$$

(5)
$$s(w, r, m|\beta) = \arg\max_{s \geq 0} u(w - s) + \beta u(r + (1 + m)s).$$

Self 0 wants to maximize the sum of her utility from consumption in the two periods, as reflected in the objective function in (2). The zero-profit constraint (3) requires that total compensation paid across the two periods must equal the worker's product γ. By concavity of the utility function, the first-best outcome equates consumption in each of the two periods at $\gamma/2$. Self 0 chooses a contract based on her belief that self 1 will put a present-bias factor of $\hat{\beta}$ on second-period utility when choosing how much to save under the contract; her anticipated savings level is determined

by (4). Her self 1 will actually make savings decisions according to (5), using a present-bias factor of $\beta \leq \hat{\beta}$.

Consider first a sophisticated myopic worker. A sophisticated worker's self 0's beliefs about her self 1's savings are correct, since $\hat{\beta} = \beta$. The problem for a sophisticated worker's self 0 is to choose a contract that induces her present-biased self 1 to save optimally. It is easy to see that a sophisticated worker will be willing to choose $r = w = \gamma/2$ to solve her time-inconsistency problem through r and achieve the first best. This contract will give self 1 exactly what self 0 wants her to consume. Self 1 will want to consume even more than $\gamma/2$ in the first period, but the remaining $\gamma/2$ of her compensation is only paid in the second period through r.

A sophisticated worker can also achieve the first-best through m. The first-order condition for self 1's choice of savings in (5) is:

$$(6) \qquad -u'\big(w - s(w, r, m | \beta)\big) + \beta(1 + m)u'\big(r + (1 + m)s(w, r, m | \beta)\big) = 0.$$

Thus choosing m such that $1 + m = 1/\beta$ will perfectly counterbalance self 1's present-bias, inducing self 1 to make savings decisions according to self 0's preferences, i.e., to equate her consumption in the two periods. Denote this m as $m^{FB} \equiv \frac{1-\beta}{\beta}$.

In the case of a rational worker, $m^{FB} = 0$ because matching would inefficiently subsidize second-period consumption, leading to a costly distortion in rationals' intertemporal consumption choices. Rationals are better off receiving their compensation through the lump-sum payments of w and r. They are indifferent among zero-profit contracts with $r \leq \gamma/2$, since they can simply choose savings to achieve the first-best levels of consumption under any such contract.

Because both sophisticated workers and rational workers have correct beliefs about their future behavior, they receive first-best contracts in equilibrium, as we now state formally:

Proposition 1. *In equilibrium,*

    (1) Sophisticated workers choose either a matching contract with $m^* = m^{FB}$ or a non-elective contribution contract with $r^* = \gamma/2$ and achieve the first best.

(2) Rational workers choose contracts with $r^* \leq \frac{\gamma}{2}$, $w^* = \gamma - r^*$, and $m^* = 0$ and achieve the

first best.

A naive worker's self 0, in contrast, underestimates her degree of myopia and hence her need for commitment. But a naive worker also has a different motivation for using m: she overestimates how much she will save under a given m and therefore the amount of matching contributions she will receive. Even a completely naive worker, with $\hat{\beta} = 1$, who has no awareness of her time-inconsistency and therefore no demand for commitment devices per se, will nonetheless demand some amount of matching contributions due to this mistake. What level of m will the worker choose? In the behavioral contract theory literature, naive present-biased agents generally do not demand first-best commitment contracts (see, e.g., DellaVigna and Malmendier, 2004; Heidhues and Kőszegi, 2010). This is also true in our setting, as the following proposition confirms.

Proposition 2. *Assume $u(c_i)$ takes a CRRA form with coefficient of relative risk aversion equal to θ. Then in equilibrium, naive workers choose a matching contract with $m^* > 0$ such that,*

(1) If $\theta < 1$ then $m^* > m^{FB}$;

(2) If $\theta = 1$ then $m^* = m^{FB}$;

(3) If $\theta > 1$ then $m^* < m^{FB}$.

Naive workers are attracted to matching contracts through a mix of commitment motivation and overestimation motivation, but in equilibrium do not in general receive a first-best matching contract. The elasticity of intertemporal substitution (EIS), $1/\theta$, determines whether the equilibrium match over- or under-shoots first-best since it reflects the willingness of workers to tolerate unequal consumption over time. Workers with a relatively high EIS have a lower willingness to accept for tolerating such unequal consumption, and their overestimation of the value of the match results in them choosing a match that overshoots the first-best. Rather surprisingly, if $\theta = 1$, so that $u(c_i) = \ln(c_i)$, then the overestimation motivation produces the first-best commitment contract even for naive workers who underestimate their need for commitment. This counter-intuitive result of first-best commitment for naives, however, is a knife-edge special case. Indeed, we show

in the Online Appendix that log utility is not only sufficient for producing the first-best contract for all degrees of naivete, it is also necessary. That is, for every other increasing, concave function $u(\cdot)$, there exists $\hat{\beta} \in (\beta, 1]$ such that $m^* \neq m^{FB}$.

3. Heterogenous Types

Consider now the case of heterogenous worker types, $(\beta, \hat{\beta}) \in \Theta$, in which firms do not observe workers' types. For the rest of the paper, we assume that $u(c_i) = \ln(c_i)$. We know from Proposition 2 that with log utility, every $(\beta, \hat{\beta})$ type achieves efficiency when contracting on their own, so this assumption means that any deviation from efficiency we show here must result from the interaction among the types. We focus on the tractable but still analytically rich case with three types: $\Theta = \{(\beta^r, \hat{\beta}^r), (\beta^n, \hat{\beta}^n), (\beta^s, \hat{\beta}^s)\}$. A fraction κ^r of workers are rational exponential discounters ($\hat{\beta}^r = \beta^r = 1$); a fraction κ^n are naively myopic, with $\beta^n < \hat{\beta}^n \leq 1$; and a fraction κ^s are myopic but sophisticated, with $\hat{\beta}^s = \beta^s = \beta^n$. Assume that all three types have positive population shares.

Thus far we have assumed that firms can only offer simple linear matching contracts. In reality, retirement plans that include matching always also include a cap on the amount of employee contributions that will be matched, typically in the form of some percentage of the wage. In the homogenous type case, allowing such caps would make no difference in equilibrium outcomes, but with heterogenous types caps play an important role.[1] In this section we thus also expand the contract space to allow for a cap on the matched savings, $c \geq 0$. Formally, represent a contract as a 4-tuple (w, r, m, c), where a worker saving s attains first-period consumption $w - s$ and second period consumption $r + s + m \min\{s, c\}$.

Our definition of a competitive equilibrium follows Heidhues and Kőszegi (2010)'s adaptation of Rothschild and Stiglitz (1976)'s approach to the case with myopic agents:

Definition 1. *A* competitive equilibrium *is a set of contracts* $C^* = \{(w^i, r^i, m^i, c^i)\}_{i \in \{r,n,s\}}$ *such that:*

(1) Workers with $\hat{\beta} = \hat{\beta}^i$ prefer contract (w^i, r^i, m^i, c^i) to the other contracts in C^*.

---

[1]In an earlier working paper version of this paper we analyzed the heterogenous type case without caps and showed that naives pool with rationals only if they are sufficiently naive. As we show below, with caps naives pool with rationals for all levels of naivete.

(2) Each (w^i, r^i, m^i, c^i) yields nonnegative profits given the types that prefer it.

(3) There does not exist an alternative contract (w', r', m', c') that is strictly preferred to (w^i, r^i, m^i, c^i) by workers with $\hat{\beta} = \hat{\beta}^i$ for some $i \in \{r, n, s\}$ and that would make nonnegative profits if it were chosen by the worker types that strictly prefer it.

Note that if $(w^i, r^i, m^i, c^i) = (w^j, r^j, m^j, c^j)$ for some $i \neq j$ in an equilibrium, then types i and j are pooling in a single contract, and the profits of the contract reflect the presence of both types for purposes of testing the nonnegative-profit condition. We will denote the contract terms of such an equilibrium pooling contract by $w^{ij} \equiv w^i = w^j$, and so forth, in a slight abuse of our notation. Note as well that this definition is akin to focusing on pure-strategy equilibria, since each type chooses only a single contract in equilibrium. The following proposition characterizes the competitive equilibria of the model with heterogenous types.

Proposition 3. *Assume $\Theta = \{(\beta^r, \hat{\beta}^r), (\beta^n, \hat{\beta}^n), (\beta^s, \hat{\beta}^s)\}$. Competitive equilibria exist and in them:*

(1) All naives and rationals pool together in a contract such that,

(a) Savings are matched at a rate m^{rn} such that $0 < m^{rn} < m^{FB}$, up to a cap $c^{rn} = s(w^{rn}, r^{rn}, m^{rn}|\hat{\beta}^n) > 0$.

(b) Rationals save at the matching cap.

(c) Naives anticipate saving at the matching cap but in fact save strictly less than the matching cap.

(2) Sophisticates separate into a contract that delivers consumption of $\gamma/2$ in each period.

Naive and rational workers in equilibrium pool together in a contract that offers matching contributions with a matching cap set at the naives' anticipated savings level. Naive workers are attracted to this contract because they overestimate the amount of matching contributions they will receive, much like in the homogenous type case. One important difference with heterogenous types is that rational workers drive down the wage of the matching contract due to their relatively high savings rate. In the homogenous type case, naive workers are always paid their marginal product of labor,

γ. Here, the *average* total compensation of naives and rationals in the pooling contract equals γ. But because rational workers save more than naive workers and therefore receive a greater amount of matching contributions, the rationals receive compensation greater than γ while naives receive less than γ. The pooling contract thus results in cross-subsidization of rational workers by naive workers. By lowering their total compensation, this cross-subsidization results in naives doing worse in this pooling contract than they would in their best separating contract.

To see why naive workers are nonetheless willing to pool with rational workers, despite this cross-subsidization, first note that because the matching cap is set at naive workers' anticipated savings level, naive workers anticipate receiving $m^{rn}c^{rn}$ in matching contributions, which is the maximum possible matching contribution under the contract. Relative to an $m = 0$ contract, any contract with $m > 0$ must offer lower $w + r$ by the amount $m\bar{s}(w, r, m, c)$, where $\bar{s}(w, r, m, c) \equiv E[\min\{s(w, r, m, c|\beta), c\}]$ is the average actual matched savings level under the contract. Because $m^{rn}c^{rn} > m^{rn}\bar{s}(w^{rn}, r^{rn}, m^{rn}, c^{rn})$ (because naives will in fact save less than c^{rn}), matching always looks like a good deal to naives.

The precise contract terms in any equilibrium pooling contract are those preferred by the naives from among all nonnegative-profit pooling contracts. The reason is that, if not, then that contract could enter, make naives strictly better off, and still make nonnegative profits. Rationals prefer this contract to any nonnegative-profit separating contract because of the cross-subsidy it provides.

Note that our theory implies that many workers will save strictly less than the amount needed to receive the full employer match offered. If all workers received the full match available, then the overestimation mechanism we have identified would break down.

Sophisticated workers, who have correct beliefs about their future savings, understand that naives' matching contract is a bad deal and never pool in a matching contract with the rationals. They instead choose a contract that serves as a first-best commitment device but avoids attracting (and therefore cross-subsidizing) the rationals, for example one in which all of their retirement consumption is financed by non-elective employer contributions or one that offers matching such that saving at the matching cap results in first-best consumption smoothing.

4. DEFAULT CONTRIBUTION RATES

We turn now to another important feature of the structure of plan contributions: the default rule for employee contributions to the plan. In traditional defined contribution plans like 401(k)s, workers have to submit paperwork to affirmatively opt in in order to contribute to the plan. Madrian and Shea (2001) studied an employer that adopted instead a positive default by automatically enrolling new hires into contributing 3% of their salary into its retirement plan unless they affirmatively opted out. They found that automatic enrollment dramatically increased the participation rate of new hires, from 37% to 85%. They also found that the majority of participants contributed the default amount when automatically enrolled. Importantly, the strictly positive default contribution rate under automatic enrollment was stickier than the zero default of the original opt-in design. They interpret the power of the default as stemming both from simple inertia of participants and also from some participants interpreting the default contribution rate as implicit advice from the employer about the *right* amount to contribute.

Behavioral economists seized on these findings to advocate that employers adopt automatic enrollment in order to increase savings (e.g., Thaler and Benartzi, 2004; Orszag, Iwry, and Gale, 2006). A group of economists at the Brookings Institution designed and successfully lobbied for the passage of legislative reforms in the Pension Protection Act (PPA) of 2006 to remove regulatory barriers to the adoption of automatic enrollment (see Beshears, Choi, Laibson, Madrian, and Weller, 2010, for an account of the legislative process). Employers have adopted it in droves, with the percentage of Vanguard-administered plans that use automatic enrollment more than doubling from 15% in 2007 to 36% in 2014 (Vanguard, 2015). The discovery and adoption of automatic enrollment is widely regarded as having improved retirement savings outcomes—Peter Orszag has referred to it as "a stunning example of the success of behavioral economics in affecting public policy" (quoted in Beshears, Choi, Laibson, Madrian, and Weller, 2010)—but there is little empirical evidence on its effects given the actual plan design choices of employers. In this section we apply our behavioral contract theory approach to analyze *equilibrium* default contribution rates.

Suppose that in addition to offering employer contributions in the contract as above, employers can now also specify a default contribution rate $d \geq 0$ for their retirement plan. We need to enrich our behavioral type space to account for the documented stickiness of defaults. Suppose that in addition to having a $(\beta, \hat{\beta})$, each worker also has a "default-sensitivity type": active chooser, procrastinator, or advice taker, denoted by $\theta \in \{a, p, t\}$. Active choosers ($\theta = a$) behave as in our model above. Procrastinators ($\theta = p$) believe in period 0 that they will save according to their $\hat{\beta}$ but in fact in period 1 will save $s = dw$ if $d \in [0, \bar{d}]$. If $d > \bar{d}$, they revert to saving according to their β, since at a sufficiently high level of default savings, even procrastinators will bear the costs of opting out. Finally, advice takers ($\theta = t$) believe in period 0 that they will save according to their $\hat{\beta}$, but in fact in period 1 will save $s = dw$ if $d \in [\underline{d}, \bar{d}]$, with $\underline{d} > 0$, and will save according to their β otherwise. The idea is that both very low and very high defaults are implausible as advice, but within an intermediate range the worker assumes that the default was chosen in an informed way and follows it. This is consistent with the evidence documenting that a higher fraction of workers stay at a low strictly positive default than stay at a default of zero (Madrian and Shea, 2001). These behavioral assumptions could be microfounded, as in Carroll, Choi, Laibson, Madrian, and Metrick (2009), but we adopt this reduced-form approach for simplicity.

Note that we have assumed that at the time of contracting, no worker type believes that the default d in the contract will affect them. There are two motivations for this assumption. First, this can be thought of as a form of naivete by both procrastinators and advice takers. Second, it could stem from inattention—the default rule governing employee contributions to the employer's retirement plan is far down on the list of important factors to consider when choosing among job offers and hence is simply not salient at the time of contracting.

The rest of the model is as above. We focus on the heterogenous type case with two myopia types (rational and naive) crossed with the three default-sensitivity types: $\Theta = \{(\beta^r, \hat{\beta}^r), (\beta^n, \hat{\beta}^n)\} \times \{a, p, t\}$. Rationals have $\beta^r = \hat{\beta}^r = 1$ and naive myopic workers have $\beta^n < 1$ and $\hat{\beta}^n \in (\beta^n, 1]$. Denote the probability of each type by $\kappa^{ij} = \kappa^i \kappa^j$ for $i \in \{r, n\}$ and $j \in \{a, p, t\}$. The following

proposition characterizes the equilibrium of the model, focusing on the case in which there is a relatively large proportion of active choosers.[2]

Proposition 4. *Assume* $\Theta = \{(\beta^r, \hat{\beta}^r), (\beta^n, \hat{\beta}^n)\} \times \{a, p, t\}$. *There exists a* $\underline{\kappa} < 1$ *such that if* $\kappa^a > \underline{\kappa}$ *then competitive equilibria exist and in them all workers pool in a contract such that:*

(1) Savings are matched at a rate $m^{rn} > 0$, *up to a cap* $c^{rn} = s(w^{rn}, r^{rn}, m^{rn}, c^{rn} | \hat{\beta}^n) > 0$.

(2) The default contribution rate d^{rn} *is the one that minimizes average worker savings in the contract, given the other terms of the contract, and the default contribution amount, $d^{rn} w^{rn}$, is strictly below the contract's matching cap. The adoption of automatic enrollment ($d^{rn} > 0$) does not increase average second-period consumption relative to the equilibrium outcome with the contract space is exogenously restricted to contracts with $d = 0$.*

(3) All workers anticipate saving at the matching cap, however:

    (a) Naives who make an active choice in fact save strictly less than the matching cap.

    (b) Only rationals who make an active choice save at the matching cap.

As in the case without defaults, naives and rationals pool in the nonnegative-profit matching contract most preferred by naives. In equilibrium the default is designed to *minimize* worker savings, conditional on the other terms of the contract. The reason is that defaults that reduce savings also reduce the employer's matching contributions, allowing the employer to offer even higher levels of salient forms of compensation. Key to this result, of course, is our assumption that defaults are not salient at the time of contracting, so that their only substantive effect is through relaxing firms' nonnegative-profit constraint. The default contribution amount is always set below the cap on matching, since otherwise it would not reduce matching payments. Note that this implication is in sharp contrast to what a paternalistic employer would do, which is to default workers at the savings rate that maximizes their experienced utility, given the other terms in the contract. That would entail setting the default contribution rate exactly at the cap on matched savings (Bernheim, Fradkin, and Popov, 2015). Similarly, there is no incentive for firms to try to make defaults less

---

[2]If there are not enough active choosers, an uninteresting equilibrium arises in which there is no meaningful interaction between the myopia types.

sticky by forcing all workers to make an active choice, as suggested by Carroll, Choi, Laibson, Madrian, and Metrick (2009). The reason is that doing so would always increase average savings under the contract relative to the optimal default.

The savings-minimizing default for any given w, r, m, and c is either $d = 0$ or $d = \underline{d}$. The firm will offer $d = \underline{d}$ instead of $d = 0$ if the drop in savings from the advice-takers, who now save at $s = \underline{d}w$ instead of actively choosing on their own, is greater than the increase in savings from moving the procrastinators from $s = 0$ to $s = \underline{d}w$, i.e. if $\kappa^{rt}s^r + \kappa^{nt}s^n - \kappa^t\underline{d}w > \kappa^p\underline{d}w$.

A key parameter that determines whether $d^{rn} = 0$ or $d^{rn} = \underline{d}$ is naives' present-bias factor β^n, since it affects how much advice-takers will save under $d = 0$ but not under $d = \underline{d}$. This is illustrated in Figure 1. Panel (A) shows the equilibrium match and default regime as a function of β^n. As β^n increases, the average savings rate increases. As we have discussed, this makes matching more expensive, resulting in lower m^{rn}. Since the lower matching rate applies to more than just the naive active savers, the net effect is to lower overall matching payments, driving up the wage, as shown in panel (C). Finally, recall that equilibrium contract terms are determined by the preferences of the naive workers, who prefer a cap set at their anticipated savings level. As the wage increases, the naive anticipates saving more, leading to a higher matching cap, as shown in Panel (B). By increasing the average savings rate of advice-takers under $d = 0$, but not under $d = \underline{d}$, increasing β^n makes the best nonnegative-profit contract with $d = \underline{d}$ relatively more attractive, resulting in a shift from $d^{rn} = 0$ (indicated by the thin line in the figures) to $d^{rn} = \underline{d}$ (indicated by the thick line). To illustrate the effect of allowing firms to adopt automatic enrollment, we also depict with a dashed line the equilibrium outcome if the contract space is exogenously limited to contracts with $d = 0$.

In our model firms do not use automatic enrollment to paternalistically increase savings, as urged in much of the literature in behavioral economics. This is starkly illustrated in Panel (D), which shows average second-period consumption in equilibrium as a function of β^n. Somewhat counter-intuitively, average second-period consumption is invariant to changes in β^n. Indeed, as we show in the proof of Proposition 4, average second-period consumption is equal to $\gamma/2$ for all

$$— \quad d^*{=}0 \qquad \text{━━} \quad d^*{=}\underline{d}$$

(A) Equilibrium match rate.

(B) Equilibrium matching cap.

(C) Equilibrium wage.

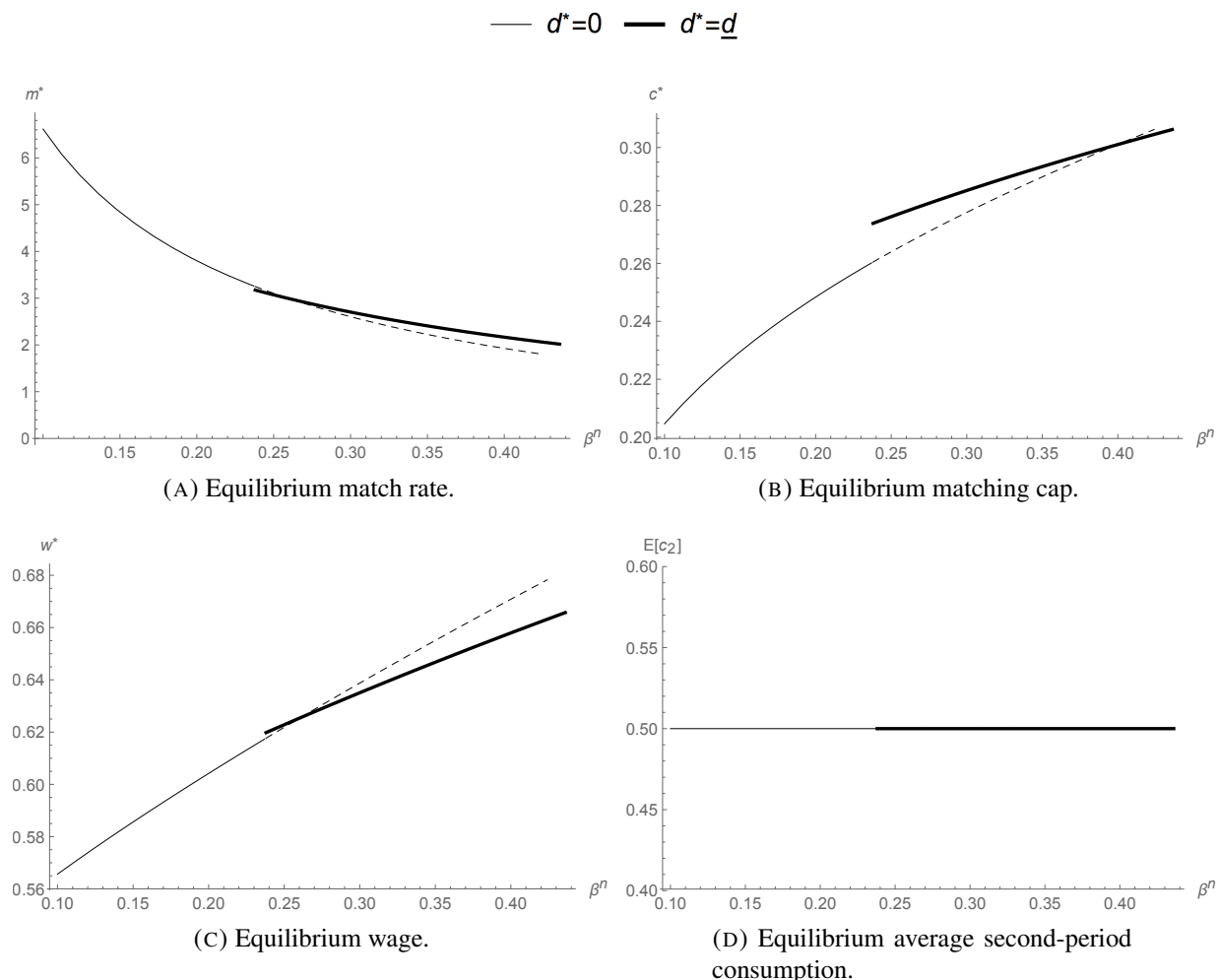(D) Equilibrium average second-period consumption.

FIGURE 1. Comparative statics on β^n with log utility. The other parameters are $\hat{\beta}^n = 1$, $\kappa^a = 0.6$, $\kappa^t = \kappa^p = 0.2$, $\kappa^r = 0.2$, $\kappa^n = 0.8$, $\underline{d} = 0.12$ and $\gamma = 1$. The thin line denotes $d^{rn} = 0$, the thick line denotes $d^{rn} = \underline{d}$, and the dashed line denotes outcome if d were exogenously set to 0.

parameter values both in the characterized equilibrium and in the equilibrium when the contract space is exogenously limited to contracts with $d = 0$. As a result, the adoption of automatic enrollment has no effect on average second-period consumption, relative to the outcome when all plans must be opt-in ($d = 0$). It does have an effect, however, on the *distribution* of savings. When $d^{rn} = \underline{d}$, advice takers save less, and procrastinators save more, relative to the outcome when all plans must have $d = 0$.

Constant consumption shares in each period is an artifact of log utility, which is a special case of CRRA utility with the coefficient of relative risk aversion (and the EIS) equal to 1. Most estimates
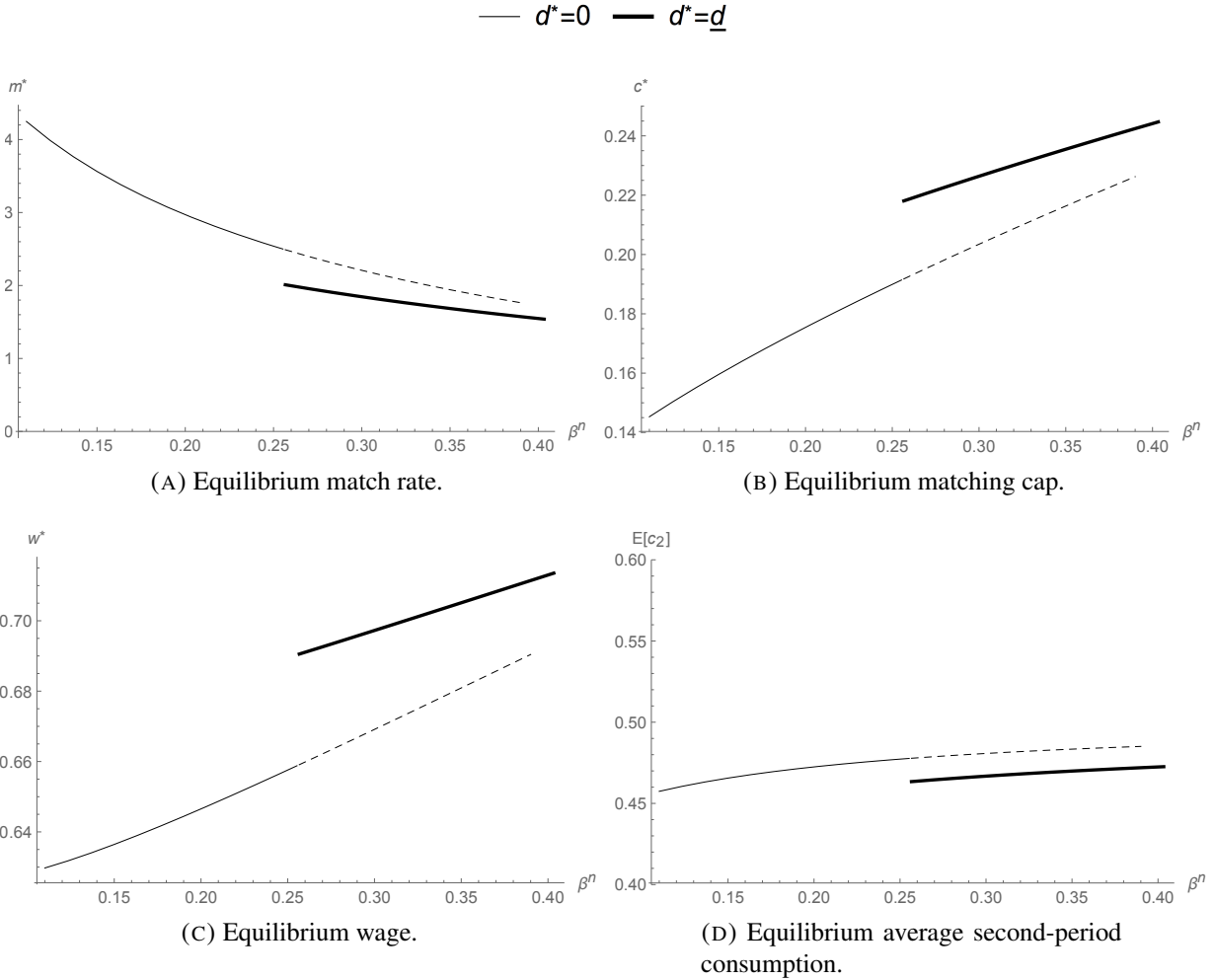
FIGURE 2. Comparative statics on β^n with CRRA utility with coefficient of relative risk aversion equal to 2.7. The other parameters are $\hat{\beta}^n = 1$, $\kappa^a = 0.8$, $\kappa^t = \kappa^p = 0.1$, $\kappa^r = 0.2$, $\kappa^n = 0.8$, $\underline{d} = 0.12$ and $\gamma = 1$. The thin line denotes $d^{rn} = 0$, the thick line denotes $d^{rn} = \underline{d}$, and the dashed line denotes outcome if d were exogenously set to 0.

of the EIS are much lower than 1. A recent estimation of intertemporal consumption preferences, for example, found a coefficient of relative risk aversion of 2.7, a β of 0.35, and an annualized exponential discount factor of 0.97 (Laibson, Maxted, Repetto, and Tobacman, 2015). Accordingly, Figure 2 shows the same set of comparative statics using CRRA utility with a coefficient of relative risk aversion equal to 2.7. The results on contract terms in panels (A) - (C) are similar to the results with log utility, but panel (D) shows that equilibrium average second-period consumption is increasing in β^n, which is a more intuitive result. Moreover, it shows that the adoption of automatic enrollment actually *lowers* retirement savings relative to if d were exogenously set to

0—the thick solid line is everywhere below the dashed line. While this result contrasts sharply with the existing literature advocating that employers automatically enroll employees in order to raise savings, it is consistent with other applications of behavioral contract theory, which show that equilibrium contracts maximize the "fictional surplus" measured by decision utility rather than the actual surplus measured by experienced utility (DellaVigna and Malmendier, 2004). For example, Baicker, Mullainathan, and Schwartzstein (2015) develop a behavioral contract theory model of health insurance focusing on contract terms that affect healthcare utilization and show that equilibrium health insurance plans are designed to minimize insurer costs rather than to maximize actual surplus.

5. EVIDENCE

The basic predictions of the model on the structure of plan contributions line up well with key facts on plan design. First, the vast majority of defined contribution plans—about 80%—offer employer matching contributions with a cap on matched savings (PSCA, 2011). Second, failure to receive the full match offered by the employer is indeed widespread, as implied by our theory. Choi, Laibson, and Madrian (2011) find that about 50% of employees do not save enough to receive their full employer match, foregoing on average 1.3% of their salary.

Third, most employers that have adopted automatic enrollment plans have chosen relatively low default contribution rates. Indeed, about three-quarters of automatic enrollment plans default workers into a 3% initial contribution rate or less (PSCA, 2011), which is the minimum initial contribution rate that qualifies for the safe harbor for such plans created by the PPA.[3] As a result, the adoption of automatic enrollment appears not to have increased overall retirement savings, contra the hopes of its advocates. While there has been no large-scale evaluation of the effect of the adoption of automatic enrollment on savings rates given the default contribution rates that employers actually choose, Choi, Laibson, Madrian, and Metrick (2004) show that the strength of the competing effects of automatic enrollment on average contribution rates vary from employer to employer and in some cases reduce savings. Furthermore, over the same period that the use of

---

[3]*See* Internal Revenue Code §401(k)(13)(C)(iii)(I).

automatic enrollment in plans administered by Vanguard doubled, the average total contribution rate of eligible employees fell from 7.9% to 7.5%, a fall that Vanguard attributes partly to "the growing use of automatic enrollment and the tendency of participants to stick with the default [contribution] rate adopted by the [employer]" (Vanguard, 2015, p.33).

Fourth, automatic enrollment plans almost always offer matching contributions, and the vast majority of those that do set a default contribution rate below the maximum amount of employee contributions that the employer matches (Beshears, Choi, Laibson, and Madrian, 2010). Even automatic enrollment plans that default workers into automatically increasing the employee contribution rate over time typically plateau at a maximum default below the employer's cap on matching contributions (Butrica and Karamcheva, 2012). These basic stylized facts about the structure of defaults relative to the structure of employer matching contributions are consistent with the equilibrium theory we have developed and inconsistent with a theory in which employers set defaults paternalistically, given the other terms of the contract.

6. Conclusion

Federal retirement savings policy has long been premised on the notion that, left to their own devices, households will make mistakes in saving for retirement (Kotlikoff, 1987). This paternalistic concern motivates both mandatory savings schemes like Social Security as well as incentive-based policy tools such as tax subsidies for retirement savings that together shape retirement savings in the United States. The special tax subsidies provided for employer-sponsored retirement savings plans amount to an attempt to harness employers to address this policy problem. As a result of these tax subsidies, each employer designs a microcosm of the broader federal policy regime through the mix of mandatory savings rules, savings incentives, default rules, and investment options they offer workers in their retirement savings plans.

Previous work in economics has considered the problems raised by mandating or subsidizing certain forms of employer benefits such as pensions and health insurance to achieve public policy goals. Summers (1989) argues, for example, that in the presence of wage rigidities, such policies can distort employment levels, in some cases disproportionately harming the very workers the

policy seeks to help. Similarly, the predominance of employer-provided health insurance, due in large part to its tax treatment, can cause an inefficient reduction in labor mobility ("job lock") (Gruber, 2000).

We identify a new type of dysfunction caused by attempts to use employers to implement social policy. We show that if workers are subject to behavioral biases that affect retirement savings decisions, then employers have incentives to cater to, rather than correct, those biases. Such biases generally imply a wedge between workers' decision utility at the time of contracting and their experienced utility that is the appropriate criterion for welfare analysis. The equilibrium in the labor market will produce plan designs that maximize the "fictional surplus" measured by workers' ex ante decision utility rather than the true surplus measured by workers' experienced utility. Our analysis thus calls into question the longstanding delegation to employers of the design of the primary tax-advantaged vehicle for retirement savings. If behavioral economists are right that workers make systematic mistakes in saving for retirement, then the labor market gives employers incentives that undermine the field's "public finance" approach to employer plan design.

While we focus in this paper on the rules that structure contributions to the plan, the same approach can be taken to other aspects of plan design. For example, in the Online Appendix we consider the set of investment options available within a retirement plan. We show that when employers contract with a mix of rational workers and naive diversifiers, in equilibrium each employer's plan offers a set of investment options that includes a low-fee option (for rationals) and higher-fee options (for naives) that are otherwise equivalent. Our approach could also be applied to other forms of employment benefits for which behavioral biases likely play an important role, such as health insurance, but we leave such an analysis for future work.

REFERENCES

BAICKER, K., S. MULLAINATHAN, AND J. SCHWARTZSTEIN (2015): "BEHAVIORAL HAZARD IN HEALTH INSURANCE," *The Quarterly Journal of Economics*, 1623, 1667.

BAR-GILL, O. (2012): *Seduction by Contract: Law, Economics, and Psychology in Consumer Markets*. Oxford University Press.

BENARTZI, S., AND R. H. THALER (2007): "Heuristics and biases in retirement savings behavior," *The journal of economic perspectives*, pp. 81–104.

BERNHEIM, B. D., A. FRADKIN, AND I. POPOV (2015): "The welfare economics of default options in 401 (k) plans," *The American Economic Review*, 105(9), 2798–2837.

BESHEARS, J., J. J. CHOI, D. LAIBSON, AND B. C. MADRIAN (2010): "the impact of employer matching on savings plan participation under automatic enrollment," in *Research Findings in the Economics of Aging*, pp. 311–327. University of Chicago Press.

BESHEARS, J., J. J. CHOI, D. LAIBSON, B. C. MADRIAN, AND B. WELLER (2010): "Public policy and saving for retirement: The "autosave" features of the Pension Protection Act of 2006," *Better living through economics*, pp. 274–290.

BUBB, R., AND R. H. PILDES (2014): "How Behavioral Economics Trims Its Sails and Why," *Harvard Law Review*, 127(6), 1637.

BUTRICA, B. A., AND N. S. KARAMCHEVA (2012): "Automatic Enrollment, Employee Compensation, and Retirement Security," *Center for Retirement Research at Boston College Working Paper*, (2012-25).

CARROLL, G. D., J. J. CHOI, D. LAIBSON, B. C. MADRIAN, AND A. METRICK (2009): "Optimal Defaults and Active Decisions," *The Quarterly Journal of Economics*, 124(4), 1639–1674.

CHOI, J. J., D. LAIBSON, AND B. C. MADRIAN (2011): "$100 bills on the sidewalk: Suboptimal investment in 401 (k) plans," *Review of Economics and Statistics*, 93(3), 748–763.

CHOI, J. J., D. LAIBSON, B. C. MADRIAN, AND A. METRICK (2004): "For Better or Worse: Default Effects and 401 (k) Savings Behavior," *Perspectives on the Economics of Aging*, p. 81.

DELLAVIGNA, S., AND U. MALMENDIER (2004): "Contract design and self-control: Theory and evidence," *The Quarterly Journal of Economics*, pp. 353–402.

——— (2006): "Paying not to go to the gym," *The American Economic Review*, pp. 694–719.

EDLIN, A. S., AND C. SHANNON (1998): "Strict monotonicity in comparative statics," *Journal of Economic Theory*, 81(1), 201–219.

GABAIX, X., AND D. LAIBSON (2006): "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *The Quarterly Journal of Economics*, pp. 505–540.

GRUBB, M. D. (2009): "Selling to overconfident consumers," *The American Economic Review*, 99(5), 1770–1807.

GRUBER, J. (2000): "Health insurance and the labor market," *Handbook of health economics*, 1, 645–706.

HEIDHUES, P., AND B. KŐSZEGI (2010): "Exploiting naivete about self-control in the credit market," *The American Economic Review*, pp. 2279–2303.

KOTLIKOFF, L. J. (1987): "Justifying public provision of social security," *Journal of Policy Analysis and Management*, 6(4), 674–696.

LAIBSON, D., P. MAXTED, A. REPETTO, AND J. TOBACMAN (2015): "Estimating Discount Functions with Consumption Choices over the Lifecycle," *Unpublished working paper*.

MADRIAN, B. C., AND D. F. SHEA (2001): "The Power of Suggestion: Inertia in 401 (k) Participation and Savings Behavior," *The Quarterly Journal of Economics*.

O'DONOGHUE, T., AND M. RABIN (2001): "Choice and procrastination," *Quarterly Journal of Economics*, pp. 121–160.

ORSZAG, P. R., J. M. IWRY, AND W. G. GALE (2006): *Aging Gracefully: Ideas to Improve Retirement Security in America*. Century Foundation.

PSCA (2011): *54th Annual Survey of Profit Sharing and 401(k) Plans*. Plan Sponsor Council of America.

ROTHSCHILD, M., AND J. STIGLITZ (1976): "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *The Quarterly Journal of Economics*, 90(4), 629–649.

SUMMERS, L. H. (1989): "Some simple economics of mandated benefits," *The American Economic Review*, 79(2), 177–183.

THALER, R. H., AND S. BENARTZI (2004): "Save more tomorrow™: Using behavioral economics to increase employee saving," *Journal of political Economy*, 112(S1), S164–S187.

VANGUARD (2015): "How America Saves 2015," Discussion paper.

APPENDIX

Proof of Proposition 1. The problem for workers with $\hat{\beta} = \beta$ simplifies to finding the zero-profit contracts that perfectly smooth her consumption across the two periods, which by concavity of the utility function is optimal. Any zero-profit contract with $r = w = \gamma/2$ achieves this and must be in the equilibrium set. Any zero-profit contract with $r > \gamma/2$ results in greater consumption in the second period than in the first, and therefore cannot be in the equilibrium set. Consider finally contracts with $r < \gamma/2$. The FOC for savings is $u'(w - s(w,r,m)) = \beta(1+m)u'(r+(1+m)s(w,r,m))$ and must be satisfied so long as $u'(w) \le \beta(1+m)u'(r)$. If this FOC is satisfied, then the only m that results in perfect consumption smoothing is $m^{FB} = \frac{1-\beta}{\beta}$. At $m = m^{FB}$, the FOC for savings is satisfied so long as $w \ge r$. Hence, any contract with $m = m^{FB}$, $w \ge r$, and that makes zero profits, which requires $w + r + m^{FB}\frac{w-r}{2+m^{FB}} = \gamma$, is also in the equilibrium. For sophisticated myopic workers, this completes the proof. For rational workers, note that $\beta = 1$ and hence $m^{FB} = 0$.

Proof of Proposition 2. Assume for now that the solution includes a contract in the portion of the zero-profit set in which the FOC for savings in period 1, given by $\beta(1+m)u'((1+m)s(w,r,m|\beta)+r) - u'(w - s(w,r,m|\beta)) = 0$, is satisfied. At the end of the proof we will show that this FOC is indeed satisfied in all equilibrium contracts. This FOC is satisfied iff $\beta(1 + m)u'(r) \ge u'(w)$, where a strict inequality implies $s(w,r,m|\beta) > 0$. Implicitly define the zero-profit wage as a function of other contract terms, $w(r,m)$, by $w + ms(w,r,m|\beta) + r - \gamma = 0$. We have by the implicit function theorem,

$$(7) \qquad \frac{\partial w(r,m)}{\partial r} = -\frac{1 + m\frac{\partial s(w,r,m|\beta)}{\partial r}}{1 + m\frac{\partial s(w,r,m|\beta)}{\partial w}} = -\frac{1}{1+m}.$$

Substituting for the wage in self 0's objective function using $w(r,m)$, we have,

$$(8) \qquad V(r,m|\hat{\beta}) \equiv u(w(r,m) - s(w(r,m),r,m|\hat{\beta})) + u(r + (1+m)s(w(r,m),r,m|\hat{\beta})).$$

Taking the partial derivative with respect to r, we have,

$$
\begin{aligned}
V_r =& [\frac{\partial w(r,m)}{\partial r} - (\frac{\partial w(r,m)}{\partial r}\frac{\partial s}{\partial w} + \frac{\partial s}{\partial r})]u'(w(r,m) - s(w(r,m),r,m|\hat{\beta})) + \\
& [1 + (1+m)(\frac{\partial w(r,m)}{\partial r}\frac{\partial s}{\partial w} + \frac{\partial s}{\partial r})]u'(r + (1+m)s(w(r,m),r,m|\hat{\beta})) = \\
& [1 + (-\frac{\partial s}{\partial w} + (1+m)\frac{\partial s}{\partial r})](1 - \hat{\beta})u'(r + (1+m)s(w(r,m),r,m|\hat{\beta})),
\end{aligned}
$$

(9)

where we have used the FOC for anticipated savings, which must also be satisfied. The leading term in brackets can be expressed as,

$$(10) \qquad [1 + ([\frac{-u''(w-s) - \hat{\beta}(1+m)^2 u''(s(1+m)+r)}{u''(w-s) + \hat{\beta}(1+m)^2 u''((1+m)s+r)}] = 0.$$

Hence, holding m fixed, self 0 is indifferent among all zero-profit contracts for which the FOC for savings is satisfied, including the one with $r = 0$. This implies that we can derive the optimal m in the subset of the contract space in which the FOC for savings is satisfied by considering contracts of the form $(w, 0, m)$. The FOCs for the Lagrangian of this restricted problem, together with the FOC for anticipated savings, imply the following condition, which pins down equilibrium m:

$$(11) \qquad \frac{s(\beta) + m^* \frac{\partial s(\beta)}{\partial m}}{1 + m^* \frac{\partial s(\beta)}{\partial w}} = \frac{s(\hat{\beta}) + (1-\hat{\beta})(1+m^*) \frac{\partial s(\hat{\beta})}{\partial m}}{(1+m^*)\hat{\beta} + (1-\hat{\beta})(1+m^*)\frac{\partial s(\hat{\beta})}{\partial w}},$$

where we suppress the contract arguments of the savings functions to economize on notation. The LHS of (11) is the slope of the zero-profit line in (m, w) space, while the RHS is the worker's MRS between m and w. Substituting $m^{FB} = \frac{1-\beta}{\beta}$ into (11) and simplifying yields,

$$(12) \qquad \frac{\beta s(\beta) + (1-\beta)\frac{\partial s(\beta)}{\partial m}}{\beta + (1-\beta)\frac{\partial s(\beta)}{\partial w}} = \frac{\beta s(\hat{\beta}) + (1-\hat{\beta})\frac{\partial s(\hat{\beta})}{\partial m}}{\hat{\beta} + (1-\hat{\beta})\frac{\partial s(\hat{\beta})}{\partial w}}.$$

This condition is satisfied when $\hat{\beta} = \beta$, so we have shown that $m^* = m^{FB}$ if $\hat{\beta} = \beta$. We now derive montone comparative statics on m^* with respect to $\hat{\beta}$ to characterize m^* for naive workers.

Substituting in for w using the zero-profit constraint, the problem becomes:

$$(13) \qquad \max_m V(m, \hat{\beta}) \equiv u\big(w(m) - s(\hat{\beta})\big) + u\big((1+m)s(\hat{\beta})\big),$$

where $w(m) = \gamma - ms(\beta)$. CRRA utility is given by $u(c) = \frac{1}{1-\theta}c^{1-\theta}$ where $\theta > 0$ and $\theta \neq 1$. The FOC for savings can be written as $s(\hat{\beta}) = w\sigma(m, \hat{\beta})$, where,

$$(14) \qquad \sigma(m, \hat{\beta}) \equiv \frac{[\hat{\beta}(1+m)]^{\frac{1}{\theta}}}{1 + m + [\hat{\beta}(1+m)]^{\frac{1}{\theta}}} = \frac{1}{1 + (1+m)^{1-\frac{1}{\theta}}\hat{\beta}^{-1/\theta}}.$$

With this notation $w = \gamma - mw\sigma(m, \beta)$, and so $w = \frac{\gamma}{1+m\sigma(m,\beta)}$. Substituting that back to the savings equation yields $s(\hat{\beta}) = \frac{\gamma\sigma(m,\hat{\beta})}{1+m\sigma(m,\beta)}$. Overall utility is then given by,

$$(15) \qquad \begin{aligned} V(m, \hat{\beta}) &= \frac{1}{1-\theta}\Big[[\frac{\gamma(1-\sigma(m,\hat{\beta}))}{1+m\sigma(m,\beta)}]^{1-\theta} + [\frac{\gamma(1+m)\sigma(m,\hat{\beta})}{1+m\sigma(m,\beta)}]^{1-\theta}\Big] \\ &= \frac{1}{1-\theta}[\frac{\gamma}{1+m\sigma(m,\beta)}]^{1-\theta}\Big[(1-\sigma(m,\hat{\beta}))^{1-\theta} + ((1+m)\sigma(m,\hat{\beta}))^{1-\theta}\Big]. \end{aligned}$$

We will now show that if $\theta < 1$ then $\frac{\partial m^*}{\partial \hat{\beta}} > 0$ and therefore $m^* > m^{FB}$. It will prove useful to work with a monotonic transformation of the objective function, $v(m, \hat{\beta}) \equiv log(V(m, \hat{\beta}))$, which represents the same preferences and hence generates the same choice function $m^*(\hat{\beta})$. It will be sufficient to show that $\frac{\partial^2 v(m,\hat{\beta})}{\partial m \partial \hat{\beta}} > 0$ (Edlin and Shannon, 1998). Taking logs, we have,

$$(16) \quad v(m, \hat{\beta}) = X - (1-\theta)log(1+m\sigma(m,\beta)) + log\Big[(1-\sigma(m,\hat{\beta}))^{1-\theta} + ((1+m)\sigma(m,\hat{\beta}))^{1-\theta}\Big],$$

where X is a constant that does not depend on m or $\hat{\beta}$.

A straightforward calculation yields the partial derivative with respect to $\hat{\beta}$:

$$(17) \qquad \frac{\partial v(m, \hat{\beta})}{\partial \hat{\beta}} = \frac{(1-\theta)}{\theta \hat{\beta}^{1+1/\theta}} \frac{\sigma^2}{1-\sigma} \left[\frac{1}{\sigma + \frac{\hat{\beta}}{1+\hat{\beta}}} \right].$$

We want to show that this function is strictly increasing in m for $\theta < 1$. When $\theta < 1$, all three terms are positive. Furthermore, the leading term is independent of m. Thus, when $\theta < 1$ the function increases in m whenever the product of the last two terms increases in m. To show that the product of the last two terms increases in m, it is sufficient to show that an increasing transformation of the last two terms increases in m. Taking the partial derivative with respect to m of the log of that product, we have,

$$(18) \qquad \frac{\partial}{\partial m} log \left(\frac{\sigma^2}{1-\sigma} \left[\frac{1}{\sigma + \frac{\hat{\beta}}{1+\hat{\beta}}} \right] \right) = \left[\frac{\frac{\hat{\beta}}{1+\hat{\beta}}(2-\sigma) + \sigma}{\sigma(1-\sigma)(\sigma + \frac{\hat{\beta}}{1+\hat{\beta}})} \right] \frac{1-\theta}{\theta}(1+m)^{-1/\theta}\sigma^2,$$

which is strictly positive whenever $\theta < 1$.

We now want to show that if $\theta > 1$ then $\frac{\partial m^*}{\partial \hat{\beta}} < 0$ and therefore $m^* < m^{FB}$. It is sufficient to show that $\frac{\partial^2 v(m, \hat{\beta})}{\partial m \partial \hat{\beta}} < 0$, which follows using the same basic argument above but noting that for $\theta > 1$, the leading constant in (17) is negative.

All that is left to show is that if $\theta = 1$ then $m^* = m^{FB}$. CRRA utility with $\theta = 1$ is given by $u(c_i) = log(c_i)$. With log utility, savings is simply $s(w|\beta) = w\frac{\beta}{1+\beta}$, independent of m. Thus, $\partial s/\partial m = 0$ and $\partial s/\partial w = \frac{\beta}{1+\beta}$. Furthermore, zero-profit wages satisfy $w^*(m) = \gamma - mw\frac{\beta}{1+\beta} = \frac{\gamma}{1+m\frac{\beta}{1+\beta}}$. Plugging these into the condition for preferring first-best matching (12), it becomes

$$(19) \qquad \frac{\beta \frac{\beta}{1+\beta} w^*}{\beta + (1-\beta)\frac{\beta}{1+\beta}} = \frac{\beta \frac{\hat{\beta}}{1+\hat{\beta}} w^*}{\hat{\beta} + (1-\hat{\beta})\frac{\hat{\beta}}{1+\hat{\beta}}},$$

which simplifies to $1/2 = 1/2$, independent of both β and $\hat{\beta}$. This shows that having CRRA utility with $\theta = 1$ is sufficient to produce $m^* = m^{FB}$ for all values of $\hat{\beta}$.

All that is left to show is that all equilibrium contracts must be such that the FOC for savings is satisfied. Suppose not. For an equilibrium contract for which this FOC is *not* satisfied, we have $s(w, r, m|\beta) = 0$ and $\partial w/\partial r = -1$. Suppose the FOC for anticipated savings is satisfied. We have,

$$V_r = \left[\frac{\partial w(r,m)}{\partial r} - \left(\frac{\partial w(r,m)}{\partial r}\frac{\partial s}{\partial w} + \frac{\partial s}{\partial r} \right) \right] u'(w(r,m) - s(w(r,m), r, m|\hat{\beta})) +$$

$$(20) \qquad \left[1 + (1+m)\left(\frac{\partial w(r,m)}{\partial r}\frac{\partial s}{\partial w} + \frac{\partial s}{\partial r} \right) \right] u'(r + (1+m)s(w(r,m), r, m|\hat{\beta})) =$$

$$\left[1 - \hat{\beta} - \hat{\beta}m + (1+m)(1-\hat{\beta})\left(\frac{\partial s}{\partial r} - \frac{\partial s}{\partial w} \right) \right] u'(r + (1+m)s(w(r,m), r, m|\hat{\beta})).$$

The leading term in brackets can be expressed as,

$$(21) \qquad 1 - \hat{\beta} - \hat{\beta}m - (1-\hat{\beta})\left[\frac{mu''(w-s)}{u''(w-s) + \hat{\beta}(1+m)^2 u''((1+m)s+r)} + 1 \right] < 0.$$

24

This implies that the supposed optimal contract is dominated by other contracts in the constraint set with a smaller r, which is a contradiction.

Finally, suppose there is a contract in the solution for which the FOCs for savings and for anticipated savings are not satisfied. In that case, $w = \gamma - r$ and the worker's problem is simply $\max_r u(\gamma - r) + u(r)$, which is maximised at $r = w = \gamma/2$, which yields utility of $2u(\gamma/2)$. The zero-profit contract using m^{FB}, $(\frac{\gamma(1+\beta)}{2}, 0, \frac{1-\beta}{\beta})$, also results in consumption $\gamma/2$ in each period. The sophisticate anticipates this consumption sequence under both the $(\frac{\gamma(1+\beta)}{2}, 0, \frac{1-\beta}{\beta})$ contract and the $(\gamma/2, \gamma/2, 0)$ contract and so is indifferent between them. Note that a myopic worker's period-0 anticipated payoff from the $(\gamma/2, \gamma/2, 0)$ contract is independent of $\hat{\beta}$. In contrast, her anticipated payoff from the $(\frac{\gamma(1+\beta)}{2}, 0, \frac{1-\beta}{\beta})$ contract is strictly increasing in $\hat{\beta}$. To show this last point, we can differentiate the period-0 objective function with respect to $\hat{\beta}$, yielding

$$\frac{\partial}{\partial \hat{\beta}}[u(w - s(w, r, m|\hat{\beta})) + u((1+m)s(w, r, m|\hat{\beta}))]$$

(22)
$$= [(1+m)u'((1+m)s(w, r, m|\hat{\beta})) - u'(w - s(w, r, m|\hat{\beta}))]\frac{\partial s(w, r, m|\hat{\beta})}{\partial \hat{\beta}}$$

$$= u'((1+m)s(w, r, m|\hat{\beta}))(1+m)(1 - \hat{\beta})\frac{\partial s(w, r, m|\hat{\beta})}{\partial \hat{\beta}} > 0,$$

where the inequality follows from the fact that $\frac{\partial s(w, r, m|\hat{\beta})}{\partial \hat{\beta}} > 0$. This implies that naive workers, who have $\hat{\beta} > \beta$, strictly prefer the matching contract $(\frac{\gamma(1+\beta)}{2}, 0, \frac{1-\beta}{\beta})$ to the $(\gamma/2, \gamma/2, 0)$ contract. This implies that all equilibrium contracts must be such that the FOC for savings is satisfied. $\qquad\square$

Proof of Proposition 3.

Lemma 1. *Consider the set of all contracts that make nonnegative profits when chosen by only the naives and rationals. The naives' preferred contracts in that set are matching contracts with a matching cap set at the naive's anticipated savings level in which rationals will save exactly to the cap but naives will save strictly less than the cap. They use $m < m^{FB}$ and deliver to both naives and rationals identical utility exceeding $2\ln(\gamma/2)$.*

Proof. For this proof, to save on notation, normalize so that $\kappa^r + \kappa^n = 1$. The naives' preferred contracts must make zero profits since otherwise the contract's wage could be increased to make naives strictly better off while still making nonnegative profits. Hence we can restrict attention to contracts that would make zero profits when chosen by only naives and rationals.

Assume for now that any contract in the constraint set that delivers the highest utility to naives—which we will refer to as an "optimal contract" for short—is a contract in which naive workers anticipate receiving matching contributions (a "matching contract"), so that $m \min(s(w, r, m, c|\hat{\beta}^n), c) > 0$. This requires that $w\hat{\beta}^n > r/(1+m)$ and $m > 0$. We will show later that naives' payoff from the best matching contacts exceeds $2\ln(\gamma/2)$. Since $2\ln(\gamma/2)$ is the highest payoff possible to any type in a non-matching contract, the optimal contract must be a matching contract.

Let $s(w, r, m, c|\beta)$ represent savings by a type-β worker. If c is small enough, the worker will save above the cap. In that case, savings satisfies $u'(w - s) = \beta u'(s + mc + r)$, which implies savings of $\frac{\beta w - mc - r}{1+\beta}$, which is, in fact, above c whenever $c < \frac{w\beta - r}{1+\beta+m}$. At the other extreme, c is large enough it will not bind and savings satisfies $u'(w - s) = (1+m)\beta u'(s(1+m) + r)$, which

implies savings of $\max\{\frac{\beta w - r/(1+m)}{1+\beta}, 0\}$, which must be below c. Between these cutoffs, the agent will simply save to the cap. In summary,

$$(23) \qquad s(w,r,m,c|\beta) = \begin{cases} \frac{\beta w - mc - r}{1+\beta} & \text{, if } c < \frac{w\beta - r}{1+\beta+m} \\ c & \text{, if } \frac{w\beta - r}{1+\beta+m} \leq c \leq \frac{\beta w - r/(1+m)}{1+\beta} \\ \max\{\frac{\beta w - r/(1+m)}{1+\beta}, 0\} & \text{, if } \frac{\beta w - r/(1+m)}{1+\beta} < c. \end{cases}$$

Denote these cutoff matching caps, $\underline{c}(\beta) = \frac{w\beta - r}{1+\beta+m}$ and $\overline{c}(\beta) = \frac{\beta w - r/(1+m)}{1+\beta}$. Note that $\underline{c}(\hat{\beta}) > \overline{c}(\beta)$ iff $\hat{\beta} > (1+m)\beta$.

The constrained optimization problem is:

$$(24) \qquad \max_{w,r,m,c} V(w,r,m,c) = \ln(w - s(w,r,m,c|\hat{\beta}^n))$$
$$+ \ln(s(w,r,m,c|\hat{\beta}^n) + m\min(s(w,r,m,c|\hat{\beta}^n),c) + r),$$

subject to,

$$(25) \qquad w = \gamma - \bar{s}(w,r,m,c)m - r,$$

where $\bar{s}(w,r,m,c) = \kappa^r \min\{c, s(w,r,m,c|1)\} + \kappa^n \min\{c, s(w,r,m,c|\beta^n)\}$ is the average matched savings.

Implicitly define the zero-profit wage as a function of the other contract parameters,

$$(26) \qquad w^*(r,m,c) = \gamma - \bar{s}(w^*(r,m,c),r,m,c)m - r.$$

The optimization problem can now be rewritten as,

$$(27) \qquad \max_{r,m,c} V(r,m,c) = \ln(w^* - s(w^*,r,m,c|\hat{\beta}^n))$$
$$+ \ln(s(w^*,r,m,c|\hat{\beta}^n) + m\min(s(w^*,r,m,c|\hat{\beta}^n),c) + r),$$

where we suppress the dependence of w^* on r, m, and c to economize on notation.

Taking r and m as given, consider what c is optimal. We will denote the c that maximizes the objective function given r and m as $c^*(r,m)$. Replacing for s, the objective function becomes,

$$(28)$$

$$V(r,m,c) = \begin{cases} \ln(w^* - \frac{\hat{\beta}^n w^* - mc - r}{1+\hat{\beta}^n}) + \ln(\frac{\hat{\beta}^n w^* - mc - r}{1+\hat{\beta}^n} + mc + r), & \text{if } c < \underline{c}(\hat{\beta}^n) \\ \ln(w^* - c) + \ln(c(1+m) + r), & \text{if } \underline{c}(\hat{\beta}^n) \leq c \leq \overline{c}(\hat{\beta}^n) \\ \ln(w^* - \frac{\hat{\beta}^n w^* - r/(1+m)}{1+\hat{\beta}^n}) + \ln((1+m)\frac{\hat{\beta}^n w^* - r/(1+m)}{1+\hat{\beta}^n} + r), & \text{if } \overline{c}(\hat{\beta}^n) < c, \end{cases}$$

where we have used the fact that $w^*\hat{\beta}^n > r/(1+m)$, since we have assumed the optimal contracts are matching contracts, to eliminate the max from the savings function in the third piece of the objective function.

The partial derivative with respect to c is,

$$(29) \qquad V_c(r,m,c) = \begin{cases} 2\frac{\frac{\partial w^*}{\partial c} + m}{w^* + mc + r} & \text{, if } c < \underline{c}(\hat{\beta}^n) \\ \frac{\frac{\partial w^*}{\partial c} - 1}{w^* - c} + \frac{1+m}{c(1+m) + r} & \text{, if } \underline{c}(\hat{\beta}^n) < c < \overline{c}(\hat{\beta}^n) \\ 2\frac{\frac{\partial w^*}{\partial c}}{w^* + r/(1+m)}, & \text{if } \overline{c}(\hat{\beta}^n) < c \end{cases}$$

By the implicit function theorem,

$$
(30) \qquad \frac{\partial w^*}{\partial c} = -\frac{m\frac{\partial \bar{s}}{\partial c}}{1 + m\frac{\partial \bar{s}}{\partial w}} = \begin{cases} -m, & \text{if } c < \bar{c}(\beta^n) \\ -mK, & \text{if } \bar{c}(\beta^n) < c < \bar{c}(1) \\ 0, & \text{if } \bar{c}(1) < c, \end{cases}
$$

where $K \equiv \frac{\kappa^r}{1 + m[\kappa^n \frac{\beta^n}{1+\beta^n} \mathbf{1}(\beta^n w^* > r/(1+m))]}$, with $\frac{\kappa^r}{1+m} < K \le \kappa^r$.

Could $c > \bar{c}(\hat{\beta}^n)$ be optimal? Replacing for $\frac{\partial w^*}{\partial c}$ from (30), the partial derivative of the objective function in this range is,

$$
(31) \qquad V_c(r, m, c) = \begin{cases} -\frac{2mK}{w^* + \frac{r}{1+m}} & , \text{if } \bar{c}(\hat{\beta}^n) < c < \bar{c}(1) \\ 0 & , \text{if } \bar{c}(1) < c. \end{cases}
$$

Thus, $c = \bar{c}(\hat{\beta}^n)$ is weakly preferred to all higher caps, and the preference is strict for $\hat{\beta}^n < 1$.

Could $c < \underline{c}(\hat{\beta}^n)$ be optimal? The derivative in this range depends on the relation between $\underline{c}(\hat{\beta}^n)$ and $\bar{c}(\beta^n)$:

$$
(32) \qquad V_c(r, m, c) = \begin{cases} 0 & , \text{if } c < \min\{\underline{c}(\hat{\beta}^n), \bar{c}(\beta^n)\} \\ \frac{2m}{w^* + mc + r}(1 - K) & , \text{if } \bar{c}(\beta^n) < c < \underline{c}(\hat{\beta}^n) \end{cases}
$$

Thus, a cap at $c = \underline{c}(\hat{\beta}^n)$ is weakly preferred to any lower cap and the preference is strict if $\bar{c}(\beta^n) < \underline{c}(\hat{\beta}^n)$.

It is left to consider caps between $\underline{c}(\hat{\beta}^n)$ and $\bar{c}(\hat{\beta}^n)$. Replacing for $\frac{\partial w^*}{\partial c}$ in the middle range of equation (29),

$$
(33) \qquad V_c(r, m, c) = \begin{cases} (1 + m)[\frac{1}{c(1+m)+r} - \frac{1}{w^* - c}] & , \text{if } \underline{c}(\hat{\beta}^n) < c < \bar{c}(\beta^n) \\ \frac{(1+m)}{c(1+m)+r} - \frac{1+mK}{w^* - c} & , \text{if } \max\{\bar{c}(\beta^n), \underline{c}(\hat{\beta}^n)\} < c < \bar{c}(\hat{\beta}^n). \end{cases}
$$

Consider the ends of this $[\underline{c}(\hat{\beta}^n), \bar{c}(\hat{\beta}^n)]$ range, where the naive anticipates saving exactly to the cap. At that lowest cap, the return to increasing the cap is given by the right-hand derivative,

$$(34)$$

$$
V_{c+}(r, m, \underline{c}(\hat{\beta}^n)) = \begin{cases} \frac{(1+m)(1-\hat{\beta}^n)[w^*(1+m)+r]}{(1+\hat{\beta}^n + m)(\underline{c}(\hat{\beta}^n)(1+m)+r)(w^* - \underline{c}(\hat{\beta}^n))} & , \text{if } \underline{c}(\hat{\beta}^n) < \bar{c}(\beta^n) \\ \frac{(1+m)(1-\hat{\beta}^n)[w^*(1+m)+r]}{(1+\hat{\beta}^n + m)(\underline{c}(\hat{\beta}^n)(1+m)+r)(w^* - \underline{c}(\hat{\beta}^n))} + \frac{m}{w^* - \underline{c}(\hat{\beta}^n)}(1 - K) & , \text{if } \bar{c}(\beta^n) < \underline{c}(\hat{\beta}^n). \end{cases}
$$

If $\bar{c}(\beta^n) < \underline{c}(\hat{\beta}^n)$, or if $\hat{\beta}^n < 1$, then this derivative is strictly positive and $c^*(r, m) > \underline{c}(\hat{\beta}^n)$. Otherwise, this derivative is zero.

Now suppose that $\bar{c}(\beta^n) \ge \underline{c}(\hat{\beta}^n)$ and consider the left-hand and right-hand derivatives at $\bar{c}(\beta^n)$. From the left,

$$
\begin{aligned}
(35) \qquad V_{c-}(r, m, \bar{c}(\beta^n)) &= \frac{-m - 1}{w^* - \frac{\beta^n w^* - r/(1+m)}{1+\beta^n}} + \frac{1 + m}{(1 + m)\frac{\beta^n w^* - r/(1+m)}{1+\beta^n} + r} \\
&= [(w^* + r/(1 + m))\frac{\beta^n}{1 + \beta^n}]^{-1}[1 - \beta^n(1 + m)],
\end{aligned}
$$

while from the right,

$$V_{c+}(r, m, \bar{c}(\beta^n)) = \frac{-mK - 1}{w^* - \frac{\beta^n w^* - r/(1+m)}{1+\beta^n}} + \frac{1+m}{(1+m)\frac{\beta^n w^* - r/(1+m)}{1+\beta^n} + r}$$

(36)

$$= [(w^* + r/(1+m))\frac{\beta^n}{1+\beta^n}]^{-1}[1 - \beta^n(1 + mK)].$$

The signs of these derivatives are ambiguous. There are two key cases to consider. First, if $m \geq \frac{\frac{1}{\beta^n} - 1}{K}$ ("big m") then $V_{c-}(r, m, \bar{c}(\beta^n)) < 0$ and $V_{c+}(r, m, \bar{c}(\beta^n)) \leq 0$, which implies that $c^*(r, m) < \bar{c}(\beta^n)$ so that both naive and rational workers will actually save enough to receive the full match in the contract. Consider the highest utility achievable in this big m case.

If $\hat{\beta}^n < 1$, any $c^*(r, m) > 0$ must satisfy the first-order condition from equating the first line of equation (33) to zero, which implies that $c^*(r, m) = \max\{\frac{w^* - r}{2+m}, 0\}$. If $r > w^*$, then this is not a matching contract and therefore utility cannot exceed $2\ln(\gamma/2)$. Otherwise, $c^*(r, m) = \frac{w^* - r}{2+m}$, and by zero profits, $w^* = \frac{\gamma}{2}(\frac{2+m}{1+m}) - r$. $r < w^*$ requires $r < \frac{\gamma}{4}(\frac{2+m}{1+m})$. All such contracts deliver anticipated consumption of $c_1 = c_2 = \gamma/2$ and utility of $2\ln(\gamma/2)$.

If $\hat{\beta}^n = 1$, then the derivatives of the objective function above imply that $c^*(r, m)$ includes any $c < \underline{c}(1) = \frac{w^* - r}{2+m}$. If $r > w^*$, then this is not a matching contract and therefore utility cannot exceed $2\ln(\gamma/2)$. Otherwise, $c^*(r, m)(1+m) + r < \gamma/2$ but the naive worker anticipates savings greater than $c^*(r, m)$ in order to equate consumption in the two periods, resulting in utility of $2\ln(\gamma/2)$.

Consider for the rest of the proof the second key case, $m < \frac{\frac{1}{\beta^n} - 1}{K}$ ("small m"), which implies that $V_{c+}(r, m, \bar{c}(\beta^n)) > 0$. Observe that for any (r, m) such that $c^*(r, m) \leq \bar{c}(\beta^n)$, the best utility that can be achieved is $2\ln(\gamma/2)$. We will show below that there exist (r, m) such that $c^*(r, m) > \bar{c}(\beta^n)$ and that delivers utility strictly greater than $2\ln(\gamma/2)$.

For caps in the interval $(\max\{\bar{c}(\beta^n), \underline{c}(\hat{\beta}^n)\}, \bar{c}(\hat{\beta}^n))$, the derivative of the objective function with respect to c, given by (33), is at first positive and decreasing in c. The left-hand derivative at the top of that interval is,

$$V_{c-}(r, m, \bar{c}(\hat{\beta}^n)) = \frac{-1 - mK}{w^* - \frac{\hat{\beta}^n w^* - r/(1+m)}{1+\hat{\beta}^n}} + \frac{1+m}{(1+m)\frac{\hat{\beta}^n w^* - r/(1+m)}{1+\hat{\beta}^n} + r}$$

(37)

$$= [(w^* + r/(1+m))\frac{\hat{\beta}^n}{1+\hat{\beta}^n}]^{-1}[1 - \hat{\beta}^n(1 + mK)].$$

This derivative is negative if $m > \frac{\frac{1}{\hat{\beta}^n} - 1}{K} \equiv \bar{m}$. In particular, it is negative in any matching contract with $\hat{\beta}^n = 1$, which together with (31) implies that $c^*(r, m) \leq \bar{c}(\hat{\beta}^n)$ for all $\hat{\beta}^n$. If this derivative is negative, then any $c^*(r, m) > 0$ satisfies the first-order condition from equating the second line of equation (33) to zero, which implies that $c^*(r, m) = \max(\frac{w^* - \frac{r}{1+m}(1+mK)}{2+mK}, 0)$. If $c = 0$, then the contract is not a matching contract. Since we have assumed the solution is a matching contract, that implies that $c^*(r, m) = \frac{w^* - \frac{r}{1+m}(1+mK)}{2+mK}$.

Finally, if $m \leq \bar{m}$, so that this derivative is positive, then $c^*(r, m) = \bar{c}(\hat{\beta}^n)$.

Substituting $c^*(r, m)$ into the objective function, the problem becomes,

(38)
$$\max_{r,m} V(r, m) = \ln(w^*(r, m) - c^*(r, m))$$
$$+ \ln(c^*(r, m)(1 + m) + r),$$

where $w^*(r, m) \equiv w^*(r, m, c^*(r, m))$. Denote the set of r that maximizes this objective function given m by the set-valued function $r^*(m)$.

Consider the highest utility achievable in each of these small m cases. First, if $m > \bar{m}$ so that $c^*(r, m)$ is in the interior, then the optimization problem becomes,

(39)
$$\max_{r,m} V(r, m) = 2 \ln(w^*(r, m) + \frac{r}{1+m}) + \ln(\frac{(1+m)(1+mK)}{(2+mK)^2}).$$

The third line of (23) gives us that $s(w^*, r, m, c^*|\beta^n) = (\frac{\beta^n w^* - \frac{r}{1+m}}{1+\beta^n}) \mathbf{1}(w^*\beta^n > \frac{r}{1+m})$. Substituting for c^* and s into the implicit definition of $w^*(r, m)$ and solving for w^* yields an explicit definition,

(40)
$$w^*(r, m) = \frac{\gamma - \frac{r}{1+m}\left(1 + m - m[\kappa^r(\frac{1+mK}{2+mK}) + \kappa^n(\frac{\mathbf{1}(w^*\beta^n > \frac{r}{1+m})}{1+\beta^n})]\right)}{1 + m[\kappa^r(\frac{1}{2+mK}) + \kappa^n(\mathbf{1}(w^*\beta^n > \frac{r}{1+m})\frac{\beta^n}{1+\beta^n})]}$$

$$= \begin{cases} \frac{\gamma}{1+m[\kappa^r(\frac{1}{2+mK})+\kappa^n(\frac{\beta^n}{1+\beta^n})]} - \frac{r}{1+m}, & \text{if } \frac{\gamma\beta^n}{1+m[\kappa^r(\frac{1}{2+mK})+\kappa^n(\frac{\beta^n}{1+\beta^n})]} > \frac{r(1+\beta^n)}{1+m} \\ \frac{\gamma - \frac{m}{1+m}\kappa^n r}{1+\frac{m\kappa^r}{2+m\kappa^r}} - \frac{r}{1+m}, & \text{otherwise.} \end{cases}$$

Replacing for this wage, the objective function becomes,

(41)
$$V(r, m) = \begin{cases} 2\ln(\frac{\gamma}{1+m[\kappa^r(\frac{1}{2+mK})+\kappa^n(\frac{\beta^n}{1+\beta^n})]}) + \ln(\frac{(1+m)(1+mK)}{(2+mK)^2}), & \text{if } \frac{\gamma\beta^n}{1+m[\kappa^r(\frac{1}{2+mK})+\kappa^n(\frac{\beta^n}{1+\beta^n})]} > \frac{r(1+\beta^n)}{1+m} \\ 2\ln(\frac{\gamma - \frac{m}{1+m}\kappa^n r}{1+\frac{m\kappa^r}{2+m\kappa^r}}) + \ln(\frac{(1+m)(1+mK)}{(2+mK)^2}), & \text{otherwise,} \end{cases}$$

where we can see by inspection that the naive's anticipated payoff is independent of r, unless r get so high that the myope quits savings, where declines in r. Thus, for interior c^*, any r satisfying $r \leq w^*(1 + m)\beta^n$ is consistent with maximization.

What is the optimal m in this case? We will maximize the first line in (41) with respect to m, but first it is convenient to simplify:

(42)
$$V(r^*, m) = 2\ln(\gamma) + \ln\left[\frac{(1+m)(1+mK)}{[2+mK+m\kappa^r+m\kappa^n[2+mK]\frac{\beta^n}{1+\beta^n}]^2}\right]$$
$$= 2\ln(\gamma/2) + \ln\left[\frac{1+m}{[1+m\kappa^n\frac{\beta^n}{1+\beta^n}][1+m(\kappa^r+\kappa^n\frac{\beta^n}{1+\beta^n})]}\right].$$

If an optimal contract includes an interior cap, then it also includes a match that satisfies,

(43)
$$\frac{1}{1+m} - \frac{\kappa^n\frac{\beta^n}{1+\beta^n}}{1+m\kappa^n\frac{\beta^n}{1+\beta^n}} - \frac{\kappa^r+\kappa^n\frac{\beta^n}{1+\beta^n}}{1+m[\kappa^r+\kappa^n\frac{\beta^n}{1+\beta^n}]} = 0.$$

This FOC results in $m = \sqrt{\frac{1+\beta^n\kappa^r}{\beta^n(\beta^n+\kappa^r)}} - 1 > 0$, which is strictly less than $m^{FB} = \frac{1-\beta^n}{\beta^n}$ and therefore also strictly less than $\frac{\frac{1}{\beta^n}-1}{K}$. If this solution exceeds the bottom of the range of m under

29

consideration, $\frac{\frac{1}{\hat{\beta}^n}-1}{K}$, then it is indeed the optimal m in that range. If not, then the payoff decreases in m throughout this range.

Consider now the case when $m \leq \bar{m}$, so that $c^*(r,m) = \bar{c}(\hat{\beta}^n) = \frac{\hat{\beta}^n w^* - r/(1+m)}{1+\hat{\beta}^n}$,

(44)
$$V(r,m) = 2\ln(w^*(r,m) + \frac{r}{1+m}) + \ln(1+m) + \ln(\frac{\hat{\beta}^n}{(1+\hat{\beta}^n)^2}),$$

and,

(45)
$$w^*(r,m) = \frac{\gamma - \frac{r}{1+m}\left(1+m-m[\frac{\kappa^r}{1+\hat{\beta}^n} + \frac{\kappa^n}{1+\beta^n}\mathbf{1}(w^*(r,m)\beta^n > \frac{r}{1+m})]\right)}{1+m[\frac{\kappa^r\hat{\beta}^n}{1+\hat{\beta}^n} + \frac{\kappa^n\beta^n}{1+\beta^n}\mathbf{1}(w^*(r,m)\beta^n > \frac{r}{1+m})]}$$

$$= \begin{cases} \frac{\gamma}{1+m[\frac{\kappa^r\hat{\beta}^n}{1+\hat{\beta}^n} + \frac{\kappa^n\beta^n}{1+\beta^n}]} - \frac{r}{1+m}, & \text{if } \frac{\gamma\beta^n}{1+m[\frac{\kappa^r\hat{\beta}^n}{1+\hat{\beta}^n} + \frac{\kappa^n\beta^n}{1+\beta^n}]} > \frac{r(1+\beta^n)}{1+m} \\ \frac{\gamma - \frac{m}{1+m}\kappa^n r}{1+m\frac{\kappa^r\hat{\beta}^n}{1+\hat{\beta}^n}} - \frac{r}{1+m}, & \text{otherwise.} \end{cases}$$

Replacing for this wage,
(46)
$$V(r,m) = \begin{cases} 2\ln(\frac{\gamma}{1+m[\frac{\kappa^r\hat{\beta}^n}{1+\hat{\beta}^n} + \frac{\kappa^n\beta^n}{1+\beta^n}]}) + \ln(1+m) + \ln(\frac{\hat{\beta}^n}{(1+\hat{\beta}^n)^2}), & \text{if } \frac{\gamma\beta^n}{1+m[\frac{\kappa^r\hat{\beta}^n}{1+\hat{\beta}^n} + \frac{\kappa^n\beta^n}{1+\beta^n}]} > \frac{r(1+\beta^n)}{1+m} \\ 2\ln(\frac{\gamma - \frac{m}{1+m}\kappa^n r}{1+m\frac{\kappa^r\hat{\beta}^n}{1+\hat{\beta}^n}}) + \ln(1+m) + \ln(\frac{\hat{\beta}^n}{(1+\hat{\beta}^n)^2}), & \text{otherwise.} \end{cases}$$

Again, in this case, the payoff is independent of r and then decreases in r once the myope stops saving.

What is the optimal m in this range? It is convenient to define $\bar{\sigma} \equiv [\kappa^r\frac{\hat{\beta}^n}{1+\hat{\beta}^n} + \kappa^n\frac{\beta^n}{1+\beta^n}]$, the average savings *rate* in the contract if $r = 0$. The first-order condition, defined by setting the partial derivative of the first line of (46) with respect to m to zero, can be written as,

(47)
$$\frac{1}{1+m} - \frac{2\bar{\sigma}}{1+m\bar{\sigma}} = 0,$$

which reduces to $m = \frac{1-2\bar{\sigma}}{\bar{\sigma}} < m^{FB} < \frac{m^{FB}}{K}$. If this solution exceeds the top of the range of m under consideration, $\frac{\frac{1}{\hat{\beta}^n}-1}{K}$, then it is indeed the optimal m in that range. If not, then the optimal m in that range is equal to $\frac{\frac{1}{\hat{\beta}^n}-1}{K}$.

We have now shown that in each of these two subcases of the small m case, the optimal m is less than m^{FB}. We also showed above that the highest utility achievable in the big m case is $2\ln(\gamma/2)$. We now show that the utility achievable in the small m case is strictly greater than $2\ln(\gamma/2)$. To see this, consider the contract with $c = \bar{c}(\hat{\beta}^n)$, $r = 0$, $m = \frac{1-2\bar{\sigma}}{\bar{\sigma}}$, and $w = w^*(r,m,c)$. This contract yields the naives utility of,

(48)
$$V(r,m) = 2\ln(\gamma/2) + \ln(\frac{\hat{\beta}^n}{(1+\hat{\beta}^n)^2}) - \ln(\bar{\sigma}(1-\bar{\sigma})) > 2\ln(\gamma/2),$$

where the inequality comes from the fact that $x(1-x)$ is increasing when $x < 1/2$ and $\bar{\sigma} < \frac{\hat{\beta}^n}{1+\hat{\beta}^n} < 1/2$. Thus, any optimal contract must be in the small m case and must have $m < m^{FB}$,

$c > \max(\underline{c}(\hat{\beta}^n), \bar{c}(\hat{\beta}^n))$, and $c \le \bar{c}(\hat{\beta}^n)$. This implies that any optimal contract uses a matching cap equal to naives' anticipated savings level such that naives save strictly less than the cap.

It is left to prove that the rationals' payoff under the optimal contract is the same as naives' payoff. Since we know that any optimal contract must offer an m in one of the two small m subcases, we will consider each subcase in turn, beginning with the interior case, when $c^*(r, m) < \bar{c}(\hat{\beta}^n) \le \bar{c}(1)$. Since payoffs are independent of r, for all r that are in used in any optimal contract (see equation (41)), it suffices to consider the case when $r = 0$. In that case, $c^*(0, m) = \frac{w^*}{2 + mK} > \frac{w^*}{2 + m} = \underline{c}(1)$, so that the rational saves just to the cap. Since naives and rationals anticipate saving the same amount, they receive the same payoff.

For the $c^*(r, m) = \bar{c}(\hat{\beta}^n)$ subcase, note that $\underline{c}(1) > \bar{c}(\hat{\beta}^n)$ whenever $\frac{\hat{\beta}^n w - \frac{r}{1+m}}{1 + \hat{\beta}^n} < \frac{w - r}{2 + m}$. Here too it suffices to consider the case where $r = 0$ (see equation 41), where the rational saves just to the cap in any optimal contract whenever,

$$
\begin{aligned}
\underline{c}(1) < \bar{c}(\hat{\beta}^n) &\iff \frac{w^*}{2 + m} < \frac{w^* \hat{\beta}^n}{1 + \hat{\beta}^n} \\
&\iff 2 + m > \frac{1 + \hat{\beta}^n}{\hat{\beta}^n}
\end{aligned}
$$
(49)

Suppose m in an optimal contract is in the interior of this subcase, so that $m = \frac{1}{\bar{\sigma}} - 2$. Then for this to be true we must have $\frac{1}{\bar{\sigma}} > \frac{1 + \hat{\beta}^n}{\hat{\beta}^n}$, which is always true. Suppose instead there is an optimal contract with m at the boundary between the two subcases, $\frac{\frac{1}{\hat{\beta}^n} - 1}{K}$. Then we have,

$$
\begin{aligned}
\underline{c}(1) < \bar{c}(\hat{\beta}^n) &\iff 2 + \frac{\frac{1}{\hat{\beta}^n} - 1}{K} > \frac{1 + \hat{\beta}^n}{\hat{\beta}^n} \\
&\iff \frac{1 - \hat{\beta}^n}{K} > 1 - \hat{\beta}^n,
\end{aligned}
$$
(50)

which is always true. Hence in either type of optimal contract, rationals save to the cap and hence naives and rationals anticipate savings the same amount and therefore receive the same payoff. $\square$

Lemma 2. *In any competitive equilibrium, the sophisticates receive a payoff of $2\ln(\gamma/2)$.*

Proof. The contract $(\gamma/2, \gamma/2, 0, 0)$ produces a payoff of $2\ln(\gamma/2)$ and makes nonnegative profits. Thus, free entry guarantees a payoff that large. Suppose instead sophisticates that a contract (w', r', m', c') and equilibrium payoff strictly greater than $2\ln(\gamma/2)$. Then $\ln(w' - s(w', r', m'|\beta^s)) + \ln(r' + (1 + m')s(w', r', m'|\beta^s)) > 2\ln(\gamma/2)$. By the concavity of the log function, this implies total consumption, and thus total compensation, exceeds γ. Under any contract, sophisticates save weakly less than any other type, so this implies that average compensation of the types who choose the contract (w', r', m', c') is also strictly greater than γ, implying negative profits, contradicting our supposition of an equilibrium. $\square$

Lemma 3. *In any competitive equilibrium, in any contract preferred by only rationals, rationals receive a payoff of $2\ln(\gamma/2)$.*

Proof. We omit the proof since it follows the proof of Lemma 2 nearly exactly. $\square$

Let C_{rn}^* represent the set of contracts preferred by the naives from the set of all nonnegative-profit contacts when pooled with the rationals. A profile consisting of a contract from C_{rn}^*, preferred by the naives and rationals, and the contract $(\gamma/2, \gamma/2, 0, 0)$, preferred by the sophisticates is an equilibrium. Clearly, no alternative contract can be strictly preferred by sophisticates or rationals, on their own, and make nonnegative profits since no nonnegative-profit separating contract can deliver a total compensation above γ, yielding utility no better than $2\ln(\gamma/2)$. Furthermore, no nonnegative-profit pooling contract can be strictly preferred by the sophisticates, since they can never receive total consumption exceeding γ, even in a pooling contract, since they will always save weakly less than average. Finally, there can be no alternative nonnegative-profit pooling contract strictly preferred by both naives and rationals by the definition of C_{rn}^*. Finally, there cannot be an entrant that is strictly preferred by only the naives. We proved in Lemma 1 that the naives and rationals get the same payoff in the proposed equilibrium. In any contract, rationals' payoff under the contract is weakly greater than naives' payoff. Thus any contract preferred by the naives to the equilibrium contract will also be preferred by the rationals.

Any equilibrium contract set must include a contract from C_{rn}^* preferred by naives and rationals and one contract that delivers $c_1 = c_2 = \gamma/2$ to sophisticates and no other contracts. By Lemma 2, in any equilibrium the sophisticates receive a payoff of $2\ln(\gamma/2)$. The only way for them to receive that in equilibrium is through a contract that delivers $c_1 = c_2 = \gamma/2$. The contracts in C_{rn}^* do not deliver consumption of $c_1 = c_2 = \gamma/2$ to sophisticates since they offer a matching rate less than m^{FB}.

Naives must receive a payoff at least as high as they receive in the contracts in the set C_{rn}^*, which we showed above is strictly greater than $2\ln(\gamma/2)$. Otherwise, a contract in the set C_{rn}^* could enter and make nonnegative profits given that naives and (potentially) rationals would strictly prefer it.

Third, in any equilibrium, rationals must pool with naives. Suppose not. By Lemma 3, rationals attain $2\ln(\gamma/2)$ in any equilibrium separating contract. However, we have already shown that naives must receive an equilibrium payoff at least as high as they receive in contracts in the set C_{rn}^*, which is greater than $2\ln(\gamma/2)$. Rationals' payoff under that contract is weakly greater than naives' payoff. This implies that rationals would strictly prefer the contract preferred by naives.

Fourth, rationals and naives must pool in a contract in the set C_{rn}^*. Suppose not. We have already shown that rationals must pool with naives, that implies that rationals and naives pool in some contract not in C_{rn}^*. But then a contract in the set C_{rn}^* could profitably enter, make naives strictly better off (by definition of C_{rn}^*).

$\square$

Proof of Proposition 4.

Lemma 4. *Any equilibrium matching contract must employ the default that minimizes average matched savings under the contract, given the other contractual terms. That default is either 0 or $\underline{d}$.*

Proof. Suppose the proposition is not true and there is some competitive equilibrium in which some matching contract (w, r, m, c, d) includes a default that does not minimize average matched savings. First, suppose that contract is preferred by rationals and naives. Consider the alternative contract (w', r, m, c, d') that uses the default d' that minimizes matched savings and a slightly higher wage $w' > w$. That alternative contract would be strictly preferred by both rationals and

naives. Moreover, there exists a $w' > w$ such that that alternative contract would still make nonnegative profits since lower matching payments are made under the alternative contract. This implies a failure of the free entry condition of equilibrium.

Second, suppose the equilibrium contract (w, r, m, c, d) is preferred by only one type, either naives or rationals. This implies that the other type strictly prefers some other contract in the equilibrium. This implies that there exists an alternative contract (w', r, m, c, d') that is strictly preferred by the type that prefers (w, r, m, c, d) from among the equilibrium contracts and that the other type does not prefer to its preferred equilibrium contracts, and that makes nonnegative profits. This implies a failure of the free entry condition of equilibrium.

To see that the default that minimizes average matched savings in a matching contract is either 0 or $\underline{d}$, first note that, relative to $d = 0$, any default in $(0, \underline{d})$ or greater than $\bar{d}$ results in the same savings outcomes for advice takers and active savers but strictly higher savings for procrastinators. Similarly, relative to $d = \underline{d}$, any default in $(\underline{d}, \bar{d}]$ results in the same savings outcomes for active savers but strictly higher savings for procrastinators and advice takers.

□

Lemma 5. *The highest utility rationals can achieve from among the set of all contracts that make nonnegative profits when taken up by only the rationals exceeds* $2\ln(\gamma/2)$ *and approaches* $2\ln(\gamma/2)$ *as* κ^a *approaches 1.*

Proof. Consider contracts in which the worker anticipates receiving matching contributions, since the best payoff attainable in a non-matching contract is $2ln(\gamma/2$. The characterization of savings, the cap, and non-discretionary contribution, and the zero-profit wage follows that in Equations 23 - 41 of Lemma 1 nearly identically, with two differences. First, the problem is that of a worker with $\beta = \hat{\beta} = 1$. Second, average matched savings will be affected by the default, so that

$$\bar{s}(w, r, m, c, d) = \begin{cases} \left[\kappa^a + \kappa^t\right] \min(c, s(w, r, m, c)) + \kappa^p \min(c, dw) & \text{if} \quad d < \underline{d} \\ \kappa^a \min(c, s(w, r, m, c)) + \left[\kappa^p + \kappa^t\right] \min(c, dw) & \text{if} \quad d \in [\underline{d}, \bar{d}] \\ \min(c, s(w, r, m, c)) & \text{if} \quad d > \bar{d}, \end{cases}$$

which is important for its implication for the zero-profit wage $w^*(r, m, c, d) = \gamma - \bar{s}(w^*, r, m, c, d)m - r$.

Following Lemma 1, we can show that $r^*(m, d) = 0$, $\underline{c} < c^*(r^*, m, d) < \bar{c}$, and letting $K(\underline{d}) \equiv \frac{\kappa^a}{1+m(1-\kappa^a)\underline{d}}$ and $K(0) \equiv \kappa^a + \kappa^t$, $c^*(r, m, d)$ satisfies the first-order condition

(51)
$$\frac{1 + mK(d)}{w^* - c^*} = \frac{1}{c^*},$$

or $c^* = \frac{w^*}{2 + mK(d)}$. Making these substitutions,

(52)
$$w^*(0, c^*, m, d) \equiv w^*(m, d) = \begin{cases} \frac{\gamma}{2} \frac{2+mK(0)}{1+mK(0)}, & \text{if } d = 0 \\ \frac{\gamma}{2} \frac{2+mK(\underline{d})}{1+m[\kappa^a+(1-\kappa^a)\underline{d}]}, & \text{if } d = \underline{d} \end{cases}$$

and we can substitute this into the optimization problem

$$\max_{m,d} V(m, d) = ln(w^*(m, d) - c^*(m, d)) + ln((1 + m)c^*(m, d))$$

(53)
$$= \begin{cases} 2\ln(\frac{\gamma}{2}) + \ln(\frac{1+m}{1+mK(0)}), & \text{if } d = 0 \\ 2\ln(\frac{\gamma}{2}) + \ln(\frac{1+m}{[1+m[\kappa^a+(1-\kappa^a)\underline{d}]][1+m(1-\kappa^a)\underline{d}]}), & \text{if } d = \underline{d} \end{cases}$$

It remains to characterize the optimal match, for a given default, taking the derivative

$$(54) \qquad V_m(m, d) = \begin{cases} -\frac{(\kappa^a + \kappa^t)}{1 + m[\kappa^a + \kappa^t]} + \frac{1}{1+m}, & \text{if } d = 0 \\ -\left[\frac{(1-\kappa^a)\underline{d}}{1+m(1-\kappa^a)\underline{d}} + \frac{\kappa^a + (1-\kappa^a)\underline{d}}{1 + m[\kappa^a + (1-\kappa^a)\underline{d}]}\right] + \frac{1}{1+m}, & \text{if } d = \underline{d}. \end{cases}$$

Inspection reveals that the payoff in the $d = 0$ contract is strictly increasing in m for all m. Thus, in that case, the payoff can be made arbitrarily close to the limit of $V(m, 0)$ as m approaches infinity, which by inspection of the first line of the value function (53) is $2\ln(\gamma/2) - \ln(\kappa^a + \kappa^t)$.

Under an automatic-enrollment strategy with $d = \underline{d}$, the optimal match is positive (assuming $\underline{d} < 1/2$), and satisfies $m^* + 1 = \sqrt{(\frac{1-\sigma_1}{\sigma_1})(\frac{1-\sigma_2}{\sigma_2})}$, where $\sigma_1 = \kappa^a + (1-\kappa^a)\underline{d}$ and $\sigma_2 = (1-\kappa^a)\underline{d}$. Plugging this in to the value function,

$$
\begin{aligned}
(55) \quad V(m^*, \underline{d}) &= 2\ln(\gamma/2) - 2\ln[\sqrt{(1-\sigma_1)\sigma_2} + \sqrt{\sigma_1(1-\sigma_2)}] \\
&= 2\ln(\gamma/2) - 2\ln[(1-\kappa^a)\sqrt{(1-\underline{d})\underline{d}} + \sqrt{\kappa^a + (1-\kappa^a)^2\underline{d}(1-\underline{d})}],
\end{aligned}
$$

which decreases in κ^a, reaching $2\ln(\gamma/2)$ as $\kappa^a \to 1$.

Both these payoffs exceed $2\ln(\gamma/2)$, the best a rational can do in a non-matching contract, so the rational's preferred zero-profit contract is a matching contract of the sort outlined. But these payoffs approach $2\ln(\gamma/2)$ from above at κ^a approaches 1. $\qquad \square$

Lemma 6. *Consider the set of contracts that make nonnegative profits when taken up by both naives and rationals. There exists a $\underline{\kappa} < 1$ such that if $\kappa^a > \underline{\kappa}$ then naives' preferred contracts in that set are matching contracts such that: (1) the matching cap is set at the naives' anticipated savings level; (2) naives who make an active choice in fact save strictly less than the matching cap; (3) rationals who make an active choice save at the matching cap; (4) the default contribution rate is the one that minimizes average worker savings in the contract, given the other terms of the contract, and the default contribution amount, dw, is strictly below the contract's matching cap; and (5) it achieves a payoff for both rationals and naives of at least $2\ln(\gamma/2) + \ln(\frac{\hat{\beta}^n}{(1+\hat{\beta}^n)^2}) - \ln(\bar{\sigma}(1-\bar{\sigma})) > 2\ln(\gamma/2)$, where $\bar{\sigma} = \kappa^n \frac{\beta}{1+\beta} + \kappa^r \frac{\hat{\beta}}{1+\hat{\beta}}$.*

Proof. Consider contracts in which the naive worker anticipates receiving matching contributions, since the best payoff attainable in a non-matching contract is $2ln(\gamma/2)$. The characterization of savings and the calculation of how the zero-profit wage adjust with the cap follows that in Equations 23 - 30 of Lemma 1 nearly identically, with the sole difference that average matched savings will be affected by the default, so that

$$
\bar{s}(w, r, m, c, d) = \begin{cases} \left[\kappa^a + \kappa^t\right]\min(c, s(w, r, m, c)) + \kappa^p \min(c, dw) & \text{if} \quad d < \underline{d} \\ \kappa^a \min(c, s(w, r, m, c)) + \left[\kappa^p + \kappa^t\right]\min(c, dw) & \text{if} \quad d \in [\underline{d}, \bar{d}] \\ \min(c, s(w, r, m, c)) & \text{if} \quad d > \bar{d}, \end{cases}
$$

which is important for its implication for the zero-profit wage $w^*(r, m, c, d) = \gamma - \bar{s}(w^*, r, m, c, d)m - r$.

We can express $\frac{\partial w^*}{\partial c}$ as,

(56)

$$\frac{\partial w^*}{\partial c} = \begin{cases} -\frac{m\kappa^a}{1+m(1-\kappa^a)\underline{d}}, & \text{if } c < \bar{c}(\beta^n) \text{ and } d = \underline{d} \text{ with } \underline{d}w^* < c \\ -m(\kappa^a + \kappa^t), & \text{if } c < \bar{c}(\beta^n) \text{ and } d = 0 \\ \frac{-m\kappa^r\kappa^a}{1+m[\kappa^n\kappa^a\frac{\beta^n}{1+\beta^n}\mathbf{1}(\beta^n w^* > r/(1+m))+(1-\kappa^a)\underline{d}]}, & \text{if } d = \underline{d} \text{ with } \underline{d}w^* < c \text{ and } \bar{c}(\beta^n) < c < \bar{c}(1) \\ \frac{-m\kappa^r(\kappa^a+\kappa^t)}{1+m[\kappa^n(\kappa^a+\kappa^t)\frac{\beta^n}{1+\beta^n}\mathbf{1}(\beta^n w^* > r/(1+m))]}, & \text{if } d = 0 \text{ and } \bar{c}(\beta^n) < c < \bar{c}(1) \\ 0, & \text{if } \bar{c}(1) < c \text{ and either } d = 0 \text{ or } d = \underline{d} \text{ with } \underline{d}w^* < c, \end{cases}$$

where we do not need to consider the $d = \underline{d}$ case with $\underline{d}w^* > c$.

Denote the optimal c, given r, m, and d, as $c^*(r, m, d)$. Consider first whether $c > \bar{c}(\hat{\beta}^n)$ could be optimal. By inspection of V_c (given in (29) above), together with the calculation in (56), $c = \bar{c}(\hat{\beta}^n)$ is weakly preferred to all higher caps, for either default choice, and the preference is strict for $\hat{\beta}^n < 1$.

Now consider whether $c < \underline{c}(\hat{\beta}^n)$ could be optimal. The calculation in (56) shows that $|\frac{\partial w^*}{\partial c}| < m$, for either default choice. Given this fact, inspection of V_c in (29) allows us to rule out any cap below $\underline{c}(\hat{\beta}^n)$, for either default choice, as the naive's objective function strictly increases in the cap.

It is left to consider caps between $\underline{c}(\hat{\beta}^n)$ and $\bar{c}(\hat{\beta}^n)$, where the naive will anticipate saving just to the cap. In this range,

(57) $$V_c(r, m, c, d) = \frac{(1+m)}{c(1+m)+r} - \frac{1+mK(d,c)}{w^*-c}$$

where $K(d, c) \equiv \frac{-1}{m}\frac{\partial w^*}{\partial c}$ is strictly between zero and 1 and depends on both the default and whether the cap exceeds $\bar{c}(\beta^n)$.

Consider first the lowest cap in this range, $\underline{c}(\hat{\beta}^n)$. The right-hand derivative there is given by,

(58)

$$V_{c+}(r, m, \underline{c}(\hat{\beta}^n), d) = \frac{(1+m)}{\frac{w^*\hat{\beta}^n-r}{1+\hat{\beta}^n+m}(1+m)+r} - \frac{1+mK(d,\underline{c}(\hat{\beta}^n))}{w^*-\frac{w^*\hat{\beta}^n-r}{1+\hat{\beta}^n+m}}$$

$$= \frac{\frac{1}{\hat{\beta}^n}(1+m)-(1+mK(d,\underline{c}(\hat{\beta}^n)))}{\frac{w^*(1+m)+r}{1+\hat{\beta}^n+m}} > 0,$$

and therefore $c^*(r, m, d) > \underline{c}(\hat{\beta}^n)$.

Now suppose $\bar{c}(\beta^n) > \underline{c}(\hat{\beta}^n)$. For $\underline{c}(\hat{\beta}^n) < c < \bar{c}(\beta^n)$, V_c is strictly decreasing in c. It is also strictly decreasing in c when $\bar{c}(\beta^n) < c < \bar{c}(\hat{\beta}^n)$. The derivatives from the right and left at $\bar{c}(\beta^n)$ are:

(59) $$V_{c+}(r, m, \bar{c}(\beta^n), d) = \frac{1-(1+mK(d,\bar{c}(\beta^n)_+))\beta^n}{\frac{\beta^n}{1+\beta^n}[w^*+\frac{r}{1+m}]}$$

(60) $$V_{c-}(r, m, \bar{c}(\beta^n), d) = \frac{1-(1+mK(d,\bar{c}(\beta^n)_-))\beta^n}{\frac{\beta^n}{1+\beta^n}[w^*+\frac{r}{1+m}]},$$

where $K(d, c_+)$ and $K(d, c_-)$ are the values of $K(d, c)$, on the right and left side of c, respectively.

Note that, for given $d \in \{0, \underline{d}\}$, $K(d, c_+) < K(d, c_-)$, so the derivative from the right is greater than the derivative from the left. Thus, if the derivative from the left is positive, then so is the

35

derivative from the right. In general, the signs of these derivatives are ambiguous, turning crucially on m.

First, if $m \geq \frac{\frac{1}{\beta^n}-1}{K(d,\bar{c}(\beta^n)_+)}$ ("big m"), then $V_{c-}(r,m,\bar{c}(\beta^n),d) < 0$ and $V_{c+}(r,m,\bar{c}(\beta^n),d) \leq 0$, which implies that $c^*(r,m,d) < \bar{c}(\beta^n)$ so that both naive and rational workers will actually save enough to receive the full match in the contract. It is easy to bound the highest payoff for naives possible in this subset of the contract space. To do so, note that the set of nonnegative-profit contracts that satisfy $c^*(r,m,d) < \bar{c}(\beta^n)$ considered here is a strict subset of the constraint set considered in lemma 5, and moreover rationals' payoff from any given contract is always weakly greater than naives' payoff. Therefore, the highest payoff possible in this big m case must be weakly less than the highest payoff possible to rationals in the constraint set considered in lemma 5, which approaches $2ln(\gamma/2)$ as κ^a approaches 1.

Consider for the rest of the proof the second key case, $m < \frac{\frac{1}{\beta^n}-1}{K(d,\bar{c}(\beta^n)_+)}$ ("small m"), which implies that $V_{c+}(r,m,\bar{c}(\beta^n),d) > 0$ so that $c^*(r,m,d) > \bar{c}(\beta^n)$, and if $c^*(r,m,d) < \bar{c}(\hat{\beta}^n)$ it must satisfy $V_c(r,m,c,d) = 0$, where V_c is defined as in (57) and,

$$(61) \qquad K(d,c) = K(d) = \begin{cases} \frac{\kappa^r \kappa^a}{1+m[\kappa^n \kappa^a \frac{\beta^n}{1+\beta^n}\mathbf{1}(\beta^n w^*>r/(1+m))+(1-\kappa^a)\underline{d}]}, & \text{if } d = \underline{d} \text{ with } \underline{d}w^* < c \\ \frac{\kappa^r(\kappa^a+\kappa^t)}{1+m[\kappa^n(\kappa^a+\kappa^t)\frac{\beta^n}{1+\beta^n}\mathbf{1}(\beta^n w^*>r/(1+m))]}, & \text{if } d = 0 \,, \end{cases}$$

where we can drop the dependence on c since in the range of caps under consideration it no longer depends on c.

At $\bar{c}(\hat{\beta}^n) = \frac{w^*\hat{\beta}^n - r\frac{r}{1+m}}{1+\hat{\beta}^n}$,

$$(62) \qquad V_{c-}(r,m,\bar{c}(\hat{\beta}^n),d) = \frac{1 - \hat{\beta}^n(1+mK(d))}{\frac{\hat{\beta}^n}{1+\hat{\beta}^n}(w+\frac{r}{1+m})}.$$

This derivative is negative if $m > \frac{\frac{1}{\hat{\beta}^n}-1}{K(d)}$. In particular, it is negative in any matching contract with $\hat{\beta}^n = 1$, which together with (56) and (56) implies that $c^*(r,m,d) \leq \bar{c}(\hat{\beta}^n)$ for all $\hat{\beta}^n$. If this derivative is negative, then any $c^*(r,m,d) > 0$ satisfies the first-order condition,

$$(63) \qquad \frac{1+mK(d)}{w^*-c^*} = \frac{1+m}{(1+m)c^*+r}$$

or $c^*(r,m,d) = \frac{w^*-(1+mK(d))\frac{r}{1+m}}{2+mK(d)}$. If $c = 0$, then the contract is not a matching contract, and since we have assumed the solution is a matching contract, that implies that this first-order condition is satisfied in any optimal contract. Finally, if $m \leq \frac{\frac{1}{\hat{\beta}^n}-1}{K(d)} \equiv \bar{m}(d)$, then this derivative is positive and $c^*(r,m,d) = \bar{c}(\hat{\beta}^n)$.

Substituting $c^*(r,m,d)$ into the objective function, the problem becomes,

$$(64) \qquad \max_{r,m,d} V(r,m,d) = \ln(w^*(r,m,d) - c^*(r,m,d)) \\ + \ln(c^*(r,m,d)(1+m)+r),$$

where $w^*(r,m,d) \equiv w^*(r,m,c^*(r,m,d),d)$. Denote the r that maximizes this objective function given m and d by the potentially set-valued function $r^*(m,d)$.

Consider the optimal r in each of these small m cases. First, if $m > \bar{m}(d)$ so that $c^*(r, m, d)$ is in the interior, then the optimization problem becomes,

(65)
$$\max_{r,m,d} V(r, m, d) = 2\ln(w^*(r, m, d) + \frac{r}{1+m}) + \ln(\frac{(1+m)(1+mK(d))}{(2+mK(d))^2}).$$

The third line of (23) gives us that $s(w^*, r, m, c^*, d|\beta^n) = (\frac{\beta^n w^* - \frac{r}{1+m}}{1+\beta^n})\mathbf{1}(w^*\beta^n > \frac{r}{1+m})$. Substituting for c^* and s into the implicit definition of $w^*(r, m, d)$ and solving for w^* yields,

(66)
$$w^* = \begin{cases} \frac{\gamma - \frac{mr}{1+m}\kappa^p}{1+m(\kappa^a+\kappa^t)[\frac{\kappa^r}{2+mK(0)}+\kappa^n\frac{\beta^n}{1+\beta^n}]} - \frac{r}{1+m}, & \text{if } d = 0 \text{ and } \frac{\beta^n(\gamma - \kappa^p\frac{rm}{1+m})}{1+m(\kappa^a+\kappa^t)[\frac{\kappa^r}{2+mK(0)}+\kappa^n\frac{\beta^n}{1+\beta^n}]} > \frac{r(1+\beta^n)}{1+m} \\ \frac{\gamma - \frac{mr}{1+m}((\kappa^a+\kappa^t)\kappa^n + \kappa^p)}{1+m(\kappa^a+\kappa^t)[\frac{\kappa^r}{2+mK(0)}]} - \frac{r}{1+m}, & \text{if } d = 0 \text{ and } \frac{\beta^n(\gamma - \kappa^p\frac{rm}{1+m})}{1+m(\kappa^a+\kappa^t)[\frac{\kappa^r}{2+mK(0)}+\kappa^n\frac{\beta^n}{1+\beta^n}]} \leq \frac{r(1+\beta^n)}{1+m} \\ \frac{\gamma - \frac{mr}{1+m}(1-\kappa^a)(1-\underline{d})}{1+m\kappa^a[\frac{\kappa^r}{2+mK(\underline{d})}+\kappa^n\frac{\beta^n}{1+\beta^n}]+m(1-\kappa^a)\underline{d}} - \frac{r}{1+m}, & \text{if } d = \underline{d} \text{ and } \frac{\beta^n(\gamma - \frac{mr}{1+m}(1-\kappa^a)(1-\underline{d}))}{1+m\kappa^a[\frac{\kappa^r}{2+mK(\underline{d})}+\kappa^n\frac{\beta^n}{1+\beta^n}]+m(1-\kappa^a)\underline{d}} > \frac{r(1+\beta^n)}{1+m} \\ \frac{\gamma - \frac{mr}{1+m}(\kappa^a\kappa^n + (1-\kappa^a)(1-\underline{d}))}{1+m\kappa^a[\frac{\kappa^r}{2+mK(\underline{d})}]+m(1-\kappa^a)\underline{d}} - \frac{r}{1+m}, & \text{if } d = \underline{d} \text{ and } \frac{\beta^n(\gamma - \frac{mr}{1+m}(1-\kappa^a)(1-\underline{d}))}{1+m\kappa^a[\frac{\kappa^r}{2+mK(\underline{d})}+\kappa^n\frac{\beta^n}{1+\beta^n}]+m(1-\kappa^a)\underline{d}} \leq \frac{r(1+\beta^n)}{1+m} \end{cases}$$

In all four cases $w^* + \frac{r}{1+m}$ is strictly decreasing in r, and therefore the objective function in (65) is strictly decreasing in r, which implies that $r^*(m, d) = 0$. It remains to consider the optimal m in this range. Define $V(m, d) \equiv V(r^*(m, d), m, d)$. We will maximize this objective function with respect to m, but first it is convenient to simplify:

(67)
$$V(m, d) = \begin{cases} 2\ln(\gamma) + \ln\left[\frac{(1+m)(1+mK(0))}{[2+mK(0)+m(\kappa^a+\kappa^t)[\kappa^r+(2+mK(0))\kappa^n\frac{\beta}{1+\beta}]]^2}\right], & \text{if } d = 0 \\ 2\ln(\gamma) + \ln\left[\frac{(1+m)(1+mK(\underline{d}))}{[2+mK(\underline{d})+m[\kappa^a\kappa^r+(2+mK(\underline{d}))(\kappa^a\kappa^n\frac{\beta}{1+\beta}+(1-\kappa^a)\underline{d})]]^2}\right], & \text{if } d = \underline{d} \end{cases}$$

$$= \begin{cases} 2\ln(\gamma/2) + \ln\left[\frac{1+m}{[1+m(\kappa^a+\kappa^t)(\kappa^n\frac{\beta}{1+\beta})][1+m(\kappa^a+\kappa^t)(\kappa^r+\kappa^n\frac{\beta}{1+\beta})]}\right], & \text{if } d = 0 \\ 2\ln(\gamma/2) + \ln\left[\frac{1+m}{[1+m[\kappa^a\kappa^n\frac{\beta}{1+\beta}+(1-\kappa^a)\underline{d}]][1+m[\kappa^a\kappa^r+\kappa^a\kappa^n\frac{\beta}{1+\beta}+(1-\kappa^a)\underline{d}]]}\right], & \text{if } d = \underline{d}. \end{cases}$$

It's convenient to define,

(68)
$$\sigma_1(d) = \begin{cases} (\kappa^a + \kappa^t)\kappa^n\frac{\beta^n}{1+\beta^n}, & \text{if d=0} \\ \kappa^a\kappa^n\frac{\beta^n}{1+\beta^n} + (1-\kappa^a)\underline{d}, & \text{if d=}\underline{d} \end{cases}$$

$$\sigma_2(d) = \begin{cases} \sigma_1(d) + (\kappa^a + \kappa^t)\kappa^r, & \text{if d=0} \\ \sigma_1(d) + \kappa^a\kappa^r, & \text{if d=}\underline{d}. \end{cases}$$

With this notation, when $m > \bar{m}(d)$, so $c^* = \frac{w^*}{2+mK(d)} < \bar{c}(\hat{\beta}^n)$, we can rewrite

(69)
$$w^*(m, d) = \frac{\gamma(2 + mK(d))}{1 + m\sigma_2(d) + (1 + mK(d))(1 + m\sigma_1(d))},$$

so that

(70)
$$V(m, d) = 2ln(\gamma/2) + ln(1 + m) - ln(1 + m\sigma_1(d)) - ln(1 + m\sigma_2(d)),$$

and

(71)
$$V_m(m, d) = \frac{1}{1+m} - \frac{\sigma_1(d)}{1+m\sigma_1(d)} - \frac{\sigma_2(d)}{1+m\sigma_2(d)}.$$

37

From a direct application of the quadratic theorem, the maximizing match satisfies $m^* = \sqrt{\frac{(1-\sigma_1(d))(1-\sigma_2(d))}{\sigma_1(d)\sigma_2(d)}} - 1$.

Consider now the case where $m \leq \bar{m}(d)$, so that $c^*(r,m,d) = \bar{c}(\hat{\beta}^n) = \frac{\hat{\beta}^n w^* - r/(1+m)}{1+\hat{\beta}^n}$. Following the approach taken for the $c^*(r,m,d) < \bar{c}(\hat{\beta}^n)$ case from equations (65) to (66), it is easy to show that the objective function is also decreasing in r in this case, so that $r^*(m,d) = 0$ in this range of m as well. We can thus write $w^*(r^*,m,d)$ as,

(72) $$w^*(m,d) \equiv w^*(r^*,m,d) = \begin{cases} \dfrac{\gamma}{1+m(\kappa^a+\kappa^t)(\kappa^r \frac{\hat{\beta}^n}{1+\hat{\beta}^n}+\kappa^n \frac{\beta^n}{1+\beta^n})}, & \text{if } d = 0 \\ \dfrac{\gamma}{1+m(\kappa^a(\kappa^r \frac{\hat{\beta}^n}{1+\hat{\beta}^n}+\kappa^n \frac{\beta^n}{1+\beta^n})+(1-\kappa^a)\underline{d})}, & \text{if } d = \underline{d}, \end{cases}$$

and,

(73) $$V(m,d) \equiv V(r^*,m,d) = \ln(w^*(m,d)(1 - \frac{\hat{\beta}^n}{1+\hat{\beta}^n})) + \ln((1+m)w^*(m,d)\frac{\hat{\beta}^n}{1+\hat{\beta}^n})$$

$$= 2\ln(\gamma) + \ln((1 - \frac{\hat{\beta}^n}{1+\hat{\beta}^n})\frac{\hat{\beta}^n}{1+\hat{\beta}^n}) + \ln(1+m) - 2\ln(1+m\sigma(d)),$$

where,

(74) $$\sigma(d) = \begin{cases} (\kappa^a+\kappa^t)(\kappa^r \frac{\hat{\beta}^n}{1+\hat{\beta}^n} + \kappa^n \frac{\beta^n}{1+\beta^n}), & \text{if } d = 0 \\ \kappa^a(\kappa^r \frac{\hat{\beta}^n}{1+\hat{\beta}^n} + \kappa^n \frac{\beta^n}{1+\beta^n}) + (1-\kappa^a)\underline{d}, & \text{if } d = \underline{d} \end{cases}$$

is the average savings rate in a contract with default d. Thus,

(75) $$V_m(m,d) = \frac{1}{1+m} - \frac{2\sigma(d)}{1+m\sigma(d)},$$

and it is a simple maximization to show that the optimal matching rate is $m^*(d) = \frac{1-2\sigma(d)}{\sigma(d)}$.

It's easy to see that the value function is continuous at $\bar{m}(d)$, since $\frac{w^*}{2+\bar{m}(d)K(d)} = \bar{c}(\hat{\beta}^n)$, by definition. We will now show that the derivative of the value function also exists at $\bar{m}(d)$. This claim is true iff the the derivative from the right (71) is equal to the derivative from the left (75),

(76)
$$\frac{1}{1+\bar{m}} - \frac{\sigma_1(d)}{1+\bar{m}\sigma_1(d)} - \frac{\sigma_2(d)}{1+\bar{m}\sigma_2(d)} = \frac{1}{1+\bar{m}} - \frac{2\sigma(d)}{1+\bar{m}\sigma(d)}$$
$$\iff [\frac{\sigma_1(d)+\sigma_2(d)}{2} - \sigma(d)][\hat{\beta}^n\sigma_2(d) - \sigma_1(d)] = (1-\hat{\beta}^n)[\sigma(d)(\frac{\sigma_1(d)+\sigma_2(d)}{2}) - \sigma_1(d)\sigma_2(d)].$$

Dividing into default cases, this is true in the $d = 0$ case if

$$\iff (\kappa^a+\kappa^t)\kappa^r(\frac{1-\hat{\beta}^n}{2(1+\hat{\beta}^n)})[\hat{\beta}^n\kappa^r(\kappa^a+\kappa^t) - (1-\hat{\beta}^n)\sigma_1(0)] =$$

(77)
$$(1-\hat{\beta}^n)\kappa^r(\kappa^a+\kappa^t)[\sigma_1(0)(\frac{\hat{\beta}^n-1}{2(1+\hat{\beta}^n)}) + \frac{\hat{\beta}^n}{2(1+\hat{\beta}^n)}\kappa^r(\kappa^a+\kappa^t)]$$

$$\iff \frac{\hat{\beta}^n\kappa^r(\kappa^a+\kappa^t) - (1-\hat{\beta}^n)\sigma_1(0)}{2(1+\hat{\beta}^n)} = \frac{\hat{\beta}^n\kappa^r(\kappa^a+\kappa^t) - (1-\hat{\beta}^n)\sigma_1(0)}{2(1+\hat{\beta}^n)}.$$

It is also true in the $d = \underline{d}$ case, where the algebra is nearly identical to the $d = 0$ case, merely replacing $\sigma_1(\underline{d})$ for $\sigma_1(0)$ and κ^a for $(\kappa^a + \kappa^t)$.

The continuity of the derivative, together with the fact that the derivative is positive at $m = 0$ and everywhere decreasing, suffices to show that there is a unique maximizer in m and that it will satisfy one of the first-order conditions $V_m(d, m) = 0$.

Consider the contract with $c = \bar{c}(\hat{\beta}^n)$, $r = 0$, $m = \frac{1 - 2\sigma(0)}{\sigma(0)}$, $d = 0$ and $w = w^*(r, m, c, d)$. Our results above show that this contract yields the naives utility of,

$$(78) \qquad V(r, m) = 2\ln(\gamma/2) + \ln(\frac{\hat{\beta}^n}{(1 + \hat{\beta}^n)^2}) - \ln(\sigma(0)(1 - \sigma(0))).$$

As κ^a goes to 1, this payoff asymptotes from above to the analogous payoff from the case without defaults considered in lemma 1, given in (48), which is strictly greater than $2\ln(\gamma/2)$. This implies that there exists a $\underline{\kappa}$ such that for all $\kappa^a > \underline{\kappa}$, any optimal contract must be in the small m case and must have $c > \max(\underline{c}(\hat{\beta}^n), \bar{c}(\beta^n))$, and $c \le \bar{c}(\hat{\beta}^n)$ so that the matching cap equals naives' anticipated savings level and naives in fact save strictly less than the cap.

For sufficiently large $\kappa^a < 1$, the rationals' payoff under the optimal contract is the same as naives' payoff. We have shown that, for sufficiently large $\kappa^a < 1$, any optimal contract must offer an m in one of the two small m subcases. Following the same argument as at the end of the proof of lemma 1, the c in any optimal contract must be strictly greater than $\underline{c}(1)$. Hence for sufficiently large $\kappa^a < 1$, in any optimal contract, rationals save to the cap and hence naives and rationals anticipate saving the same amount and therefore receive the same payoff strictly greater than $2\ln(\gamma/2)$.

Following the argument in the proof of lemma 4, the default of any optimal contract must minimize matched savings under the contract, given the other terms of the contract. In particular, any optimal contract must have $dw < c$, since otherwise $d = 0$ would reduce matched savings. Note that for sufficiently large $\kappa^a < 1$, in any optimal contract all savings are matched and thus the default must also minimize average savings under the contract.

Finally, we show that average second-period consumption is equal to $\gamma/2$ both in equilibrium and in the equilibrium in which the contract space is restricted to contracts with $d = 0$. Taking $d \in \{0, \underline{d}\}$ as given, let $\bar{\sigma}(d)$ denote the average savings rate in the naive's preferred contract given that default (w^*, m^*, r^*, d, c^*). We know that there is no non-matched savings in this contract and $r^* = 0$, so average second period consumption is simply $w^*\bar{\sigma}(d)(1 + m^*)$, and zero-profits implies that $w^* = \frac{\gamma}{1 + m^*\bar{\sigma}(d)}$. Substituting for this wage, average second-period consumption is given by

$$(79) \qquad \bar{C}_2 = \gamma(\frac{\bar{\sigma}(d) + \bar{\sigma}(d)m^*}{1 + \bar{\sigma}(d)m^*}).$$

It's easy to check that if $m^* = \frac{1}{\bar{\sigma}(d)} - 2$, then $\bar{C}_2 = \gamma/2$.

If that contract is such that $m^* < \bar{m}$, we know that $\bar{\sigma}(d) = \sigma(d)$ from (74), and $m^* = \frac{1}{\bar{\sigma}(d)} - 2$. If the contract is such that $m^* > \bar{m}$, then

$$(80) \qquad \bar{\sigma}(d) = \begin{cases} (\kappa^a + \kappa^t)(\kappa^n \frac{\beta}{1+\beta} + \kappa^r \frac{c^*}{w^*}), & \text{if } d = 0 \\ \kappa^a(\kappa^n \frac{\beta}{1+\beta} + \kappa^r \frac{c^*}{w^*}) + (1 - \kappa^a)\underline{d}, & \text{if } d = \underline{d} \end{cases}$$

$$= \sigma_1(d) + \frac{(\sigma_2(d) - \sigma_1(d))(1 + m\sigma_1(d))}{2 + m(\sigma_1(d) + \sigma_2(d))}.$$

Thus,

$$(81) \qquad \frac{1}{\bar{\sigma}(d)} - 2 = \frac{2 - \sigma_1(d) - \sigma_2(d) + m[\sigma_1(d) + \sigma_2(d) - 2\sigma_1(d)\sigma_2(d)]}{\sigma_1(d) + \sigma_2(d) + 2m\sigma_1(d)\sigma_2(d)} - 1,$$

solving $\frac{1}{\bar{\sigma}(d)} - 2 = m$ for m yields $m = \sqrt{\frac{(1-\sigma_1(d))(1-\sigma_2(d))}{\sigma_1(d)\sigma_2(d)}} - 1$, exactly the match derived from the first-order condition in equation (71). Thus, in this case too $\bar{C}_2 = \gamma/2$.

$\square$

Let C_{rn}^* represent the set of contracts preferred by the naives from the set of all nonnegative-profit contacts when pooled with the rationals. Consider an equilibrium consisting of a contract from C_{rn}^*. We will show that if κ^a is sufficiently close to 1, then such a contract is an equilibrium. If not, there must be some alternative contract (w, r, m, c, d) that is strictly preferred by one or more types and would make nonnegative profits if chosen by the types that strictly prefer it.

First, note that by Lemmas 5 and 6, for κ^a sufficiently close to 1, the highest payoff rationals could achieve in an alternative nonnegative-profit contract preferred by only rationals is below their payoff from a contract in C_{rn}^*. So there cannot be an alternative nonnegative-profit contract strictly preferred by only rationals. Second, there can be no nonnegative-profit pooling contract that is strictly preferred by both naives and rationals by the definition of C_{rn}^*. Thus, the only type of alternative contract that remains to be considered is one that is strictly preferred by only the naives. But no such entrant can exist. We proved in Lemma 6 that the naives and rationals get the same payoff in the proposed equilibrium. In any contract, rationals' payoff under the contract is weakly greater than naives' payoff. Thus any contract preferred by the naives to the equilibrium contract will also be preferred by the rationals.

Finally, any equilibrium must consist of a contract from C_{rn}^*. To see this, first, note that in equilibrium, naives must receive a payoff at least as high as they receive in the contracts in the set C_{rn}^*, which we showed above is strictly greater than $2\ln(\gamma/2)$. Suppose not. Then a contract in the set C_{rn}^* could enter, make naives strictly better off, and make nonnegative profits given that naives and (potentially) rationals would choose it, which is a contradiction.

Second, if $\kappa^a < 1$ is sufficiently high, then in any equilibrium, rationals must pool with naives. Suppose not. Then by Lemma 5, we can choose κ^a close enough to 1 to make rationals' highest payoff in a nonnegative-profit separating contract arbitrarily close to $2\ln(\gamma/2)$. However, we have already shown that naives must receive an equilibrium payoff at least as high as they receive in contracts in the set C_{rn}^*, which is greater than $2\ln(\gamma/2)$. In any contract, rationals' payoff under the contract is weakly greater than naives' payoff. This implies that rationals would strictly prefer the contract preferred by naives, which is a contradiction.

Third, rationals and naives must pool in a contract from the set C_{rn}^*. Suppose not. Then a contract in the set C_{rn}^* could enter, make naives strictly better off, and make nonnegative profits given that naives and (potentially) rationals would choose it, which is a contradiction with the definition of equilibrium.

Finally, note that by Lemma 6, contracts in C_{rn}^* have all of the characteristics asserted in the statement of the proposition.

$\square$