Exploitative Contracting in a Life Cycle Savings Model*

Tomasz Sulka†

February 23, 2023

Abstract

This paper analyses the interaction between a present-biased individual and a profit-maximising financial provider in order to examine the properties of exploitative savings contracts and the impact of common policy interventions. Using a tractable theoretical model, I find that naïve present-biased agents are offered contracts that are 'inefficiently cheap' (low-yield, low-fee) when the income effect of an interest rate change dominates in the agent's utility function, and 'inefficiently expensive' (high-yield, high-fee) otherwise. Subsequently, I embed the interaction with a pension provider in a numerical life-cycle framework with hyperbolic discounting. The calibrated model indicates that due to naiveté, the contract offered in market equilibrium is Pareto-inefficient. Exploitative contracting reduces the agent's pension wealth by 8%, lowering expected annual consumption in retirement by 3%. The associated loss of consumer welfare corresponds to 0.23% of annual consumption.

JEL: D14, D15, D49, D91, E21

Keywords: Exploitative contracting. Life cycle model. Saving. Pensions. Present bias.

1 Introduction

Pension reforms in the OECD countries tend to reduce the generosity of public benefits and increase the importance of private pensions. For instance, many countries had introduced mandatory, or quasi-mandatory, private pensions, achieving nearly universal coverage of the working-age

*I am hugely indebted to my thesis advisors, Philipp Kircher and Jan Grobovšek, for their continuous support, advice, and kindness. I would like to thank Liang Bai, Rachel Forshaw, Daniel Gottlieb, Paul Heidhues, Ed Hopkins, Johannes Johnen, Tatiana Kornienko, Mats Köster, Botond Kőszegi, Peter Maxted, Mariann Ollar, Frank Schilbach, Carl Singleton, Christopher Stapenhurst, Adam Szeidl, Ina Taneva, and presentation audiences for their valuable inputs. I am grateful to Jeremy Tobacman for generously sharing the simulation code. I would like to thank the ESRC (UK) for financial support.

†DICE, University of Düsseldorf. Email: sulka@dice.hhu.de

population (OECD, 2021). Importantly, the underlying private pension arrangements are predominantly of the defined-contribution (DC) type, under which assets accumulated by the time of retirement directly determine the amount of pension benefit. Because a typical DC pension grants an individual substantial freedom to choose how much to save and how to invest their assets, working-age individuals today bear much more responsibility for their financial security in retirement than the preceding generations.

Given the changing policy environment, a natural question arises of whether individuals are capable of selecting an appropriate pension arrangement and preparing themselves financially for retirement. Note that decisions characterised by high degrees of complexity, intertemporal nature of trade-offs, long planning horizons, and infrequent feedback are particularly prone to various psychological factors. The literature on behavioural household finance (Beshears, Choi, Laibson, and Madrian, 2018; Campbell, 2016) documents that wealth accumulation is indeed affected by self-control problems (e.g., Ameriks, Caplin, Leahy, and Tyler, 2007; Ashraf, Karlan, and Yin, 2006) and procrastination (e.g., Benartzi and Thaler, 2007; Chetty, Friedman, Leth-Petersen, Nielsen, and Olsen, 2014; Choi, Laibson, Madrian, Metrick, and Poterba, 2004; Madrian and Shea, 2001; Thaler and Benartzi, 2004). From a theoretical perspective, these prevalent behaviours can be captured by a model of present-biased preferences (Strotz, 1955), which predicts undersaving due to the self-control problem (Laibson, 1994, 1997; Diamond and Köszegi, 2003) as well as procrastination on tasks with small current costs and substantial future benefits (O'Donoghue and Rabin, 1999a,b, 2001). Regarding life-cycle applications, accounting for the present bias improves the fit to the consumption-saving patterns observed in the data, relative to a standard model with exponential discounting (Angeletos, Laibson, Repetto, Tobacman, and Weinberg, 2001; Laibson, Maxted, Repetto, and Tobacman, 2018).[1]

The approach prevailing in the literature is to analyse the biased financial decisions without considering the supply side of the market for savings products, treating the alternatives available to an agent as exogenously given or designed by a benevolent social planner.[2] This paper goes beyond these assumptions by analysing the interaction between a present-biased individual and a profit-maximizing pension provider. In the first part of the paper, I analyse a simplified model, which allows for analytical tractability and provides some necessary intuition. When the rate of return on savings determines the agent's valuation of a contract, the exploitative contracts aimed at naïve present-biased individuals are either 'inefficiently cheap' (low-yield, low-fee) or 'inefficiently expensive' (high-yield, high-fee), depending on whether the income or the substitution effect of an interest rate change dominates in the agent's utility function. In

[1]Goda, Levy, Manchester, Sojourner, and Tasoff (2019) provide evidence of an empirical link between a direct measure of present bias and a level of retirement savings.

[2]For example, Choi, Laibson, Madrian, and Metrick (2003) and Carroll, Choi, Laibson, Madrian, and Metrick (2009) study socially optimal defaults, while Pavoni and Yazici (2017) derive optimal capital taxation when present bias changes over the life cycle.

the second part of the paper, I embed the interaction with a pension provider in a rich life-cycle framework with hyperbolic discounting, which allows to assess the quantitative importance of contractual design and choice in a realistic environment. The results indicate the prevalence of Pareto-inefficient contract terms in market equilibrium, in line with the qualitative prediction of the simple model. This inefficiency in contractual choice lowers the agent's pension wealth by 8% and reduces expected annual consumption in retirement by 3%.

In the simple model, the financial provider offers a savings contract that specifies a rate of return on the agent's savings and a fee charged for the service, arguably the two most important features of a private pension arrangement (Office of Fair Trading, 2014). The individual evaluates the contract proposed by the provider, taking into account predicted saving over the duration of the contract. When the provider can observe the agent's characteristics and tailor the contract terms accordingly, a contract offered to a sophisticated present-biased agent is efficient in the sense that it maximises social surplus. By contrast, a contract offered to an agent who is (at least partially) naïve about his own present bias is distorted in a way which increases provider's profits by exploiting the agent's naiveté, as in DellaVigna and Malmendier (2004).

Intuitively, a naïve present-biased agent is overly optimistic about his future propensity to save for any contract parameters, and thus his willingness to pay for a contract exceeds the willingness to pay of a sophisticated agent. This could lead to the hasty conclusion that, compared to sophisticates, naifs would be offered expensive savings products with generous rates of return. The optimal distortion is more subtle than that, however, because the provider exploits naiveté about the present bias by offering contract terms that appear attractive as long as the individual saves a lot. This is achieved by distorting the interest rate downwards when the income effect dominates and upwards when the substitution effect dominates. In case when the income effect is stronger, a naïve present-biased agent finds a cheap savings product with low returns and low fees attractive, because he mistakenly expects to counteract the low yield with higher savings and thus benefit from a discounted fee. In case when the substitution effect is stronger, a naïve present-biased agent finds an expensive savings product attractive, because he mistakenly expects to save more when the rate of return is high, justifying higher charges.

In the second part of the paper, I embed the choice of a pension provider in a numerical life-cycle model with hyperbolic discounting. To simulate household behaviour, I modify and recalibrate the framework developed by Laibson et al. (2018).[3] I subsequently add and calibrate the firm side of the model by matching several regularities reported in the empirical literature on operation of pension funds (Basu and Andrews, 2014; Bateman and Mitchell, 2004; Bauer, Cremers, and Frehen, 2010; Bikker and De Dreu, 2009; Bikker, Steenbeek, and Torracchi, 2012;

[3]More precisely, I analyse choices of a representative household parametrised by the short-run discount factor $\beta = 0.503$, the long-run discount factor $\delta = 0.988$, and the CRRA parameter $\theta = 1.105$. At the benchmark, I follow the approach in Laibson et al. (2018) and assume complete naiveté, i.e. $\hat{\beta} = 1$.

Financial Conduct Authority, 2017; OECD, 2017).

The analysis reveals substantial magnitude of the forecasting errors about future saving. At the beginning of the life cycle, a naïve agent expects to retire with pension wealth almost twice as large as the actual average wealth holdings. These forecasting errors are then reflected in the agent's choice of a savings contract. As a measure of efficiency, I adopt a conservative Pareto criterion that asks whether there exist alternative contract parameters that would improve the agent's (actual) utility while retaining the profits earned by the providers in an imperfectly competitive equilibrium. I find that the equilibrium contract is inefficiently 'cheap', consistent with the theoretical prediction. Exploitative contracting lowers the agent's wealth at retirement by 8% and reduces his expected annual consumption in retirement by 3%, generating a consumption-equivalent welfare loss of 0.23%, or $230, per annum. To my knowledge, this is the first estimate of the quantitative importance of exploitative contracting obtained in a realistic life-cycle setting. The calibrated model also allows to compare the magnitude of the effects associated with inefficient contractual choice to those of naiveté and present bias itself. While choosing a Pareto-efficient contract would increase the naïve agent's pension wealth by 8%, becoming sophisticated would increase it by 36% and eradicating the present bias altogether would increase it by 87%, conditional on the contract parameters.

Finally, I simulate the effects of three common policy interventions. Increasing the degree of competition improves consumer welfare at the cost of provider's profits. By contrast, a binding ceiling on fees precludes the market from providing higher-yielding contracts and diminishes consumer welfare. Increasing the liquidity of the pension asset exacerbates the overvaluation of savings contracts by a naïve agent and results in lower wealth accumulation. However, consumer welfare increases monotonically when the pension asset is made more liquid, because the present-biased agent no longer simultaneously accumulates illiquid pension wealth and carries a credit card debt, suffering from the difference in the interest rates.

Related literature

This paper contributes to the literature on behavioural industrial organisation, reviewed recently in Heidhues and Kőszegi (2018). The numerical part of the present paper constitutes the first attempt to structurally assess the quantitative importance of exploitative contracting in a realistically complex setting. In addition, the simple model provides a meaningful condition that determines whether linear contracts aimed at naïve agents are 'inefficiently cheap' or 'inefficiently expensive', relative to the contracts chosen by sophisticates. While already DellaVigna and Malmendier (2004) noted that the direction of such exploitative distortions is ambiguous in general, the underlying condition is shown to depend on relative strength of the substitution and income effects, thus highlighting the importance of interactions between parameters traditionally seen as 'classical' and 'behavioural'. Such considerations have thus far received little attention from the literature on behavioural contracting (Kőszegi, 2014).

Two recent papers on exploitative contracting are complementary to the exercise presented here. Bubb and Warren (2020) analyse the optimal design of a workplace pension plan when the workers are present-biased and under-appreciate the stickiness of the default. They demonstrate that the difference between exploitative and benevolent employers boils down to whether they choose a default that minimises or one that maximises the employer's matching contribution. Gottlieb and Zhang (2021) study a finite-horizon problem of optimal (unrestricted) contracting under present bias. They show that in each period the optimal contract consists of two alternatives, a fictitious and an actual choice, and because the welfare loss from time-inconsistency materialises only in the final period, it vanishes as the contracting horizon becomes sufficiently long.

This paper also contributes to the literature on household finance (Attanasio and Weber, 2010; Beshears et al., 2018; Campbell, 2016). The novelty of the present paper lies in accounting for the supply side of the market for financial products within the framework of a numerical life-cycle model. Given the increasing importance of individual pension choices and the fact that the standard model fails to replicate the observed dispersion in wealth accumulation, the numerical results presented here quantify the impact of contractual choice and various policy interventions on key financial outcomes.

Related work by Lusardi, Michaud, and Mitchell (2017) evaluates the quantitative importance of endogenous financial knowledge accumulation. In their calibrated model, financial literacy is a component of human capital that enables an agent to obtain higher expected returns on investments, albeit at an exogenously given cost. By contrast, the present paper analyses choices of an agent who understands the available contracts perfectly, but mispredicts his own behaviour conditional on contract parameters. More fundamentally, introducing the supply side of the market, and thus endogenising 'prices', allows to conduct welfare analysis and policy counterfactuals while accounting for the market response.

The remainder of this paper proceeds as follows. Section 2 analyses the simple model. Section 3 presents a life-cycle model with choice of a pension provider and the numerical results. Section 4 concludes. Derivations of the main results are provided in the appendix, while the analysis of extensions and a complete set of numerical results are relegated to the online appendix.

2 Simple model

In this section, I set up and analyse a simple model of contracting between a present-biased saver and a profit-maximising financial provider in order to develop necessary intuition, before introducing a much more complex model of life cycle savings and solving the provider's problem numerically. I demonstrate that the direction of the exploitative distortion depends on the curvature of the agent's utility function, and then discuss the robustness of this result to many alternative specifications of the model, including varying the degree of competition and observability of the individual characteristics. Theoretical analysis of three different policy interventions is provided in the online appendix.

2.1 Baseline formulation

Consider a market for savings products in which an individual ("he") interacts with a monopolistic financial provider ("she") who can observe the agent's characteristics and tailor the contract terms accordingly. Motivated by an observation that investment performance and associated charges are the most important features of a private pension arrangement (Office of Fair Trading, 2014), suppose that a linear savings contract P is characterised by two parameters - an interest rate r that the provider offers on the agent's savings and a flat-rate ('per-period') fee f that is charged for the service.

The simple model has three stages. In period 0, the contract is proposed by the provider and evaluated by an individual. If the individual rejects the provider's offer, he obtains reservation utility of $\underline{u}$. This is interpreted as a level of utility associated with reliance on a public pension system and is therefore type-independent. If the individual accepts the offer, he makes a saving decision in period 1, taking the contract terms as given. Retirement occurs deterministically in period 2. Separating contract evaluation from the saving decision is meant to reflect the fact that private pension wealth is typically accumulated over a long time horizon, which limits the power of any initial commitment. For example, even if the individual was able to select a specific contribution rate at the contracting stage, he might expect to regularly revise it upwards as he nears retirement.

At the contracting stage in period 0, the agent's preferences are represented by the following quasi-linear utility function:

$$U_0 = u(c_1) + \delta u(c_2) - f$$

where $u(\cdot)$ is an instantaneous utility function satisfying strict concavity, continuous differentiability, and the Inada condition $\lim_{c \to 0} u'(c) = +\infty$. Parameter $\delta \in (0, 1)$ is the (long-term) discount factor. Let $Y > 0$ denote the individual's take-home pre-retirement income. Then, from the intertemporal budget constraint $c_1 + \frac{c_2}{(1+r)} \leq Y$, the maximisation of U_0 amounts to the choice of savings $s \geq 0$, with $c_1 = Y - s$ and $c_2 = (1 + r)s$. Since $Y > 0$ and $u(\cdot)$ satisfies the Inada condition, this maximisation problem has an interior solution. Under a quasi-linear formulation, the solution is independent of the fee f.

In period 1, when saving takes place, the agent is present-biased and maximises the following utility function, taking contract parameters r and f as given:

$$U_1 = u(c_1) + \beta\delta u(c_2) - f$$

where $\beta \in (0,1]$ is the present bias parameter (Laibson, 1997; Strotz, 1955). Typically, present bias reflects the impact of non-normative factors that make an agent over-value his current consumption relative to future consumption, such as temptation or myopia. The lower β, the larger the self-control problem. For $\beta = 1$, on the other hand, the agent's behaviour is not affected by the present bias. It is thus standard in the literature to interpret U_0 as a normative measure of consumer welfare (see Heidhues and Kőszegi, 2010; O'Donoghue and Rabin, 1999a, 2001; Spiegler, 2011).

Preferences of a present-biased agent exhibit time-inconsistency. Clearly, a consumption bundle that maximises utility function U_1 for $\beta < 1$ differs from a consumption bundle that would maximise U_0. For that reason, the beliefs about the propensity to save in period 1 are crucial when evaluating the contract from the perspective of period 0. Following O'Donoghue and Rabin (2001), denote period-0 singleton beliefs regarding the future present bias by $\hat{\beta} \in [\beta, 1]$. That is, in period 0 an individual believes that his period-1 self will act so as to maximise U_1 parametrised by $\hat{\beta}$. Agents with beliefs $\hat{\beta} > \beta$ are called 'naïve' because they underestimate the magnitude of their present bias. The greater $\hat{\beta}$, the more severe their naiveté. On the other hand, agents with beliefs $\hat{\beta} = \beta$ are 'sophisticated' in the sense that they are fully aware of their present bias.

Let $s(\hat{\beta}, r)$ denote the level of saving that maximises U_1 parametrised by $\hat{\beta}$, given the rate of return r. For conciseness, let the agent's valuation of a savings contract $P = (r, f)$ conditional on his belief $\hat{\beta}$ be represented by $\hat{U}_0 \equiv U_0(s(\hat{\beta}, r), r, f)$. Similarly, contract valuation net of the fee is $\hat{V}_0 \equiv V_0(s(\hat{\beta}, r), r)$, so that $\hat{U}_0 = \hat{V}_0 - f$. The following (standard) result states that naif's optimistic perception of future propensity to save makes him overvalue any savings contract. Intuitively, naiveté about the present bias implies that for any contract parameters, the agent's period-0 self believes that his future savings will be more in line with the current preference.

Lemma 1: *For a strictly concave and continuously differentiable $u(\cdot)$, $s(\hat{\beta}, r)$ is increasing in $\hat{\beta}$. Consequently, $\hat{U}_0$ is increasing in $\hat{\beta}$ for any r, f.*

The short proof is provided in the appendix. In the remainder of the paper, I assume a CRRA utility formulation:

$$u(x) = \tfrac{x^{1-\theta} - 1}{1-\theta}$$

parametrised by $\theta > 0$. This formulation features prominently in the literature on intertemporal choice in the life-cycle context (see Attanasio and Weber, 2010; Browning and Lusardi, 1996).[4]

[4] CRRA stands for 'constant relative risk aversion'. Such utility function is also characterised by a constant elasticity of intertemporal substitution, given by the inverse of θ.

In general, the response of optimal savings to changes in the interest rate is ambiguous. On the one hand, a higher interest rate makes future consumption cheaper relative to current consumption. This generates the substitution effect, according to which optimal savings would increase in the interest rate. On the other hand, under a higher interest rate lower savings are required to support any given level of future consumption. This gives rise to the income effect, according to which optimal savings would decrease in the interest rate. From this perspective, a CRRA utility formulation is particularly useful, as the size of parameter θ determines whether the substitution or the income effect dominates. For $\theta < 1$, the substitution effect is stronger and optimal savings increase in the interest rate. For $\theta > 1$, the income effect is stronger and optimal savings decrease in the interest rate. For the special case of $\theta = 1$, the CRRA utility function takes a logarithmic form and optimal savings are invariant to the interest rate.

Introducing the supply side of the model, suppose that a savings contract is offered by a financial provider implicitly assumed to be a manager of a pension fund. Pension funds manage most of the private pension assets accumulated in the OECD countries (OECD, 2017).[5] Moreover, pension funds are able to offer tailor-made products, both via the retail market and workplace arrangements. Thus, the issues of optimal contract design and regulation appear particularly relevant when applied to pension funds.

The empirical literature divides the total provision costs of pension funds into administrative and investment costs (Basu and Andrews, 2014; Bateman and Mitchell, 2004; Bauer et al., 2010; Bikker and De Dreu, 2009; Bikker et al., 2012). There is consistent evidence that the administrative costs are increasing in the size of an individual pension pot (controlling for the number of pension plan participants) and that they include a substantial fixed cost component. The investment costs are increasing in factors that raise the (expected) rate of return, such as the portfolio share of stocks, the number of available investment options, and the 'quality and complexity' of a plan. These findings seem plausible even if the financial providers are not able to consistently generate abnormal positive returns on risky components of their portfolios by 'exerting effort'. For example, the strategy of holding a particular mix of stocks with the intention of replicating performance of a particular market index over time is plausibly more costly to execute than simply acquiring high-grade bonds.[6]

In order to capture the intuition that it is costly for the financial provider to promise a high rate of return on her clients' savings, suppose that a savings contract is offered by the provider who maximises the following profit function:

[5]This proportion equals 59% in the US, 54% in Canada, 100% in the UK, 97% in Australia, 61% in Japan, and 100% in the Netherlands.

[6]A further concern might be that the current model abstracts from financial risk on the consumer's side, even though most investment strategies are inherently risky. Taken literally, the provider's costs would reflect the investment risk associated with promising a given rate of return to the agent (Antolín, Payet, Whitehouse, and Yermo, 2011).

$$\pi = f - K(r, s)$$

where the cost of the service conditional on savings $K(\cdot, s)$ is increasing, continuously differentiable, and strictly convex in the offered interest rate. Note that a specific cost function will be proposed and calibrated in the quantitative part of the paper, while below I discuss several modifications to this stylised view of the pension industry.

Then, the optimal (profit-maximising) savings contract solves the following problem:

$$\max_{r,f} \ \pi = f - K(r, s), \text{ s.t.:}$$

1. $s = s(\beta, r)$
2. $V_0(s(\hat{\beta}, r), r) - f \geq \underline{u}$

The first constraint means that the provider has a correct expectation of the agent's future savings given the parameters of the contract. It is standard in the exploitative contracting literature to assume that the firm is able to predict the individual's behaviour more accurately than the individual himself (Heidhues and Kőszegi, 2018; Spiegler, 2011). The second constraint is a version of the participation constraint, which says that for the agent to accept the contract in period 0, his *perceived* utility from signing it cannot be lower than the reservation utility. Once accepted, the agent takes the contract parameters as given in period 1.[7]

Solution to the provider's problem is derived under mild technical assumptions.

Assumption 1: *Solution to the provider's problem is obtained under:*

1. $r > 0$

2. $\theta > \underline{\theta} \equiv 1 + \frac{log(\hat{\beta}\delta)}{log(1+r)}$

These parametric assumptions are sufficient to derive a sharp prediction relating the optimal contract design to the relative strength of the substitution and income effects. More precisely, Assumption 1 is sufficient to pin down how the overvaluation of a contract by a naïve agent

[7] Gottlieb and Smetters (2021) study optimal contracts with one-sided commitment in the context of the insurance market. The assumption of two-sided commitment made here appears more appropriate in the context of the market for pension products, where regulation often prevents liquidation and discourages switching providers. In addition, individuals display substantial inertia in their pension choices, even if adjustments are freely allowed (Beshears et al., 2018). See Heidhues, Kőszegi, and Murooka (2021) for a model that captures a failure of a present-biased agent to switch providers.

changes with the offered rate of return. Note that for $\delta < 1$, $\hat{\beta} \leq 1$, and $r > 0$, the threshold $\underline{\theta}$ lies strictly below 1.[8]

Finally, within this simple framework, I measure social surplus as a sum of utility attained by the agent (according to U_0) and the provider's profits. Under the assumption of quasi-linearity, this measure of efficiency coincides with an objective of a time-consistent individual who has access to the provider's technology of converting savings into future wealth and is a function of r only. Contracts that maximise social surplus are called 'efficient', while those that do not are called 'inefficient'.

Proposition 1, the main result of this section, characterises the qualitative properties of the profit-maximising contract, denoted $P^* = (r^*, f^*)$.

Proposition 1: *Suppose $u(\cdot)$ and $K(\cdot,\cdot)$ are as above and Assumption 1 holds. Then, in a monopolistic market with perfect observability:*

1. *For any $\hat{\beta}$, the fee f^* is increasing in r. For any r, the fee f^* is increasing in $\hat{\beta}$.*
2. *A sophisticated agent obtains an efficient contract. Due to the monopolistic power of the provider, the agent obtains utility of $\underline{u}$.*
3. *A naïve agent obtains an inefficient contract. The direction of the exploitative distortion is given by:*

$$\frac{d\,r^*}{d\hat{\beta}} \begin{cases} > 0 & \text{for } \theta < 1, \\ = 0 & \text{for } \theta = 1, \\ < 0 & \text{for } \theta > 1 \end{cases}$$

4. *The agent's utility as well as social surplus are decreasing in the degree of naiveté $\hat{\beta}$, while the provider's profits are increasing in $\hat{\beta}$.*

The derivation is provided in the appendix. The first statement establishes that the willingness to pay for a contract is increasing in the offered rate of return for any degree of naiveté. In addition, since a naïve agent overestimates future savings, he would accept a higher fee for

[8] Assuming that the agent has access to a costless savings account with an interest rate of zero would guarantee that $r > 0$ at the provider's optimum, see the discussion in section 2.2 and the online appendix. The second part of Assumption 1 disciplines the overestimation of future saving by a naif by imposing an upper bound on the agent's elasticity of intertemporal substitution, see the proof of Proposition 1. Intuitively, with highly elastic preferences, the overestimation of future savings can be non-monotonic in the interest rate, for example if the actual savings 'catch up' with the predicted savings, bounded from above by Y, once the interest rate is high enough. In the quantitative part of this paper, I set $\delta\hat{\beta} = 0.988$ and $r = 0.05$ per annum, which implies that the theoretical predictions apply to parameter constellations with $\theta > 0.75$.

any given interest rate. This form of exploitation generates the so-called 'distributional effect' (DellaVigna and Malmendier, 2004).

Regarding efficiency, note that the profit-maximising contract maximises the *perceived* social surplus, which coincides with the actual surplus only in the case of a fully sophisticated individual ($\hat{\beta} = \beta$). Then, the contract offered by the provider is efficient. Under monopolistic pricing, the provider is able to extract the entire surplus and the agent obtains utility equal to $\underline{u}$.

By contrast, the provider exploits naiveté about the present bias by distorting the interest rate away from first-best. More precisely, the third statement of Proposition 1 indicates that the profit-maximising interest rate is monotonically increasing in the degree of naiveté for $\theta < 1$, but monotonically decreasing in naiveté for $\theta > 1$. In the knife-edge case of logarithmic utility ($\theta = 1$), the interest rate is independent of $\hat{\beta}$.

What is the intuition behind these qualitative results? In short, since a naïve present-biased agent is over-optimistic about his future propensity to save, the provider takes advantage of the agent's mistake by offering contract terms that appear attractive as long as the individual saves a lot. When the income effect is stronger ($\theta > 1$), a naïve agent finds a cheap savings product with low returns and low fees attractive, because he mistakenly expects to counteract the low return with higher savings and thus benefit from a discounted fee. When the substitution effect is stronger ($\theta < 1$), a naïve present-biased agent finds an expensive savings product attractive, because he mistakenly expects to save more when the rate of return is high, justifying higher charges.

As a result of exploitative contracting, a naïve agent obtains utility below $\underline{u}$. The resulting loss of consumer welfare is increasing (continuously) in the degree of naiveté and arises from two distinct sources. In addition to the distributional effect mentioned above, the distortion of the interest rate away from first-best results in the negative 'efficiency effect' (DellaVigna and Malmendier, 2004).

The dependence of the direction of exploitative distortion on the size of the CRRA parameter highlights the role of interactions between 'classical' (θ) and 'behavioural' ($\hat{\beta}$) preference parameters in generating theoretical predictions. While in their seminal paper DellaVigna and Malmendier (2004) already noted that the direction of the exploitative distortion to two-part-tariff contracts aimed at naïve agents is ambiguous, their setting did not provide a meaningful condition that would pin it down.[9]

Applications. Beyond the choice between stylised savings contracts, the mechanisms under-

[9]Restricting attention to two-part tariffs is without the loss of generality in DellaVigna and Malmendier (2004), where the good is non-divisible and preferences are additive. Subsequent literature has focused primarily on choices from menu contracts to characterise the outcomes attained by sophisticated and naïve agents (Heidhues and Kőszegi, 2018; Spiegler, 2011).

lying Proposition 1 are likely to be relevant in other settings. Most importantly, the trade-off between fixed costs and returns to savings could play a role in the design of multidimensional compensation packages, whereby the employer can offer to match the worker's contributions (up to a specified threshold), thus increasing the effective rate of return. While naïve workers would accept a greater wage reduction in exchange for any match, the curvature of their utility function determines whether this type of overvaluation is exacerbated with an inefficiently 'cheap' or an inefficiently 'generous' match.[10] Similarly, support for subsidies to private savings, financed with lump-sum taxation, depends on the voter's expectation of taking advantage of the subsidy. In this case, naiveté about the present bias might distort the valuation of the subsidy, which realistically benefits only a minority of 'active savers' (Chetty et al., 2014).

Furthermore, even though a lot of attention has been devoted to the effects of automatic enrolment on pension plan participation and contribution rates, there is relatively little research trying to understand the impact of the default investment allocation, or the menu of funds more generally, on the extent of inertia displayed by treated savers. For example, in the seminal paper by Madrian and Shea (2001), the treated employees were automatically contributing 3% of their salary into a money market fund with low returns. Acknowledging alternative reasons for why the employer might offer such a default investment allocation (e.g., potential liability issues), the results underlying Proposition 1 show that as long as a typical worker is characterised by $\theta > 1$ and at least partially naïve, he would overvalue the utility from accepting the default more if the default investment offers low yield. This would only strengthen any co-existing rationales for inertia, such as switching costs.

Finally, the trade-off between fixed costs and higher (expected) returns to savings is present in the model of endogenous financial knowledge accumulation of Lusardi et al. (2017). The complementary results of the current paper help understand when a naïve individual would over- or under-invest in financial literacy.

Identification of the risk aversion parameter. Which of the two cases outlined in Proposition 1 appears empirically more relevant? Although the CRRA utility formulation is one of the most commonly used in the literature, there seems to be little consensus regarding the 'right' size of the CRRA parameter. This is to be expected, as under this parsimonious formulation, a single parameter governs risk aversion as well as elasticity of intertemporal substitution. Nonetheless, most of the available point estimates (derived under both exponential and hyperbolic discount-

[10]Of course, there are other biases that are likely to play a role when assessing the value of a workplace pension. Lack of financial literacy combined with agency problems results in dominated financial products being offered in many plans (Ayres and Curtis, 2014). Furthermore, inertia in pension choices implies that a profit-maximising employer can select default options in a way that does not maximise the benefit accruing to the worker (Bubb and Warren, 2020). As demonstrated in one of the extensions, the interactions between different mistakes can have non-trivial implications for the contract design.

ing) indicate $\theta > 1$, while many numerical life-cycle models assume $\theta > 2$. This would suggest that inefficiently cheap savings products prevail in the market with naïve consumers.[11]

In terms of identifying the key parameter θ, Laibson et al. (2018) jointly estimate the set of preference parameters (β, δ, θ) by matching 12 empirical moments reflecting borrowing and wealth accumulation over the life cycle. The best fit to their data is achieved when setting $\theta = 1.255$ (see below). In a seminal contribution, Gourinchas and Parker (2002) set up a buffer-stock model with one liquid asset, income uncertainty, and a CRRA utility formulation in order to estimate time and risk preferences by matching moments of consumption data. Interestingly, they find either $\theta = 0.514$ or $\theta = 1.397$, depending on the weighting matrix employed in the estimation.[12]

2.2 Extensions

The baseline model analysed above provides the cleanest illustration of the impact that income and substitution effects have on the exploitative contract design. Given the stylised formulation of the model, it is natural to ask whether its main prediction is robust to relaxing some of the simplifying assumptions. In a large majority of cases, the answer is affirmative. Below, I briefly summarise the extensions that are analysed formally in the online appendix.

Simplifying technical assumptions. Assuming a quasi-linear utility function and exogenous valuation of the outside option is common in the literature as it substantially simplifies the exposition. Relaxing quasi-linearity does not affect the qualitative prediction of the model, but it does result in a different threshold $\check{\theta} < 1$, above which the exploitative savings contract is inefficiently cheap, rather than inefficiently expensive. Under endogenous valuation of the agent's outside option, e.g. a costless savings account at a bank, the negative distributional effect associated with naiveté is mitigated, and even turns positive for $\theta > 1$. Nonetheless, the inefficient distortion to contract terms aimed at a naïve agent carries over from the baseline model.

Degree of competition. Assuming perfect competition results in zero-profit pricing and therefore does away with the negative distributional effect. However, the inefficient distortion to the rate of return offered to a naïve agent remains unchanged. That is because the equilibrium con-

[11] Conversely, attempting to infer θ from contract parameters observed in the field would be too heroic, given the long list of plausible confounding factors.

[12] These point estimates correspond to an average, or 'representative', household, although risk attitudes are most likely heterogeneous. For a summary of the existing empirical evidence see Attanasio and Weber (2010). Examples of calibration exercises that assume $\theta > 2$ include Angeletos et al. (2001) and Scholz, Seshadri, and Khitatrakun (2006). For an alternative approach, see Bansal and Yaron (2004) who assume the Epstein and Zin (1989) preferences and separately identify risk aversion and intertemporal elasticity of substitution based on asset market data.

tract again maximises the *perceived* social surplus and competition affects only the distribution of the surplus. Analogously, an identical exploitative distortion arises in a Hotelling model of imperfect competition between two homogeneous providers.

Variable fees. Variable fees, proportional to the agent's contributions, assets, or interest earnings are commonly observed in the pension industry (Dobronogov and Murthi, 2005; Tapia and Yermo, 2008). Importantly, variable fees might make the provider's incentives more aligned with the agent's preferences. For example, with variable fees proportional to the agent's accumulated wealth, the provider has a greater incentive to encourage high levels of saving and to provide a high rate of return. In the online appendix, I analyse the optimal contract design when, in addition to the flat-rate fee, the provider is able to charge a variable fee proportional either to the agent's savings or accumulated assets. The introduction of a variable fee proportional to the agent's savings shifts the optimal rate of return in the direction that encourages higher (actual) savings, independent of naiveté. However, because the naïve agent's valuation of a contract offer is less sensitive to the variable fee, the provider optimally charges a higher variable fee when dealing with a naif. Intuitively, while a naïve agent fully appreciates the direct negative impact of a variable fee on the utility he will derive from the contract, he fails to take into account the indirect negative effect that the fee will have on his future saving.[13] The fact that the optimal variable fee increases with the agent's naiveté only amplifies the magnitude of the baseline exploitative distortion to the rate of return. The same conclusion is reached when the provider charges a variable fee proportional to the agent's accumulated assets.

Pension industry's business model. The profit function assumed above is meant to reflect the incentives of the pension provider in a stylised way, but a plausible alternative view is that the provider's source of revenues is the difference between the realised rate of return on assets under management and the payout promised to the agent. In particular, this would be the case under a defined-benefit (DB) pension arrangement, but not under a DC contract. Accordingly, suppose that the provider can obtain a rate of return r^P on all her investments at a convex cost $\tilde{K}(r^P)$, provide a rate of return $0 \leq r^C \leq r^P$ to the agent, and charge a flat-rate fee f. When the provider invests all capital under management and discounts future profits by $\delta^P \in (0, 1]$, the profit function is:

$$\tilde{\pi} = \delta^P[(1 + r^P)(f + s) - (1 + r^C)s] - \tilde{K}(r^P) = \delta^P[(1 + r^P)f + (r^P - r^C)s] - \tilde{K}(r^P)$$

where s and f are determined by the same constraints as in the baseline problem. Now, in addition to catering to the agent's naiveté, the provider has an incentive to collect a high fee,

[13]More formally, a (fully) naïve agent mistakenly applies the envelope theorem to calculate the impact of the change in contract parameters on the resulting utility from a contract.

which she then re-invests. Similarly to the baseline model, the fee that a monopolistic provider can charge is increasing in r^C (for given $\hat{\beta}$) as well as in $\hat{\beta}$ (for given r^C). For $\theta < 1$, this new incentive is therefore aligned with the baseline exploitative distortion whereby a naif is offered an inefficiently high rate of return. For $\theta > 1$, by contrast, this new incentive is at odds with the baseline exploitative distortion. If the motive to collect high fees due to the possibility of re-investment dominates the original incentive to exacerbate the naïve agent's overvaluation of a contract relative to the provider's cost of service, then the provider would offer an inefficiently high rate of return also for $\theta > 1$. Nevertheless, it cannot be ruled out that the prediction of Proposition 1 carries over to this extension, with the underlying parametric condition reflecting the shape of the provider's cost function.[14]

Financial (il)literacy. A growing body of empirical and theoretical work highlights an important role of financial literacy in economic decision-making and documents a pervasive lack thereof (e.g., Lusardi and Mitchell, 2014; Lusardi et al., 2017). The online appendix outlines an extension of the model that allows for misinterpretation of the contract terms by the agent. More specifically, the provider advertises 'headline' and 'actual' (small-print) contract parameters, either the fee or the interest rate, and a financially unsophisticated agent overestimates the probability of the 'headline' parameters applying. Although introducing financial illiteracy affects the optimal contract design in a predictable way (i.e., the headline terms appear very attractive), the exploitative distortion due to the agent's naiveté about the present bias carries over from the baseline model in a qualitative sense. In case the agent underestimates the fee he will end up paying to the provider, financial illiteracy exacerbates the magnitude of the baseline distortion, but the opposite is true when the agent is overoptimistic about the expected returns. Although preliminary, this analysis suggests that interactions between various biases may have non-trivial implications for financial decision-making and contract design.

Contracts with unused options. Offering menu contracts can create a more pronounced wedge between the agent's anticipated and actual choices, and lower consumer welfare discretely even for small degrees of naiveté (e.g., Heidhues and Kőszegi, 2010).[15] In this paper, however, I

[14]When solving the model, I represent the provider's problem as a choice of the interest rate promised to the agent r^C and the 'wedge' between the total return on the provider's investments and the promised payout, $t \equiv r^P - r^C \geq 0$. For any given r^C, the provider optimally invests in a greater wedge when contracting with a naif willing to pay a high fee. Then, for the original exploitative distortion to carry over for $\theta > 1$, t has to be 'strongly increasing' in $\hat{\beta}$ so that the total rate of return $r^C + t$ is increasing in $\hat{\beta}$, even when r^C decreases. As it turns out, the sensitivity of the optimal wedge to $\hat{\beta}$ depends on the steepness of the provider's marginal cost function. If the marginal cost is sufficiently steep, the incentive to collect a high absolute fee from the naif dominates the original incentive to exacerbate his overvaluation of a contract.

[15]Gottlieb and Zhang (2021) show that this result does not necessarily extend to settings with long enough contracting horizon.

restrict attention to simple contracts without such 'unused options', where the fee and the interest rate are independent of the agent's choice of savings. That is for two reasons. First, private pension products tend to be heavy regulated, which often restricts the contracting space. Non-discriminatory access to pension plans "that have a significant role in providing retirement income and that are significantly subsidised by the state" is one of core principles of private pension regulation according to the OECD (2016). Accordingly, the US Employee Retirement Income Security Act (ERISA) includes a non-discrimination rule which states that qualified workplace pension plans must offer unrestricted access to the same investment options to all employees. In particular, this precludes conditioning the set of investment options available to an employee, or the fees she is paying, on her savings rate.[16] Second, this simplifying assumption enables to link the theoretical analysis of the profit-maximising contract design to an exercise embedding exploitative contracting into a numerical life-cycle framework. Numerical life-cycle models with multiple state variables, such as the one I am employing later on, usually not only assume that the parameters of the economic environment do not depend on the agent's actions, but also take them as exogenously given. Nevertheless, note that the analysis of contracting under financial (il)literacy already sheds some light on the additional effects of allowing the provider to offer menu contracts, albeit in a simplified way. Unsurprisingly, offering a rate of return that is contingent on the agent's savings would only exacerbate the welfare loss from naiveté. Thus, the linear contracts studied here limit the scope for exploitation, at least in the setting of the simple model.

Imperfect observability. I also relax the assumption of perfect observability of the agent's characteristics by the provider. First, when contracting with a population of indistinguishable agents who differ in their actual magnitude of the present bias but hold identical beliefs about the present bias, the provider necessarily offers a pooling contract. As a result of pooling, the agents characterised by a greater degree of naiveté can be either better or worse off, depending on the direction of the original exploitative distortion and the sign of a cross-partial derivative of the provider's cost function.

Second, I characterise the optimal screening contract when the provider is facing a population of indistinguishable agents who differ in their beliefs but are characterised by the same actual present bias, as in Eliaz and Spiegler (2006). This contract offers the naifs the same exploitative interest rate as under perfect observability, but for a lower fee due to the incentive compatibility constraint. Furthermore, the interest rate aimed at sophisticates is distorted away from first-best in a way that creates a larger wedge between the contract terms aimed at the two types. Depending on parameters of the model, in particular the distribution of types and the magnitude

[16] Despite the fact that ERISA imposes fiduciary duty on sponsors of private pension plans, offering a large enough menu of investment options provides them with a legal 'safe harbour' and effectively grants the sponsors a substantial degree of flexibility in designing the menu (Ayres and Curtis, 2014).

of the present bias, the monopolistic provider might find it optimal to offer a pooling contract or exclude sophisticates from the market, rather than screen. What is important, however, is that independent of the provider's contract design strategy, naifs are (weakly) better off, relative to the case of perfect observability, while sophisticates obtain (weakly) less efficient contract terms.

3 Life-cycle model with choice of a pension provider

Building on the qualitative insights of the simple model, in this section I evaluate the quantitative importance of exploitative contracting in a complex setting of a realistically calibrated life-cycle model. The life-cycle model employed here has two main building blocks. First, the household side which is based on the existing work on dynamic decision-making under hyperbolic discounting, both theoretical and quantitative (Angeletos et al., 2001; Harris and Laibson, 2001; Laibson, 1994, 1997; Laibson et al., 2018). Second, the novel firm side which is informed by the empirical literature on operation of pension funds (Basu and Andrews, 2014; Bateman and Mitchell, 2004; Bauer et al., 2010; Bikker and De Dreu, 2009; Bikker et al., 2012).

I begin by outlining an extended theoretical contracting framework that accounts for additional factors that are relevant in the context of a life cycle. Subsequently, I calibrate the model and discuss the impact of naiveté on contractual choice and resulting outcomes, in particular pension wealth accumulation and consumer welfare. Having considered a range of robustness checks, I simulate the impact of three widespread policy interventions on efficiency, consumer welfare, and industry profits, while taking the market response into account.

3.1 Theoretical foundation

Consider the following model of life-cycle savings, preceded by a contracting stage. At the contracting stage, financial providers representing an imperfectly competitive industry simultaneously make their contract offers which specify the rate of return on the agent's retirement savings and the underlying fees. Having selected the provider, if any, the agent goes through his life cycle taking the contract parameters as given.

The structure of the household's consumption-savings problem closely follows the framework of Laibson et al. (2018). An economic agent is alive for a maximum of T periods. Let Y_t for $t \in \{1, 2, ..., T\}$ denote the agent's disposable income in period t. Each period, the agent chooses his current consumption as well as net investments into illiquid and liquid assets. The illiquid asset Z is characterised by a rate of return R^Z, a negative income flow reflecting the underlying fees, and a schedule of age-dependent withdrawal penalties. The agent is not allowed to borrow against the illiquid asset. Thus, the illiquid asset essentially plays a role of a funded pension scheme, whereby the agent chooses his contributions prior to retirement and incrementally withdraws from the asset after retirement. The liquid asset X, on the other hand, imposes no charges or withdrawal penalties. The agent may hold negative amounts of asset X up to a certain limit, which illustrates credit card borrowing. The appreciation rate of the liquid asset thus depends on whether its holdings are positive or negative. Positive holdings earn return R^X, while negative holdings generate borrowing cost of R^{CC}.

In period t, the agent's objective is to maximise the expected discounted lifetime utility:

$$U_t = u(C_t) + \mathbb{E}_t\left[\beta \sum_{s=t+1}^{T} \delta^{s-t} u(C_s)\right],$$

subject to the borrowing and budget constraints spelled out below. Here, $u(C_t) = = n_t \frac{(\frac{C_t}{n_t})^{1-\theta}-1}{1-\theta}$ is an instantaneous CRRA utility function which accounts for household size n_t. The household size changes deterministically throughout the life cycle. The expectation operator is taken with respect to stochastic outcomes, that is survival and disposable income. The agent derives utility from leaving a bequest, which is proportional to the offspring's utility from an infinite stream of consumption that can be financed with assets held at the time of death. For notational simplicity, the bequest motive is suppressed in the above formulation of lifetime utility.[17]

No borrowing on the illiquid asset is captured by the constraint $Z_t \geq 0$. The borrowing limit on the liquid asset is proportional to the average income at age t, which is captured by $X_t \geq -\lambda \bar{Y}_t$ for some $\lambda > 0$. Let I_t^i denote the agent's net investment into asset i in period t. Then, the dynamic budget constraint associated with the illiquid asset is $Z_{t+1} = (1 + R^Z)$ $(Z_t + I_t^Z)$. The dynamic budget constraint associated with the liquid asset is $X_{t+1} = (1 + R)$ $(X_t + I_t^X)$, where $R = R^X$ if $X_t + I_t^X \geq 0$, and $R = R^{CC}$ otherwise. Let κ_t denote the withdrawal penalty associated with the illiquid asset. Then, the static budget constraint implies:

$$C_t = Y_t - I_t^Z - I_t^X + \kappa_t \min(I_t^Z, 0)$$

For illustration, consider a one-asset environment, in which case the choice of a completely naïve decision-maker in period t satisfies the following first-order condition:

$$u'(C_t) \geq \mathbb{E}_t\left[(1 + R^Z)\ \beta\delta\ u'(C_{t+1})\right]$$

mistakenly expecting that all future choices will satisfy:

$$u'(C_{t+s}) \geq \mathbb{E}_{t+s}\left[(1 + R^Z)\ \delta\ u'(C_{t+s+1})\right]$$

[17] Denoting average working-life income by $\bar{Y}$, the utility from leaving a bequest consisting of X_t and Z_t is equal to $b(X_t, Z_t) = \alpha \times \frac{\delta}{1-\delta}[u(\bar{Y} + R^X(\max\{0, X_t + (1-\kappa_t)Z_t\})) - u(\bar{Y})]$, where α is a weight attached to the bequest motive and κ_t is an age-dependent withdrawal penalty.

In other words, a naïve present-biased agent fully gives in to his self-control problem, supposing that the problem will disappear in future periods. Compared to an individual who discounts exponentially, a naïve present-biased agent under-saves in every period and is overoptimistic about future saving, thus overestimating the utilisation of a savings contract. Due to this mechanism, naiveté may again lead to inefficiencies in contractual choice.[18]

Regarding the firm side of the model, suppose that the financial product underlying accumulation of the illiquid asset Z is provided by a pension fund. As mentioned in the previous section, the empirical literature divides total costs of pension funds into administrative and investment costs.[19] The estimated elasticity of administrative costs with respect to assets under management varies from 0.19 to 0.87, reflecting not only cross-country differences in market conditions, but also differences in data quality and definition of administrative costs across studies. Although the investment costs have been shown to increase in factors typically associated with a higher expected rate of return (i.e., the proportion of assets invested in stocks, the number of available investment options, and the 'quality and complexity' of a pension plan), the literature offers arguments both in favour and against scale effects in investment costs. The former include fixed costs and a greater bargaining power of larger pension funds, while the latter refer to limits to arbitrage and the liquidity risk. Attempting to account for these observations, I assume that the provider's total costs in period t are given by:

$$K\left(R^Z, Z_t\right) = k_1 Z_t^{\gamma_1} + k_2 (R^Z - R^X)^{\gamma_2}$$

where $k_1, \gamma_1, k_2, \gamma_2 > 0$. The first term illustrates the administrative cost component. The second term illustrates the investment cost component, which is assumed to be increasing in the difference between the offered rate of return R^Z and a 'risk-free' rate of return on the liquid asset R^X which the agent, and supposedly the provider, may access costlessly. Collected fees f constitute the provider's revenue and thus the per-period profit is $\pi_t = f - K\left(R^Z, Z_t\right)$.

What would be an appropriate way to model competition in this context? On the one hand, industry reports suggest a low degree of market concentration. For example, the 10 largest

[18]The comparison with a behaviour of a sophisticated present-biased agent is not as straightforward. In a setting with at least two periods of saving, present bias generates a strategic conflict between subsequent selves, since the agent's preferences are time-inconsistent. Consequently, the typical approach is to find a pure-strategy subgame-perfect Nash equilibrium of the intrapersonal game (Laibson, 1994, 1997). As there may be multiple equilibria, and the policy functions are often discontinuous, Harris and Laibson (2001) derive an approximate solution called Hyperbolic Euler Equation. Salanié and Treich (2006) use a three-period model to demonstrate that under the CRRA formulation, sophisticates save more than naifs for $\theta > 1$, but the opposite is true for $\theta < 1$. With a logarithmic utility function, savings are independent of naiveté.

[19]For example, Bikker and De Dreu (2009) define investment costs as directly related to asset management (e.g., wages of portfolio managers and analysts, trading costs). Administrative costs include all other operational expenses related to record-keeping, communication with participants, marketing, and compliance with existing legal requirements (e.g., wages, rent).

pension funds held only 8.5% of all private pension assets in the US in 2016, and the 20 largest pension funds in the world managed 17% of all assets (Willis Towers Watson, 2017). At the same time, markets for financial products often feature substantial markups (Beshears et al., 2018; Financial Conduct Authority, 2017).[20]

Given these caveats, I analyse a Hotelling-style model of imperfect competition among homogeneous providers, as in Heidhues and Kőszegi (2010). Two providers, called A and B, are located at endpoints of a unit interval. The providers are otherwise identical. There is a single consumer who is ex ante equally likely to find himself at any point along the interval. The providers simultaneously make their contract offers and the agent evaluates them at time $t = 0$, that is before the beginning of the life cycle. The contracts specify the parameters of the illiquid asset Z, i.e. the rate of return R^Z and the per-period fee f. The contracts are exclusive and not renegotiable, with the parameters binding throughout the entire life cycle. Conditional on his realised location $x \in [0, 1]$, the agent derives the following *perceived* utility from signing a contract with provider A and provider B, respectively:

$$\hat{U}_0(P_A) = V_0(\mathbf{C}(\hat{\beta}, R^Z_A, f_A)) - \xi x$$

$$\hat{U}_0(P_B) = V_0(\mathbf{C}(\hat{\beta}, R^Z_B, f_B)) - \xi(1 - x)$$

where $\hat{U}_0(P)$ is computed by plugging in the expected life-cycle consumption path $\mathbf{C} = (C_1(\hat{\beta}, R^Z) - f, C_2(\hat{\beta}, R^Z) - f, ..., C_T(\hat{\beta}, R^Z) - f)$ into the lifetime utility function with exponential discounting:

$$V_0 = \mathbb{E}_0 \sum_{s=1}^{T} \delta^s u(C_s - f)$$

Thus, as in the simple model, the agent is not directly affected by the present bias at the contract evaluation stage. The expected consumption path $\mathbf{C}$ depends on the agent's future saving behaviour and as such is a function of the offered rate of return R^Z and belief about the present bias $\hat{\beta}$, among other factors. While the utility function is no longer quasi-linear, for computational feasibility I impose that the impact of fees on retirement savings is negligible, i.e. $\frac{dC_t}{df} = 0$.[21] Finally, parameter ξ measures the agent's 'distance aversion'. In the context of a choice of a financial product, this can be interpreted as reflecting the agent's tendency to select a 'default' provider. If the agent rejects both offers, he obtains reservation utility υ corresponding to no access to the illiquid asset.

[20] A plausible interpretation put forward in the literature is that consumers' search costs allow financial providers to price their products above the zero-profit level (Choi, Laibson, and Madrian, 2010; Hortaçsu and Syverson, 2004).

[21] This avoids re-computing household behaviour over an entire life cycle for each fee level separately when calibrating the model. As a robustness check, I verify that this simplifying assumption is in fact innocuous.

The providers maximise the present value of expected profits, discounting future profits using the risk-free rate R^X. Thus, the objective of provider $i = A, B$ is:

$$\mathbb{E}_0\ \pi_i = \mathbb{E}_0\ \sum_{t=1}^{T} \frac{1}{(1+R^X)^t}\ \pi_{i,t}$$

where the expectation is taken with respect to the agent's survival and disposable income. The problem of provider i is then:

$$\max_{R_i^Z, f_i} \quad \mathcal{P} \times \mathbb{E}_0\ \pi_i \quad \text{s.t.:}$$

1. $Z_t = Z_t(\beta, R_i^Z)$
2. $\mathcal{P} = \mathbb{P}(\hat{U}_0(P_i) \geq \underline{u})$
3. $\mathbb{E}_0\ \pi_i \geq 0$

The first constraint says that when computing the agent's wealth accumulation path, the provider takes into account the true present bias parameter β, in addition to the offered rate of return R_i^Z. The second constraint is the individual rationality constraint, which reflects the probability that the agent accepts firm i's offer given his realised location x. In the above, $\underline{u} = \max\ \{\hat{U}_0(P_{-i}), \upsilon\}$ is dependent on the contract parameters offered by a competitor and the value of the outside option. The third constraint assures that the providers enter the market only when they make non-negative expected profits. The following result outlines the theoretical predictions of the extended model.

Proposition 2: *In a symmetric equilibrium of a life-cycle model with choice of a pension provider:*

1. *The contract parameters cater to the agent's naiveté by maximising the perceived consumer-firm surplus.*
2. *The contract parameters change continuously with parameter ξ. For low enough ξ, the equilibrium contracts coincide with those that would be provided by a perfectly competitive market. For high enough ξ, the equilibrium contracts coincide with one that would be provided by a monopolist.*

The derivation is provided in the appendix. Similarly to the simple model, the exploitative contracts offered in equilibrium maximise the perceived consumer-firm surplus for any degree of competition. These contracts are thus inefficiently distorted when the agent is naïve about the present bias. The purpose of the following numerical exercise is to evaluate the quantitative importance of this inefficiency.

As an equilibrium result, the Hotelling-style model spans all degrees of competition according to a single parameter of distance aversion ξ. For low values of ξ, the agent has low disutility from 'shopping around' and thus the provider's market share is sensitive to the offered contract parameters. In the limit, this induces the providers to supply contracts that maximise the agent's percieved utility subject to the zero-profit condition, as a perfectly competitive market would. For higher values of ξ, by contrast, the agent values proximity of the provider and the market shares become less sensitive to the offered contract terms. In the limit, this allows the providers to maximise profits subject to the agent's participation constraint, effectively enjoying monopolistic powers.

3.2 Calibration

To simulate the agent's behaviour, actual and anticipated, I employ a numerical life-cycle framework developed by Laibson et al. (2018). In what follows, I first discuss the household side calibration. Next, I add and calibrate the firm side of the model.

Except for the specification of the illiquid asset, parameters of the agent's economic environment are exogenous and taken from Laibson et al. (2018), who calibrate those to the US data. This includes parameters governing the stochastic income and survival processes as well as deterministic age-dependent household size and retirement age.[22] Positive holdings of the liquid asset X generate return of $R^X = 2.79\%$ per annum, while its negative holdings generate a borrowing cost of $R^{CC} = 11.52\%$.[23]

Under the specification preferred by Laibson et al. (2018), the illiquid asset Z generates no interest earnings, but provides a proportional consumption flow of 5%. In order to make asset Z reflect the properties of private pension wealth, rather than housing, I make the following two modifications to the numerical model. First, I assume that the illiquid asset appreciates over time at a rate R^Z, instead of generating a consumption flow. Second, I change parameters governing the evolution of liquidation costs over time, so that they are high during the agent's working life, but decrease sharply at retirement.[24]

[22] The disposable income process is defined as a sum of a polynomial in age, an AR(1), and a random shock, with the underlying parameters estimated using the PSID data. The income process accounts for transfers, in particular for Social Security which provides a replacement rate of around 0.70 for an average household. Retirement occurs deterministically at the age of 64.

[23] R^X is equal to a long-run average of real yield on AAA bonds. R^{CC} reflects the average quarterly interest rate reported by the Fed, adjusted for the bankruptcy rate and inflation.

[24] In Laibson et al. (2018), the liquidation costs evolve over time according to the function $0.5/(1 + e^{\frac{\text{age}-50}{10}})$, which starts at the value slightly below 0.50 at the age of 20 and then decreases monotonically. By contrast, I assume that the liquidation costs are given by $0.47629/(1 + e^{\frac{\text{age}-60}{3}})$, which starts at the same value initially and

The structural model is used by Laibson et al. (2018) to estimate key preference parameters (β, δ, and θ) of an average household by matching the data on wealth accumulation and credit card borrowing among US households using the Method of Simulated Moments (Gourinchas and Parker, 2002). Under their benchmark specification, the best match to the data is achieved by setting $\beta = 0.5054$, $\delta = 0.9872$, and $\theta = 1.2551$. Given that I slightly modify the economic environment, it is reasonable to ask whether those estimates remain valid. I therefore locally recalibrate the household preference parameters under the new specification of the illiquid asset.[25] The recalibrated model fits the empirical moments well and the parameter estimates are hardly affected, see Table 5 in the appendix. More precisely, I find that the deviation from the empirical moments is now minimised for $\beta = 0.5030$, $\delta = 0.9880$, and $\theta = 1.1051$, which are all well within the margin of statistical error of the original estimates, see Table 3 in Laibson et al. (2018). The recalibrated model under-predicts wealth accumulation early on in the life-cycle, but over-predicts it for households aged 51-60, reflecting the difference in incentives to accumulate illiquid wealth.

The simulated paths of consumption and wealth accumulation correspond to the actual behaviour of a typical household observed in the data. To model the naïve agent's (erroneous) forecast of future wealth accumulation, I follow Laibson et al. (2018) and assume complete naiveté, thus simulating an otherwise identically parametrised household with $\hat{\beta} = 1$.[26]

Introduction and calibration of the firm side of the model constitute the main contribution of the present exercise. The firm side calibration requires specifying the firm's cost function and the Hotelling parameter. I take parameter γ_1, the elasticity of administrative costs, directly from the empirical literature and set it equal to 0.5. The remaining four free parameters $(k_1, k_2, \gamma_2, \xi)$ are calibrated jointly to meet numerical targets regarding the cost-to-assets ratio, share of administrative costs in the total cost, provider's markup, and to target the interest rate offered in equilibrium. As a benchmark case, I target the cost-to-assets ratio of 0.005, the share of administrative costs of 0.50, the markup of 0.20, and the equilibrium interest rate of 5%, calculated at the means where applicable (Basu and Andrews, 2014; Bateman and Mitchell, 2004; Bauer et al., 2010; Bikker and De Dreu, 2009; Bikker et al., 2012; Financial Conduct Authority, 2017; OECD, 2017). While the choice of these target moments is discussed in the online appendix, note that the numerical results are robust to alternative target values.

falls more slowly during the working life, but then drops sharply at retirement. The two functions cross at the retirement age of 64.

[25]This is done by iterating over a fine grid of parameter values, initially centered around the original estimates. The grid is subsequently adjusted to reflect any fit-improving changes to the initial parameter values.

[26]Recall that under (partial) sophistication, the agent's policy functions are often discontinuous. Laibson et al. (2018) and Angeletos et al. (2001) argue that the assumption of complete naiveté improves the numerical stability of the results, while providing similar insights to a model with full sophistication. I explore alternative assumptions about $\hat{\beta}$ in one of the robustness checks.

Table 1: Firm side calibration

Jointly calibrated parameters	**Value**	**Target moment**	**Moment value**
Admin cost multiplier k_1	1.5083	Share of admin costs	0.50
Investment cost elasticity γ_2	5.75	Eqm. interest rate R^Z	5%
Investment cost multiplier k_2	2.649×10^{12}	Cost-to-assets ratio	0.005
Hotelling parameter ξ	0.4815	Markup	0.20
Set parameters	**Value**	**Source**	
Admin cost elasticity γ_1	0.5	Bateman, Mitchell 2004; Bikker et al. 2012	
CRRA parameter θ	1.1051	Author's calibration (Table 5)	
Short-run discount factor β	0.5030	Author's calibration (Table 5)	
Long-run discount factor δ	0.9880	Author's calibration (Table 5)	
Beliefs about present bias $\hat{\beta}$	1	Laibson et al. 2018	
Risk-free interest rate R^X	2.79%	Laibson et al. 2018	

The equilibrium of the model is found numerically in the following way. Using the simulated household behaviour as an input, I allow the providers to offer different rates of return centred around the target value R^{Z*}. Specifically, I consider a grid of 13, equally spaced values $R^Z \in \{3.5\%, 3.75\%, ..., 6.5\%\}$.[27] For each constellation of parameters that satisfies the cost and markup targets at R^{Z*}, I check whether there exist any alternative contract parameters that would increase the agent's *perceived* utility while holding the provider's expected profit at least constant. This includes both price-based deviations (i.e., charging a different fee for R^{Z*}) and return-based deviations (i.e., offering a different R^Z along the provider's iso-profit curve). If there exist no such mutually profitable deviations, then R^{Z*} is indeed offered in equilibrium.

Intuitively, while parameters k_2 and γ_2 jointly determine the cost-to-assets ratio, parameter k_1 is calibrated to match the share of administrative costs and parameter ξ allows to match the markup. To rule out mutually profitable deviations, the provider's cost function needs to be 'steep enough'. Thus, the equilibrium interest rate determines γ_2. Table 1 reports the jointly calibrated firm side parameters, corresponding target moments, and the set parameters.

[27] This spans the range of values considered by Laibson et al. (2018) and appears realistic given average returns to private pension assets, see the online appendix.

Non-targeted moments. The numerical model replicates some non-targeted estimates reported by the empirical literature reasonably well. The equilibrium fee is equal to 0.6% of assets, measured at the mean. This is well within the range of fee levels observed in the pension industry, which are typically between 0.5% and 1.7% (Dobronogov and Murthi, 2005; OECD, 2017; Tapia and Yermo, 2008). The absolute (dollar) value of the calibrated administrative costs, which is equal to \$1511 on average, is somewhat high, but it lies close to the upper end of a range of values reported in the literature.[28] The cost-to-assets ratio, which is calibrated to a value of 0.005 at the means, varies from 0.0035 to 0.0289 over the life cycle, with an average of 0.009, which also falls within the typical range observed across studies (Bateman and Mitchell, 2004; Bauer et al., 2010; Bikker and De Dreu, 2009; Bikker et al., 2012; OECD, 2017).

3.3 Quantitative results

Magnitude of the forecasting error. A naïve present-biased agent is overoptimistic about his future propensity to save and thus utilisation of any savings contract. The magnitude of this forecasting error determines the extent of inefficiency prevalent in equilibrium. Figure 1 plots the agent's expected and actual wealth holdings over the life cycle, averaged over income shocks and conditional on survival. The visible divergence reflects the forecasting error made by a naïve household at the contract evaluation stage.

Several observations can be made. First, the magnitude of the forecasting errors about future wealth holdings is large. At retirement at the age of 64, the expected holdings of illiquid wealth are overestimated by 87%, while the expected holdings of total wealth are overestimated by over 120%. Second, the period-by-period forecasting errors compound. As a result, the total forecasting error increases over time. Third, in addition to levels, a naïve agent also misperceives the composition of his wealth, erroneously expecting to hold some liquid wealth, while in fact he ends up accumulating virtually all wealth in form of the illiquid asset. Note that for an extended interval, the value of the illiquid asset exceeds the value of total wealth. That is because the present-biased household simultaneously accumulates illiquid wealth and borrows on the credit card.[29]

The qualitative predictions of the simple model stem from the observation that the profit-maximising provider offers contract terms that appear attractive as long as the individuals saves

[28]Bikker and De Dreu (2009) note a substantial heterogeneity in administrative costs of the Dutch pension funds, which vary from \$53 to \$1509. These bounds are slightly narrower for a sample of larger pension funds from four countries in Bikker et al. (2012), ranging from \$30 to \$674. For Australia, Bateman and Mitchell (2004) report the range from \$105 to \$897.

[29]Such naiveté-driven patterns are conceptually consistent with survey evidence, according to which over two thirds of respondents consider their current savings rate for retirement to be "too low". Among those who additionally declare the willingness to increase their savings rate within the next few months, only 14% actually do so (Choi, Laibson, Madrian, and Metrick, 2002).

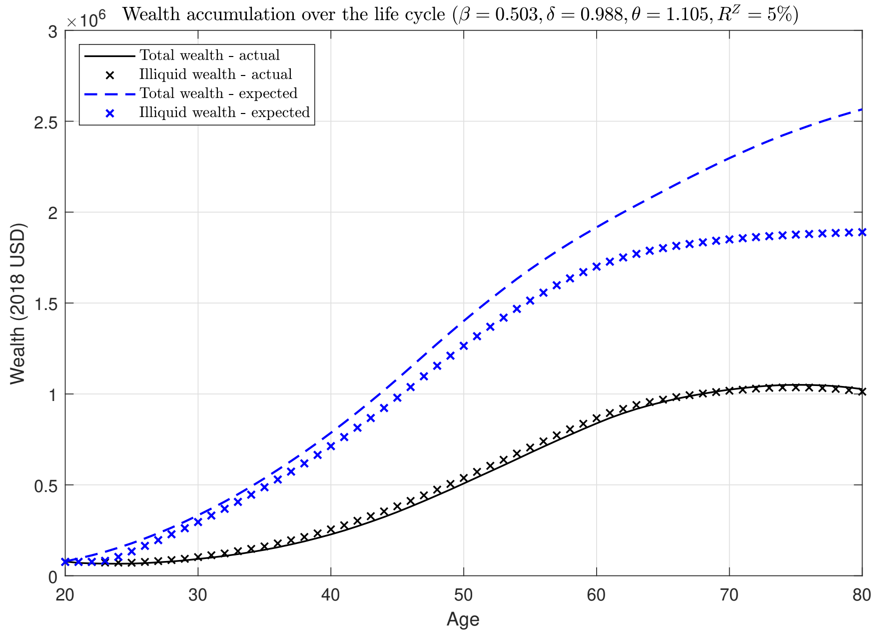

Figure 1: Wealth accumulation paths (expected and actual)

Source: Author's calculations.

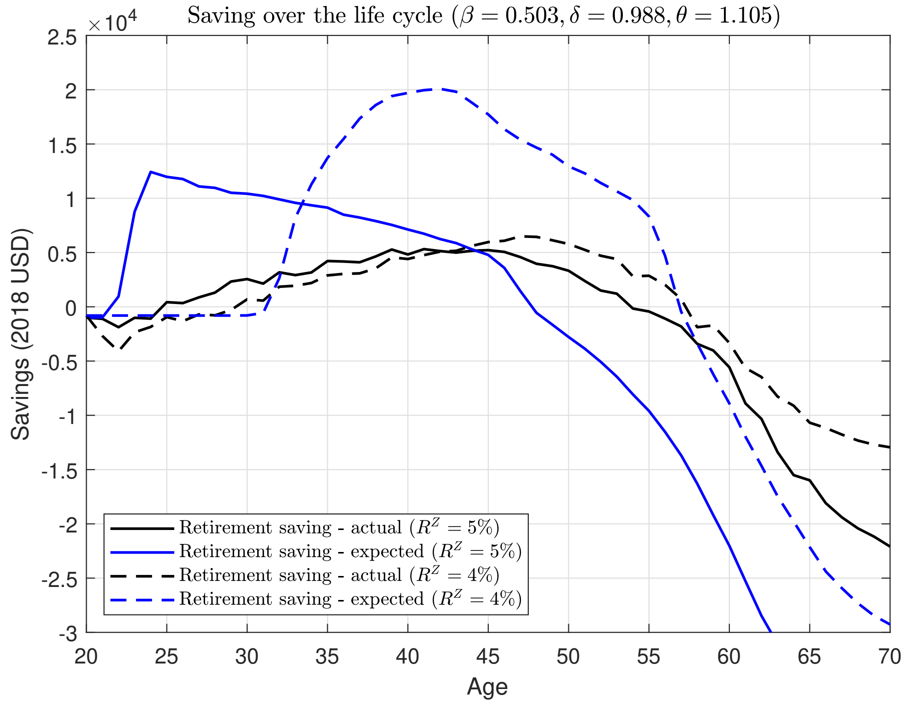

Figure 2: Retirement saving paths (expected and actual)

Source: Author's calculations.

a lot. The exploitative distortions to the rate of return exacerbate the overvaluation of a contract by a naïve agent, relative to the provider's actual cost of service. It is important to realise that in a dynamic environment the relationship between the interest rate and (cumulative) savings is not necessarily monotonic. That is because the agent is deciding not only how much to save but also *when* to save, and any such adjustments to the savings horizon can distort the textbook relationship between the interest rate and cumulative savings. With this caveat in mind, the intuition stemming from the simple model appears to extend to a dynamic life-cycle setting. For the benchmark calibration of the household with $\theta = 1.105$, the misprediction of future saving, measured as the difference between expected and actual average retirement savings, is indeed *decreasing* in the rate of return on the illiquid asset, except for very high rates of return (over 5.75%). An illustration is provided in Figure 2, which plots the actual and expected retirement savings for two different rates of return. The naïve agent's misprediction is clearly amplified under the lower rate of return of 4%, because the actual savings do not adjust upwards as strongly as the agent anticipates, relative to the case with the higher rate of return of 5%.

Properties of the equilibrium contract. According to the target, the savings contract offered in equilibrium provides a rate of return $R^{Z*} = 5\%$. Is this an efficient outcome? When the agent's utility function is not quasi-linear, any social welfare function aggregating the agent's utility and the provider's profits would rely on a questionable comparison between utils and monetary amounts. Consequently, in the context of a life-cycle model my preferred measure of efficiency is a Pareto criterion that asks whether there exist any alternative contract parameters that would improve the agent's *actual* welfare, while keeping the provider's profit at least constant.[30] If such parameters do not exist, I call a contract 'Pareto-efficient'. As long as there exists a Pareto-improving alternative, any other efficiency benchmark based on a social welfare function would also indicate inefficiency, but the converse is not true. The Pareto criterion thus constitutes a conservative efficiency benchmark which avoids attaching arbitrary weights to the outcomes attained by the agent and the provider. The only reason why the equilibrium contract may fail to be Pareto-efficient is the agent's naiveté about the present bias, which leads to erroneous predictions of future utilisation and misevaluation of the offered contract terms.

Equilibrium of the calibrated model indeed turns out to be Pareto-inefficient. A Pareto-improvement could be achieved by an offer of a higher rate of return, namely 5.25%. Thus, the exploitative contract offered in equilibrium is inefficiently cheap, in line with the qualitative prediction of the simple model.

Interestingly, the small difference in contract parameters (5.25% vs. 5%) translates into

[30] More precisely, given the calibration of the provider's cost function, I calculate fees associated with alternative rates of return that would preserve the provider's profit at the current (equilibrium) level to derive the relevant iso-profit curve.

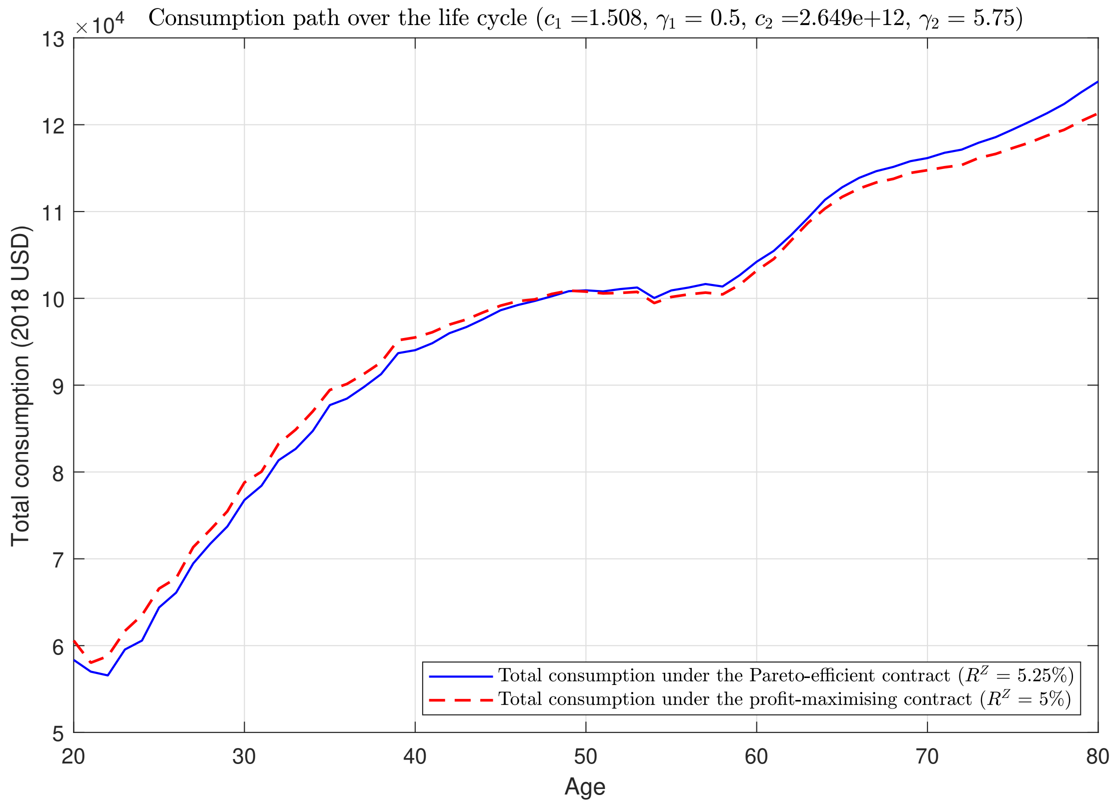

Figure 3: Contract terms and consumption paths
Source: Author's calculations.

substantial differences in pension wealth holdings and retirement consumption. Choosing an inefficiently cheap savings contract ultimately reduces the agent's pension wealth by 8%. This reflects not only the effect of compound interest, but also different consumption and wealth accumulation paths induced by the two contracts.[31] As shown in Figure 3, while the Pareto-efficient contract slightly lowers consumption during the working life, it supports visibly higher consumption in retirement. Choosing this contract would increase the agent's annual consumption in retirement by 3% on average, while lowering his annual consumption during the working life by less than 1%.

A consumption-equivalent (CE) loss of consumer welfare arising from exploitative contracting equals 0.23% per annum. In other words, to be compensated for choosing an inefficient contract, the agent's consumption would have to increase by 0.23% in every period of the model. This corresponds to approximately $230 of consumption foregone annually, which does not increase the provider's profit. To my knowledge, this is the first structural estimate of the quantita-

[31] A simple calculation shows that every dollar saved at the age of 40 earning interest of 5.25% p.a. yields 5.9% higher total returns by the age of 64, compared to interest of 5% p.a.

tive importance of exploitative contracting obtained within a rich life-cycle model. Although the estimated CE welfare loss may appear small, it is not negligible. Its modest size reflects the conservative efficiency criterion and discounting of future outcomes in the lifetime utility function.[32]

Regarding the latter point, keep in mind that the efficient savings contract would support higher consumption in retirement at the expense of slightly lower consumption during the working life. According to the lifetime utility function with *exponential* discounting and $\delta = 0.988$, the agent is indifferent between one util obtained at the beginning of a working life and almost exactly two utils obtained at the beginning of retirement, thus imposing an onerous requirement on all 'more expensive' contracts. Consequently, a substantial difference in the consumption paths induced by the two contracts translates into a modest estimate of the CE welfare loss. Arguably, the consumption-based metric should be more relevant for policy-makers concerned about under-consumption in retirement.

The calibrated model also allows to compare the quantitative impact of contractual choice, holding the agent's naiveté fixed, with the impact of naiveté and present bias itself. For the targeted rate of return of 5%, a sophisticated present-biased household with the same preference parameters β, δ, and θ would accumulate 36% more pension wealth by the time of retirement. Thus, the impact of contractual choice (8%) appears visibly smaller than the impact of naiveté itself. Eradicating the present bias altogether would increase the accumulated pension wealth by 87%, relative to the case of naïve present bias.

3.3.1 Robustness checks

The results discussed above are robust, both in qualitative and in quantitative sense, to changing the target moments, the provider's cost function, the degree of competition, and the fee structure. Alternative assumptions about the household's degree of risk aversion also indicate the prevalence of Pareto-inefficient contract terms, but the quantitative impact of exploitative contracting on savings and consumer welfare depends on the size of the CRRA parameter. Assuming partial naiveté alleviates concerns related to Pareto-inefficiency of market outcomes. The intuition is not straightforward, however, and the model with partial naiveté strengthens the rationale for policy interventions based on distributional concerns. A complete set of the underlying numerical results is relegated to the online appendix.

Sensitivity to target values. The benchmark results are robust to alternative target values

[32] In a recent contribution, Gottlieb and Zhang (2021) study a dynamic problem of optimal (unrestricted) contracting and find that the cost of naiveté materialises only in the final period of the model. Consequently, the associated welfare loss vanishes as the time horizon grows (as long as $\delta < 1$). It is therefore unclear whether the estimate of the welfare loss would necessarily be larger under unrestricted contracting.

and assumed elasticity of administrative costs.

Firm's cost function. I consider two alternative specifications of the provider's cost function:

$$K\left(R^Z, Z_t\right) = k_1 + k_2(R^Z - R^X)^{\gamma_2}$$

and

$$K\left(R^Z, Z_t\right) = k_1 + k_2[(R^Z - R^X)Z(t)]^{\gamma_2}$$

The first alternative replaces the administrative cost component which depends on assets under management with a fixed administrative cost. The second formulation maintains the assumption of fixed administrative cost, but models the investment cost as a function of the agent's capital earnings in excess of the risk-free rate. The numerical results regarding the extent of inefficiency and the impact of exploitative contracting are unaffected by these alternative specifications.[33]

Degree of competition. The implications of the model are practically unaffected when the firm side calibration targets an average markup of zero, illustrating the assumption of perfect competition. The CE measure of loss of consumer welfare associated with exploitative contracting is somewhat larger at 0.26%, rather than 0.23%.

Under another extreme assumption of monopoly, the contract that is offered in equilibrium is simply the one that maximises the provider's profits subject to extracting the entire perceived surplus from the agent. In this case, the agent would be better off paying the monopolistic fee for a much higher rate of return of 6%, rather than 5%. The associated CE measure of loss of consumer welfare is very high at 1.14%. In this case, however, the shift from the inefficiently cheap to the more expensive contract can no longer be interpreted as a Pareto improvement, because the monopolist makes a strictly higher profit when offering the lower rate of return.

Variable fees. Assuming that each savings contract imposes a variable fee equal to 0.5% of assets substantially diminishes the importance of fixed fees. The equilibrium contract is again inefficiently cheap, but accounting for variable fees results in a larger negative impact of exploitative contracting on retirement wealth (13%).

[33] I have also performed a calibration assuming an alternative objective function of the provider, following the theoretical extension of an alternative business model of the pension industry. For parameter constellations that satisfy the cost and markup targets, the optimal (profit-maximising) wedge between the returns of the consumer and the provider turns out to be zero. Intuitively, when the cost function of the provider is steep enough so that a naïve agent is not willing to compensate her for rates of return exceeding 5%, the provider herself also finds the cost of generating the wedge prohibitively high.

Household risk aversion. I follow robustness checks conducted by Laibson et al. (2018) and locally re-calibrate the household side of the model under alternative assumptions of $\theta = 0.5$ and $\theta = 2.0$. Similarly to their results, the calibrated parameter β is decreasing in the assumed CRRA parameter, while the calibrated δ increases. I then repeat the exercise of calibrating the firm side of the model and calculating the impact of exploitative contracting. In either case, the contract offered in equilibrium is Pareto-inefficient, although the impact of exploitative contracting on consumer welfare and wealth accumulation depends on the calibration. Under $\theta = 2.0$, the extent of inefficiency is larger and exploitative contracting is associated with a CE loss of consumer welfare of 0.42%. Under $\theta = 0.5$, the CE welfare loss from exploitative contracting is lower at 0.13%, but because the agent's savings are much more sensitive to the rate of return, choosing an inefficient contract lowers the agent's pension wealth by a staggering 57%. Even though for $\theta < 1$ the substitution effect dominates within any period, due to the timing effects inherent to the dynamic consumption-saving problem, the cumulative savings are not monotonically increasing in R^Z. For that reason, an imperfectly competitive market provides a contract that is inefficiently cheap also in this case. However, note for completeness that a monopolist would again find it profit-maximising to offer an inefficiently expensive contract when interacting with a household parametrised by $\theta = 0.5$, but an inefficiently cheap one when a household is parametrised by $\theta = 2.0$.

Partial naiveté. Following Laibson et al. (2018), the benchmark calibration exercise assumes complete naiveté, i.e. $\hat{\beta} = 1$. I analyse a departure from this assumption by considering two values of $\hat{\beta} < 1$ capturing decreasing degrees of naiveté. This applies both at the contract evaluation stage as well as throughout the life cycle. As it turns out, partial naiveté has very little impact on the agent's actual behaviour.[34] In marked contrast, the expected wealth holdings *increase*, relative to the case of complete naiveté. That is because a sophisticated agent with a moderate present bias would take into account time-inconsistency of his future selves and accumulate a substantially larger proportion of his wealth in the form of an illiquid asset, relative to an agent who discounts exponentially. As a result, the willingness to pay for a savings contract of a partially naïve agent is substantially higher than that of his fully naïve counterpart. Then, in order for the equilibrium contract to offer a rate of return of 5%, the calibration returns a cost function of the provider that is much steeper than in the benchmark case. Consequently, there exist no Pareto-improving contract terms along the provider's iso-profit curve. In sum, even though partial naiveté alleviates concerns related to Pareto-inefficiency of equilibrium outcomes, distributional rationale for a regulatory intervention becomes stronger.

[34]This is not surprising, since the utility function is close to logarithmic (Salanié and Treich, 2006).

Table 2: Ceiling on fees

Fraction of benchmark fee	**Market R^Z**	**Consumer welfare (CE)**	**Provider's profits**
1.0	5%	-	-
0.98	4.75%	-0.89%	-
0.76	4.50%	-2.03%	-
0.62	4.25%	-3.39%	-
⋮			
0.45	3.50%	-6.43%	-
0.40	3.50%	-6.34%	-9%
0.35	3.50%	-6.12%	-27%

Notes: Changes in consumer welfare and provider's profits are calculated relative to the *laissez-faire* outcomes.

3.3.2 Policy interventions

This section analyses the impact of three widespread policy interventions, namely introducing the ceiling on fees, regulating competition, and controlling the liquidity of the pension asset. Each case accounts for the market response to the policy, taking the benchmark calibration of the model as given.

Ceiling on fees. Consider an absolute ceiling on fees, expressed as a fraction of the calibrated equilibrium fee, which applies independently of the offered rate of return. Intuitively, imposing a ceiling on fees can only distort the provider's offers in favour of cheaper savings contracts with low returns.[35] Table 2 presents breakpoints, at which gradually lower interest rates are offered in equilibrium.

Clearly, an effective ceiling on fees cannot resolve the issue of inefficiently cheap contracts prevailing in the market. What is more, in an imperfectly competitive setting, the providers are optimally charging fees strictly below the agent's reservation price. Then, the intervention does not even achieve redistribution from the providers to the agent. Instead, the providers are able to retain their profits by offering low-yielding contracts and charging the highest allowed fees, thus passing the entire loss of efficiency onto the consumer. The ceiling on fees enforces some redistribution only when the providers are already offering the minimal R^Z, i.e. 3.5%.

[35]This is formally confirmed in the setting of the simple model, see Lemma A1.

Table 3: Regulating competition

Markup	**Market** R^Z	**Consumer welfare (CE)**	**Provider's profits**
0	5%	+0.71%	−62%
0.05	5%	+0.53%	−47%
0.10	5%	+0.36%	−31%
0.15	5%	+0.18%	−16%
0.20	5%	-	-
0.25	5%	−0.18%	+16%
0.30	5%	−0.36%	+31%

Notes: Changes in consumer welfare and provider's profits are calculated relative to the *laissez-faire* outcomes.

Regulating competition. Consider an intervention that affects the underlying parameter ξ, effectively changing the markup that the providers can charge at the target interest rate of 5%. Table 3 shows the effects of gradually increasing, or decreasing, the degree of competition. Note that this exercise is different from one of the robustness checks above, in that the firm's cost function is given by the benchmark calibration.

Promoting competition redistributes from the provider to the consumer, although the contract offered in equilibrium is characterised by the same rate of return.[36]

Liquidity of pension wealth. Suppose that the policy-maker regulates the liquidity of private pension wealth by increasing or reducing penalties for early withdrawals. On the one hand, the penalties act as a commitment device against early withdrawals for individuals with a self-control problem. On the other hand, more liquid wealth holdings allow to react to income shocks. In the calibrated model, the provider treats those penalties as exogenously given, i.e. the level of liquidity of pension wealth is not part of the savings contract. This is motivated by the observation that the private market would typically under-provide commitment, thus exacerbating exploitation.[37] In reality, the regulatory policies indeed seem to dictate a minimum

[36]This does not necessarily hold for any calibration of the model. When the agent's utility is concave in fees, his willingness to pay a premium for a higher-yielding contract is not independent of the absolute fee level, in contrast to the case of quasi-linear preferences. In general, changing the degree of competition could have not only distributional, but also efficiency consequences.

[37]A social planner designing the withdrawal penalties would consider the trade-off between commitment and

Table 4: Liquidity of retirement savings

Withdrawal penalty	**Market** R^Z	**Efficient** R^Z	**Consumer welfare (CE)**	**Savings at retirement**	**Provider's profits**
0.25	5.25%	5.25%	+2.58%	−14%	+14%
0.50	5.25%	5.25%	+1.45%	−7%	+18%
0.75	5.25%	5.25%	+0.68%	+1%	+18%
1.0	5%	5.25%	-	-	-
1.25	5%	5%	−0.34%	+6%	−13%

Notes: Changes in consumer welfare and provider's profits are calculated relative to the *laissez-faire* outcomes.

level of commitment (OECD, 2016). In the US, for example, sponsors of the 401(k) pension plans are able to impose specific vesting rules, but nevertheless early withdrawals made before the age of 59.5 are subject to an additional tax of 10%.

Suppose that under different policy regimes the shape of the withdrawal penalty schedule is maintained (see footnote 24), but the level changes relative to the benchmark calibration. Following the intervention, the market provides a contract that maximises the consumer's perceived utility while retaining the provider's profits that are not competed away. Results of this exercise are presented in Table 4.

When pension wealth is made more liquid, a naïve present-biased agent expects to accumulate more of asset Z since withdrawals are not as costly, but for this very reason ends up accumulating less than in the benchmark case. Consequently, the agent's willingness to pay increases and he ends up selecting a more expensive contract. For similar reasons, the providers' profits increase as some of the additional revenues are not competed away.[38]

Importantly, Table 4 indicates that consumer welfare increases monotonically as pension wealth becomes more liquid. This somewhat surprising result arises from the fact that the agent no longer simultaneously accumulates illiquid pension wealth and carries a large credit card debt, suffering from the difference in the interest rates. Quantitatively, this positive effect,

flexibility and, in a typical case, decide against full flexibility (Amador, Werning, and Angeletos, 2006). A profit-maximising provider, by contrast, would recognise that a completely naïve agent has no taste for commitment and offer a fully flexible savings arrangement. Similarly, a partially naïve individual would choose insufficient, but costly, commitment, which would also exacerbate exploitation (Heidhues and Kőszegi, 2009). In a survey paper, Bryan, Karlan, and Nelson (2010) discuss other reasons why private markets might not provide sufficient commitment.

[38]For a given Hotelling parameter ξ, the modified pattern of wealth accumulation by the agent implies that the new equilibrium outcome lies along a different iso-profit curve. That is because the agent's valuation and the provider's costs associated with a particular rate of return are both affected.

together with selection of a more expensive contract, appears to dominate the negative effect of progressively lower retirement savings.[39] More generally, these results highlight a well-known point that the interplay between various forms of wealth should be carefully accounted for when evaluating policy interventions in the domain of household finance.

4 Conclusion

This paper studies contracting between a present-biased individual and a profit-maximising financial provider. In the first part of the paper, I analyse a simplified model to show that the exploitative savings contracts aimed at a naïve present-biased agent are either 'inefficiently cheap' (low-yield, low-fee) or 'inefficiently expensive' (high-yield, high-fee), depending on whether the income or the substitution effect of an interest rate change dominates in the agent's utility function. Intuitively, that is because the profit-maximising provider caters to the agent's naiveté by offering contract terms that appear attractive as long as the individual saves a lot. In the second part of the paper, I embed the interaction with a pension provider in a rich life-cycle framework with hyperbolic discounting (Laibson et al., 2018) in order to assess the quantitative importance of contractual design and choice. The results indicate the prevalence of Pareto-inefficient contract terms in market equilibrium, which lowers the agent's pension wealth by 8% and reduces expected annual consumption in retirement by 3%.

Regarding future research, the model might be extended in a number of ways in order to provide new insights and address remaining policy questions. First, what if the interactions between a firm and a consumer are repeated and the individual is allowed to either renegotiate or sign multiple contracts, as in Gottlieb and Zhang (2021)? How do the results change if the agent learns about the present bias over time, as in Ali (2011), or if the present bias itself is age-dependent, as in Pavoni and Yazici (2017)? Second, the distribution of behavioural characteristics in a population should be carefully considered (Chan, 2017). How do the equilibrium contracts reflect the composition of a population of diversely present-biased and diversely naïve agents? Third, as for the increasingly important workplace pension arrangements, what is the role of intermediation by an employer if her preferences do not necessarily coincide with those of her workers? What may ensure adequate retirement preparation of workers who frequently transition between jobs and may suffer from procrastination?

More broadly, further investigation of the quantitative effects of exploitative contracting in a setting with multiple asset classes should also prove fruitful. Recent literature at the intersection of behavioural and macro-economics examines the impact that behavioural biases may have in the housing market (Attanasio, Kovacs, and Moran, 2021; Schlafmann, 2021), modelling the self-

[39] Since the benchmark penalty schedule was not necessarily socially optimal, this explains the simultaneous improvement in consumer welfare and increase in providers' profits.

control problem using time-consistent temptation preferences (Gul and Pesendorfer, 2001). It would thus be interesting to explore the effects of contractual choice on financial outcomes across the household's balance sheet, conditional on the assumed psychological model. In a similar vein, a richer model including specific regulated savings vehicles would allow to analyse the overall impact of such policy instruments as tax incentives, the timing of taxation, or contribution limits.

Appendix

Lemma 1

A consumption schedule (c_1, c_2), where $c_1 = Y - s$ and $c_2 = (1+r)s$, that maximises U_0 satisfies the following first order condition:

$$u'(Y - s) = \delta\ (1+r)\ u'((1+r)s)$$

The expected saving $\hat{s}$, on the other hand, maximise U_1 parametrised by $\hat{\beta}$, and thus satisfy the following:

$$u'(Y - \hat{s}) = \hat{\beta}\delta\ (1+r)\ u'((1+r)\hat{s})$$

It follows that a naïve agent overestimates his future savings $\hat{s} \equiv s(\hat{\beta}, r)$ for a concave $u(\cdot)$, i.e. $d\hat{s}/d\hat{\beta} > 0$. Consequently, he overvalues any contract $P = (r, f)$. The impact of the agent's beliefs on a valuation of a contract (denoted $\hat{U}_0(P)$ for brevity) is given by:

$$\tfrac{d\hat{U}_0(P)}{d\hat{\beta}} = \tfrac{d\hat{V}_0(P)}{d\hat{\beta}} = \underbrace{\left[-u'(Y - \hat{s}) + \delta(1+r)\, u'((1+r)\hat{s})\right]}_{\geq 0}\ \underbrace{\frac{d\hat{s}}{d\hat{\beta}}}_{>0} \quad \geq 0$$

for $\hat{\beta} \leq 1$. The inequality is strict for $\hat{\beta} < 1$.

$\square$

Proposition 1

Recall the provider's problem specific to a monopolistic market with perfect observability:

$$\max_{r,f}\ \ \pi = f - K(r, s), \text{ s.t.:}$$

1. $s = s(\beta, r)$
2. $\hat{V}_0 - f \geq \underline{u}$

The second constraint is binding at the optimum since $d\pi/df > 0$. Then, the first statement in Proposition 1 follows directly from Lemma 1 and from the fact that $d\,\hat{V}_0/d\,r > 0$.[40]

Substituting for both constraints reduces the problem to:

[40] Although this is obvious under time consistency, demonstrating that $\hat{V}_0$ is strictly increasing in the rate of return under time inconsistency requires that $\frac{d\,\hat{V}_0}{d\,r} = \frac{\partial\,\hat{V}_0}{\partial\,r} + \frac{\partial\,\hat{V}_0}{\partial\,\hat{s}}\,\frac{d\,\hat{s}}{d\,r} > 0$. The possible complication arises from the fact that $\frac{\partial\,\hat{V}_0}{\partial\,\hat{s}} > 0$ for $\hat{\beta} < 1$ and $\frac{d\,\hat{s}}{d\,r} < 0$ for $\theta > 1$. However, one can use the closed-form solution for $s(\hat{\beta}, r)$ derived below to demonstrate that the positive direct effect of an increase in the rate of return always dominates the (possibly) negative indirect effect.

$$\max_r \;\; V_0(s(\hat{\beta}, r), r) - \underline{u} - K(r, s(\beta, r))$$

Define the efficiency criterion as:

$$W(P) = U_0(P) + \pi(P) = V_0(s(\beta, r), r) - f + f - K(r, s(\beta, r)) =$$

$$= V_0(s(\beta, r), r) - K(r, s(\beta, r))$$

An efficient contract thus satisfies:

$$\frac{d\, V_0(s(\beta,r),r)}{dr} = \frac{d\, K(r,s(\beta,r))}{dr}$$

while under the assumption of quasi-linear preferences, efficiency is invariant to the transfer f from the agent and the provider.

For a sophisticated agent, $\hat{\beta} = \beta$ and thus the optimal contract solves:

$$\max_r \;\; V_0(s(\beta, r), r) - \underline{u} - K(r, s(\beta, r))$$

which coincides with the efficiency criterion. This demonstrates that a sophisticated present-biased agent obtains an efficient contract.

For a naïve agent, on the other hand, $\hat{\beta} > \beta$ and the optimal contract solves:

$$\max_r \;\; V_0(s(\hat{\beta}, r), r) - \underline{u} - K(r, s(\beta, r))$$

or:

$$\frac{d\, V_0(s(\hat{\beta},r),r)}{dr} = \frac{d\, K(r,s(\beta,r))}{dr}$$

The above does not coincide with the efficiency criterion as long as

$$\frac{d^2\, V_0(s(\hat{\beta},r),r)}{d\hat{\beta}\; dr} \neq 0$$

For completeness, note that the first-order conditions indeed identify the (global) maxima as long as $V_0(s(\hat{\beta}, r), r)$ is strictly concave in r and $K(r, s(\beta, r))$ is strictly convex. To avoid imposing further conditions on possibly endogenous objects, observe that $V_0(s(\hat{\beta}, r), r)$ is strictly concave for either $\hat{\beta} \to 1$ or $\theta \to 1$, so that the envelope theorem applies and $d\,\hat{V}_0/d\,r = \partial\,\hat{V}_0/\partial\,r$. Similarly, $K(r, s(\beta, r))$ is strictly convex for $\theta \to 1$, where $d\,K(r, s(\beta, r))/dr = \partial\,K(r, s(\beta, r))/\partial r$.

Applying total differentiation to $\hat{V}_0$ yields:

$$\frac{d\, V_0(s(\hat{\beta},r),r)}{dr} = \frac{\partial\, V_0(s(\hat{\beta},r),r)}{\partial r} + \frac{\partial\, V_0(s(\hat{\beta},r),r)}{\partial s(\hat{\beta},r)}\;\frac{d\, s(\hat{\beta},r)}{dr}$$

and then:

$$\frac{d^2\, V_0(s(\hat{\beta},r),r)}{d\hat{\beta}\; dr} = \frac{\partial^2\, \hat{V}_0}{\partial \hat{s}\, \partial r} \underbrace{\frac{d\,\hat{s}}{d\,\hat{\beta}}}_{>0} + \frac{\partial^2\, \hat{V}_0}{\partial\, \hat{s}^2} \frac{d\,\hat{s}}{d\,r} \underbrace{\frac{d\,\hat{s}}{d\,\hat{\beta}}}_{>0} + \underbrace{\frac{\partial\,\hat{V}_0}{\partial\,\hat{s}}}_{\geq 0} \frac{d^2\,\hat{s}}{d\hat{\beta}\,dr}$$

where the inequalities follow from the derivation of Lemma 1.

Under the CRRA utility formulation, $\hat{s}$ has a closed form solution given by the following first order condition:

$$u'(Y - \hat{s}) = \hat{\beta}\delta\; (1 + r)\; u'((1 + r)\hat{s})$$

where $u(x) = \frac{x^{1-\theta}-1}{1-\theta}$ and $\theta > 0$. Then:

$$\hat{s} = \frac{1}{1+(\hat{\beta}\delta)^{\frac{-1}{\theta}}(1+r)^{\frac{\theta-1}{\theta}}} \times Y$$

Differentiating shows that $d\,\hat{s}/d\,\hat{\beta} > 0$, while

$$\frac{d\,\hat{s}}{d\,r} \begin{cases} > 0 & \text{for } \theta < 1 \\ = 0 & \text{for } \theta = 1 \\ < 0 & \text{for } \theta > 1 \end{cases}$$

in line with discussion in the main body of the paper regarding the relative strengths of the substitution and income effects, conditional on the CRRA parameter θ.

Taking into account that

$$\hat{V}_0 = \frac{[Y-\hat{s}]^{1-\theta}-1}{1-\theta} + \delta\frac{[(1+r)\hat{s}]^{1-\theta}-1}{1-\theta}$$

the following can be shown using algebraic rearrangements:

$$\frac{\partial^2\, \hat{V}_0}{\partial \hat{s}\, \partial r} \begin{cases} > 0 & \text{for } \theta < 1 \\ = 0 & \text{for } \theta = 1 \\ < 0 & \text{for } \theta > 1 \end{cases}$$

$$\frac{\partial^2\, \hat{V}_0}{\partial\, \hat{s}^2} < 0$$

$$\text{and } \frac{d^2\,\hat{s}}{d\hat{\beta}\,dr} \begin{cases} > 0 & \text{for } \theta < 1 \\ = 0 & \text{for } \theta = 1 \\ < 0 & \text{for } \theta > 1 \end{cases} \quad \text{for } \theta > \underline{\theta} \text{ (Assumption 1).}^{41}$$

[41]Specifically, using a Taylor expansion around $\theta = 1$, $\frac{d^2\,\hat{s}}{d\hat{\beta}\,dr}$ has the sign outlined above provided that $\theta >$

Therefore:

$$\frac{d^2\,V_0(s(\hat\beta,r),r)}{d\hat\beta\;dr} = \underbrace{\frac{\partial^2\,\hat V_0}{\partial\hat s\,\partial r}\,\frac{d\,\hat s}{d\,\hat\beta}}_{\begin{cases}>0 & \text{for } \theta<1\\ <0 & \text{for } \theta>1\end{cases}} + \underbrace{\frac{\partial^2\,\hat V_0}{\partial\,\hat s^2}\,\frac{d\,\hat s}{d\,r}\,\frac{d\,\hat s}{d\,\hat\beta}}_{\begin{cases}<0 & \text{for } \theta<1\\ >0 & \text{for } \theta>1\end{cases}} + \underbrace{\frac{\partial\,\hat V_0}{\partial\,\hat s}\,\frac{d^2\,\hat s}{d\hat\beta\,dr}}_{\begin{cases}>0 & \text{for } \theta<1\\ <0 & \text{for } \theta>1\end{cases}}$$

Importantly, for every parameter constellation that has $\hat\beta \leq 1$,

$$\left|\frac{\partial^2\,\hat V_0}{\partial\hat s\,\partial r}\right| \geq \left|\frac{\partial^2\,\hat V_0}{\partial\,\hat s^2}\,\frac{d\,\hat s}{d\,r}\right|$$

which finally yields:

$$\frac{d^2\,V_0(s(\hat\beta,r),r)}{d\hat\beta\;dr}\begin{cases}>0 & \text{for } \theta<1\\ =0 & \text{for } \theta=1\\ <0 & \text{for } \theta>1\end{cases}$$

Intuitively, the above determines the direction of changes to the interest rate that exacerbate the overvaluation of a contract by a naïve agent, captured by $d\,V_0(s(\hat\beta,r),r)/d\,\hat\beta$. This condition, combined with the assumption of convexity of $K(\cdot,s)$, implies the main result regarding the direction of the exploitative distortion $d\,r^*/d\hat\beta$.

Regarding efficiency, it is evident that while the contract for a sophisticated agent maximises social surplus, the contract for a naïve agent is inefficiently distorted, and the more so the greater the degree of naiveté. Thus social surplus $W(P^*)$ is decreasing in $\hat\beta$, holding β fixed.

At the optimum, the utility attained by the agent is:

$$U_0(P^*) = V_0(s(\beta,r^*),r^*) - f^* = V_0(s(\beta,r^*),r^*) - V_0(s(\hat\beta,r^*),r^*) + \underline{u}$$

For sophisticated agents, $\hat\beta=\beta$ and thus $U_0(P^*)=\underline{u}$ due to the provider's monopolistic power. Using total differentiation of $U_0(P^*)$:

$1+\frac{\log(\hat\beta\delta)}{\log(1+r)}$. This is clearly satisfied for any $\theta > 1-\epsilon$ for some $\epsilon > 0$ provided that $\hat\beta\delta < 1$ and $r > 0$. On the other hand, the condition is not satisfied under a combination of low θ and high $\hat\beta\delta$. Intuitively, when the intertemporal elasticity of substitution is large and beliefs are close to complete naiveté, a higher interest rate can make a naif overestimate her future savings to a lesser extent even though the substitution effect dominates. That is because the difference between savings predicted by agents differing in their naiveté eventually 'flattens out'. For example, in the extreme case of linear preferences, i.e. $\theta \to 0$, an agent would shift her entire consumption to period 2 provided that the interest rate is high enough relative to the discount factor. As the interest increases, the prediction error of a naif eventually disappears as the actual savings start coinciding with predicted savings (which are bounded from above by Y).

$$\frac{d\,U_0(P^*)}{d\,\hat\beta} = \frac{d\,V_0(P^*)}{d\,\hat\beta} - \frac{d\,\hat V_0(P^*)}{d\,\hat\beta} = \frac{\partial\,V_0}{\partial r^*}\,\frac{d\,r^*}{d\,\hat\beta} - \frac{\partial\,\hat V_0}{\partial r^*}\,\frac{d\,r^*}{d\,\hat\beta} - \frac{\partial\,\hat V_0}{\partial \hat s}\,\frac{d\,\hat s}{d\,\hat\beta} =$$

$$= \underbrace{-\frac{\partial\,\hat V_0}{\partial \hat s}\,\frac{d\,\hat s}{d\,\hat\beta}}_{\text{the `distributional effect'}} + \underbrace{\{\frac{\partial\,V_0}{\partial r^*} - \frac{\partial\,\hat V_0}{\partial r^*}\}\,\frac{d\,r^*}{d\,\hat\beta}}_{\text{the `efficiency effect'}}$$

Given the derivations above, both effects decrease the agent's utility.

Finally, it is straightforward to notice that the provider's profits are also increasing in $\hat\beta$, as $f^* = [V_0(s(\hat\beta, r), r) - \underline{u}]$ is increasing in $\hat\beta$ for every r. The provider's optimal choice of r^* only adds to this effect.

$\square$

Proposition 2

Consider the problem from the perspective of provider A:

$$\max_{R_A^Z, f_A} \quad \mathcal{P} \times \mathbb{E}_0 = \pi_A \quad \text{s.t.:}$$

1. $Z_t = Z_t(\beta, R_A^Z)$
2. $\mathcal{P} = \mathbb{P}(U_0(R_A^Z, f_A, \hat\beta) \geq \underline{u})$
3. $\mathbb{E}_0\ \pi_A \geq 0$

which can be written more concisely as:

$$\max_{R_A^Z, f_A} \quad \mathbb{P}(U_0(R_A^Z, f_A, \hat\beta) \geq \underline{u}) \times \mathbb{E}_0\ \pi_A(R_A^Z, f_A, \beta)$$

provided that $\mathbb{E}_0\ \pi_A \geq 0$ at the optimum.

Write the individual rationality constraint as:

$$\mathbb{P}[U_0(R_A^Z, f_A, \hat\beta) \geq \underline{u}] = \mathbb{P}[U_0(R_A^Z, f_A, \hat\beta) \geq U_0(R_B^Z, f_B, \hat\beta)] =$$

$$= \mathbb{P}[V_0(\mathbf{C}(\hat\beta, R_A^Z, f_A)) - \xi x \geq V_0(\mathbf{C}(\hat\beta, R_B^Z, f_B)) - \xi(1-x)] =$$

$$= \mathbb{P}[x \leq 0.5 + \tfrac{V_0(\mathbf{C}(\hat\beta, R_A^Z, f_A)) - V_0(\mathbf{C}(\hat\beta, R_B^Z, f_B))}{2\xi}] =$$

$$= 0.5 + \tfrac{V_0(\mathbf{C}(\hat\beta, R_A^Z, f_A)) - V_0(\mathbf{C}(\hat\beta, R_B^Z, f_B))}{2\xi}$$

where the last line follows from the assumption that x is distributed uniformly over $[0,1]$. This allows to write the problem of provider A as:

$$\max_{R_A^Z, f_A} \quad \{0.5 + \tfrac{V_0(\mathbf{C}(\hat{\beta}, R_A^Z, f_A)) - V_0(\mathbf{C}(\hat{\beta}, R_B^Z, f_B))}{2\xi}\} \times \mathbb{E}_0 \ \pi_A(R_A^Z, f_A, \beta)$$

Denoting the objective function by Π, this yields the following first-order conditions:

$$\tfrac{\partial \Pi}{\partial R_A^Z} = \tfrac{\partial V_0(\mathbf{C}(\hat{\beta}, R_A^Z, f_A))}{\partial R_A^Z} \tfrac{1}{2\xi} \ \mathbb{E}_0 \ \pi_A(R_A^Z, f_A, \beta) + \{0.5 + \tfrac{V_0(\mathbf{C}(\hat{\beta}, R_A^Z, f_A)) - V_0(\mathbf{C}(\hat{\beta}, R_B^Z, f_B))}{2\xi}\} \tfrac{\partial \mathbb{E}_0 \ \pi_A(R_A^Z, f_A, \beta)}{\partial R_A^Z} = 0$$

$$\tfrac{\partial \Pi}{\partial f_A} = \tfrac{\partial V_0(\mathbf{C}(\hat{\beta}, R_A^Z, f_A))}{\partial f_A} \tfrac{1}{2\xi} \ \mathbb{E}_0 \ \pi_A(R_A^Z, f_A, \beta) + \{0.5 + \tfrac{V_0(\mathbf{C}(\hat{\beta}, R_A^Z, f_A)) - V_0(\mathbf{C}(\hat{\beta}, R_B^Z, f_B))}{2\xi}\} \tfrac{\partial \mathbb{E}_0 \ \pi_A(R_A^Z, f_A, \beta)}{\partial f_A} = 0$$

which give the optimal relationship between the rate of return R_A^Z and the fee f_A:

$$\tfrac{\partial V_0(\mathbf{C}(\hat{\beta}, R_A^Z, f_A))}{\partial R_A^Z} / \tfrac{\partial \mathbb{E}_0 \ \pi_A(R_A^Z, f_A, \beta)}{\partial R_A^Z} = \tfrac{\partial V_0(\mathbf{C}(\hat{\beta}, R_A^Z, f_A))}{\partial f_A} / \tfrac{\partial \mathbb{E}_0 \ \pi_A(R_A^Z, f_A, \beta)}{\partial f_A}$$

This optimal relationship, say $f(R^Z)$, arises from the trade-off between the probability of a contract offer being accepted and profits from an accepted contract. Note the following implications. First, $f(R^Z)$ is independent of the parameter of distance aversion ξ. When disutility from 'shopping around' is linear, the pricing of contracts with different rates of return is independent of ξ (the optimal contract may nonetheless depend on ξ). Second, the optimal fee $f(R^Z)$ is determined by the agent's perceived utility from accepting a contract. Third, when the providers are homogeneous, they optimally charge the same fee for the same interest rate.

The above allows to rewrite the problem of provider A as a single-variable optimisation problem:

$$\max_{R_A^Z} \quad \{0.5 + \tfrac{V_0(\mathbf{C}(\hat{\beta}, R_A^Z, f(R_A^Z))) - V_0(\mathbf{C}(\hat{\beta}, R_B^Z, f_B))}{2\xi}\} \times \mathbb{E}_0 \ \pi_A(R_A^Z, f(R_A^Z), \beta) =$$

$$= \{0.5 + \tfrac{V_0(\mathbf{C}(\hat{\beta}, R_A^Z, f(R_A^Z))) - V_0(\mathbf{C}(\hat{\beta}, R_B^Z, f_B))}{2\xi}\} \times \mathbb{E}_0 \ \sum_{t=1}^{T} \frac{1}{(1+R^X)^t} \ [f(R_A^Z) - K\,(R_A^Z, Z_t)]$$

which yields the following first-order condition:

$$\tfrac{d\,\Pi}{d\,R_A^Z} = \tfrac{1}{2\xi} \ [\tfrac{\partial V_0(\mathbf{C}(\hat{\beta}, R_A^Z, f(R_A^Z))}{\partial R_A^Z} + \tfrac{\partial V_0(\mathbf{C}(\hat{\beta}, R_A^Z, f(R_A^Z))}{\partial f(R_A^Z)} \tfrac{d\,f(R_A^Z)}{d\,R_A^Z}] \ \mathbb{E}_0 \ \pi_A(R_A^Z, f(R_A^Z), \beta) +$$

$$+ \ \mathbb{P}[U_0(R_A^Z, f(R_A^Z), \hat{\beta}) \geq \underline{u}] \ [\mathbb{E}_0 \ \sum_{t=1}^{T} \frac{1}{(1+R^X)^t} \ (\frac{d\,f(R_A^Z)}{d\,R_A^Z} - \frac{\partial\,K(R_A^Z, Z_t)}{\partial\,R_A^Z} - \frac{\partial\,K(R_A^Z, Z_t)}{\partial\,Z_t} \frac{d\,Z_t}{d\,R_A^Z})] = 0$$

This demonstrates the following. First, homogeneous providers optimally offer the same contract terms $P_A^* = P_B^* = (R^{Z*}, f(R^{Z*}))$. Second, as in the simple model, the exploitative contracts cater to the agent's naiveté by equalising the marginal perceived benefit to the consumer with the marginal actual cost of the provider (see the second square bracket above).

Finally, the optimal contract parameters change continuously with parameter ξ. For low enough ξ, the providers' offers maximise the agent's perceived utility (subject to the zero-profit constraint binding), just as would be the case under perfect competition. That is because as $\xi \to 0^+$, the term in the upper line dominates, calling for maximisation of the consumer's perceived utility, regardless of the provider's profit. For high ξ, by contrast, the providers' market shares are insensitive to the offered contract terms and thus the providers maximise their expected profit (subject to the consumer's participation constraint binding), just as a monopolist would. That is because as $\xi \to \infty$, the term in the upper line vanishes, leading to maximisation of the provider's profit, conditional on acceptance of the contract.

$\square$

Quantitative analysis

Table 5: Household side moments and goodness of fit

	Data	**LMRT (2018)**	**Fixed pref. params**	**Recalibrated pref. params**
Calibrated β	-	0.5054	0.5054	0.5030
Calibrated δ	-	0.9872	0.9872	0.9880
Calibrated θ	-	1.2551	1.2551	1.1051
Frac. borrowing, 21-30	0.815	0.598	0.610	0.608
Frac. borrowing, 31-40	0.782	0.607	0.653	0.714
Frac. borrowing, 41-50	0.749	0.586	0.792	0.838
Frac. borrowing, 51-60	0.659	0.568	0.840	0.853
Avg. debt to income, 21-30	0.199	0.232	0.232	0.241
Avg. debt to income, 31-40	0.187	0.237	0.254	0.272
Avg. debt to income, 41-50	0.261	0.216	0.297	0.319
Avg. debt to income, 51-60	0.276	0.196	0.328	0.323
Avg. wealth to income, 21-30	1.23	1.30	0.78	1.00
Avg. wealth to income, 31-40	1.86	1.83	1.06	1.47
Avg. wealth to income, 41-50	3.24	2.94	2.47	3.09
Avg. wealth to income, 51-60	5.34	5.05	5.27	6.27
Goodness-of-fit	-	250.75	276.28	221.34

Source: Author's calculations. Column 'LMRT (2018)' uses the original simulation code to replicate Table 3 in Laibson et al. (2018). Column 'Fixed preference parameters' shows the fit of the model to the data under the modified economic environment of the present paper, taking the parameter estimates from Laibson et al. (2018). Column 'Recalibrated preference parameters' demonstrates the results of recalibrating household preference parameters under the modified economic environment. Goodness-of-fit is measured as a weighted sum of squared deviations of the simulated moments from the data moments. All other parameter values are the same as in Laibson et al. (2018).

References

ALI, S. N. (2011): "Learning Self-Control," *The Quarterly Journal of Economics*, 126, 857–893.

AMADOR, M., I. WERNING, AND G.-M. ANGELETOS (2006): "Commitment vs. Flexibility," *Econometrica*, 74, 365–396.

AMERIKS, J., A. CAPLIN, J. LEAHY, AND T. TYLER (2007): "Measuring Self-Control Problems," *American Economic Review*, 97, 966–972.

ANGELETOS, G.-M., D. LAIBSON, A. REPETTO, J. TOBACMAN, AND S. WEINBERG (2001): "The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation," *Journal of Economic Perspectives*, 15, 47–68.

ANTOLÍN, P., S. PAYET, E. WHITEHOUSE, AND J. YERMO (2011): *The Role of Guarantees in Defined Contribution Pensions*, OECD Publishing, Paris.

ASHRAF, N., D. KARLAN, AND W. YIN (2006): "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines," *The Quarterly Journal of Economics*, 121, 635–672.

ATTANASIO, O., A. KOVACS, AND P. MORAN (2021): "Temptation and Incentives to Wealth Accumulation," *NBER Working Paper No. 28938.*

ATTANASIO, O. P. AND G. WEBER (2010): "Consumption and Saving: Models of Intertemporal Allocation and Their Implications for Public Policy," *Journal of Economic Literature*, 48, 693–751.

AYRES, I. AND Q. CURTIS (2014): "Beyond Diversification: The Pervasive Problem of Excessive Fees and Dominated Funds in 401(k) Plans," *Yale Law Journal*, 124, 1476.

BANSAL, R. AND A. YARON (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *The Journal of Finance*, 59, 1481–1509.

BASU, A. AND S. ANDREWS (2014): "Asset Allocation Policy, Returns and Expenses of Superannuation Funds: Recent Evidence Based on Default Options," *Australian Economic Review*, 47, 63–77.

BATEMAN, H. AND O. S. MITCHELL (2004): "New Evidence on Pension Plan Design and Administrative Expenses: The Australian Experience," *Journal of Pension Economics & Finance*, 3, 63–76.

BAUER, R., M. CREMERS, AND R. FREHEN (2010): "Pension Fund Performance and Costs: Small Is Beautiful," *Working Paper.*

BENARTZI, S. AND R. THALER (2007): "Heuristics and Biases in Retirement Savings Behavior," *Journal of Economic Perspectives*, 21, 81–104.

Beshears, J., J. J. Choi, D. Laibson, and B. C. Madrian (2018): "Behavioral Household Finance," in *Handbook of Behavioral Economics: Applications and Foundations 1*, Elsevier, vol. 1, 177–276.

Bikker, J. A. and J. De Dreu (2009): "Operating Costs of Pension Funds: The Impact of Scale, Governance, and Plan Design," *Journal of Pension Economics & Finance*, 8, 63–89.

Bikker, J. A., O. W. Steenbeek, and F. Torracchi (2012): "The Impact of Scale, Complexity, and Service Quality on the Administrative Costs of Pension Funds: A Cross-Country Comparison," *Journal of Risk and Insurance*, 79, 477–514.

Browning, M. and A. Lusardi (1996): "Household Saving: Micro Theories and Micro Facts," *Journal of Economic Literature*, 34, 1797–1855.

Bryan, G., D. Karlan, and S. Nelson (2010): "Commitment Devices," *Annual Review of Economics*, 2, 671–698.

Bubb, R. and P. L. Warren (2020): "An Equilibrium Theory of Retirement Plan Design," *American Economic Journal: Economic Policy*, 12, 22–45.

Campbell, J. Y. (2016): "Restoring Rational Choice: The Challenge of Consumer Financial Regulation," *American Economic Review*, 106, 1–30.

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, 1639–1674.

Chan, M. K. (2017): "Welfare Dependence and Self-Control: An Empirical Analysis," *The Review of Economic Studies*, 84, 1379–1423.

Chetty, R., J. N. Friedman, S. Leth-Petersen, T. H. Nielsen, and T. Olsen (2014): "Active vs. Passive Decisions and Crowd-Out in Retirement Savings Accounts: Evidence from Denmark," *The Quarterly Journal of Economics*, 129, 1141–1219.

Choi, J. J., D. Laibson, and B. C. Madrian (2010): "Why Does the Law of One Price Fail? An Experiment on Index Mutual Funds," *The Review of Financial Studies*, 23, 1405–1432.

Choi, J. J., D. Laibson, B. C. Madrian, and A. Metrick (2002): "Defined Contribution Pensions: Plan Rules, Participant Choices, and the Path of Least Resistance," *Tax Policy and the Economy*, 16, 67–113.

——— (2003): "Optimal Defaults," *American Economic Review*, 93, 180–185.

Choi, J. J., D. Laibson, B. C. Madrian, A. Metrick, and J. M. Poterba (2004): "For Better or for Worse: Default Effects and 401(k) Savings Behavior," in *Perspectives on the Economics of Aging*, University of Chicago Press.

DellaVigna, S. and U. Malmendier (2004): "Contract Design and Self-Control: Theory and Evidence," *The Quarterly Journal of Economics*, 119, 353–402.

Diamond, P. and B. Köszegi (2003): "Quasi-Hyperbolic Discounting and Retirement," *Journal of Public Economics*, 87, 1839–1872.

Dobronogov, A. and M. Murthi (2005): "Administrative Fees and Costs of Mandatory Private Pensions in Transition Economies," *Journal of Pension Economics & Finance*, 4, 31–55.

Eliaz, K. and R. Spiegler (2006): "Contracting with Diversely Naive Agents," *The Review of Economic Studies*, 73, 689–714.

Epstein, L. G. and S. E. Zin (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57, 937–969.

Financial Conduct Authority (2017): *Asset Management Market Study - Final Report*, FCA, London.

Goda, G. S., M. Levy, C. F. Manchester, A. Sojourner, and J. Tasoff (2019): "Predicting Retirement Savings Using Survey Measures of Exponential-Growth Bias and Present Bias," *Economic Inquiry*, 57, 1636–1658.

Gottlieb, D. and K. Smetters (2021): "Lapse-Based Insurance," *American Economic Review*, 111, 2377–2416.

Gottlieb, D. and X. Zhang (2021): "Long-Term Contracting with Time-Inconsistent Agents," *Econometrica*, 89, 793–824.

Gourinchas, P.-O. and J. A. Parker (2002): "Consumption over the Life Cycle," *Econometrica*, 70, 47–89.

Gul, F. and W. Pesendorfer (2001): "Temptation and Self-Control," *Econometrica*, 69, 1403–1435.

Harris, C. and D. Laibson (2001): "Dynamic Choices of Hyperbolic Consumers," *Econometrica*, 69, 935–957.

Heidhues, P., B. Kőszegi, and T. Murooka (2021): "Procrastination Markets," *Working Paper.*

Heidhues, P. and B. Kőszegi (2009): "Futile Attempts at Self-Control," *Journal of the European Economic Association*, 7, 423–434.

——— (2010): "Exploiting Naivete about Self-Control in the Credit Market," *American Economic Review*, 100, 2279–2303.

——— (2018): "Behavioral Industrial Organization," in *Handbook of Behavioral Economics: Applications and Foundations 1*, Elsevier, vol. 1, 517–612.

Hortaçsu, A. and C. Syverson (2004): "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds," *The Quarterly Journal of Economics*, 119, 403–456.

Kőszegi, B. (2014): "Behavioral Contract Theory," *Journal of Economic Literature*, 52, 1075–1118.

Laibson, D. (1994): "Self-Control and Saving," *Massachusets Institute of Technology mimeo.*

——— (1997): "Golden Eggs and Hyperbolic Discounting," *The Quarterly Journal of Economics*, 112, 443–478.

Laibson, D., P. Maxted, A. Repetto, and J. Tobacman (2018): "Estimating Discount Functions with Consumption Choices with the Lifecycle," *Working Paper*, [Available online at: https://www.dropbox.com/s/khoyt5hz9ci1qfy/Estimating%20Discount%20Functions.pdf?dl=0].

Lusardi, A., P.-C. Michaud, and O. S. Mitchell (2017): "Optimal Financial Knowledge and Wealth Inequality," *Journal of Political Economy*, 125, 431–477.

Lusardi, A. and O. S. Mitchell (2014): "The Economic Importance of Financial Literacy: Theory and Evidence," *Journal of Economic Literature*, 52, 5–44.

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*, 116, 1149–1187.

O'Donoghue, T. and M. Rabin (1999a): "Doing It Now or Later," *American Economic Review*, 89, 103–124.

——— (1999b): "Procrastination in Preparing for Retirement," in *Behavioral Dimensions of Retirement Economics*, Brookings Institution Press and Russell Sage Foundation, 125–160.

——— (2001): "Choice and Procrastination," *The Quarterly Journal of Economics*, 116, 121–160.

OECD (2016): *Core Principles of Private Pension Regulation*, OECD Publishing, Paris.

——— (2017): *Pensions at a Glance 2017*, OECD Publishing, Paris.

——— (2021): *Pensions at a Glance 2021*, OECD Publishing, Paris.

Office of Fair Trading (2014): *Defined Contribution Workplace Pension Market Study*, The National Archives, London.

Pavoni, N. and H. Yazici (2017): "Optimal Life-Cycle Capital Taxation under Self-Control Problems," *The Economic Journal*, 127, 1188–1216.

Salanié, F. and N. Treich (2006): "Over-Savings and Hyperbolic Discounting," *European Economic Review*, 50, 1557–1570.

Schlafmann, K. (2021): “Housing, Mortgages, and Self-Control,” *The Review of Financial Studies*, 34, 2648–2687.

Scholz, J. K., A. Seshadri, and S. Khitatrakun (2006): “Are Americans Saving ”Optimally” for Retirement?” *Journal of Political Economy*, 114, 607–643.

Spiegler, R. (2011): *Bounded Rationality and Industrial Organization*, Oxford University Press.

Strotz, R. H. (1955): “Myopia and Inconsistency in Dynamic Utility Maximization,” *The Review of Economic Studies*, 23, 165–180.

Tapia, W. and J. Yermo (2008): “Fees in Individual Account Pension Systems: A Cross-Country Comparison,” *OECD Working Paper on Insurance and Private Pensions No. 27.*

Thaler, R. H. and S. Benartzi (2004): “Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving,” *Journal of Political Economy*, 112, S164–S187.

Willis Towers Watson (2017): *Global Pension Assets Study.*

Supplementary Appendix for Online Publication

This online appendix provides a formal treatment of alternative formulations and extensions of the simple model, as well as a full set of quantitative results.

A Quasi-linearity of the utility function

The utility function assumed in the baseline simple model is linear in transfer f between the agent and the provider. That is for several reasons. First, quasi-linearity implies that the interest rate r is the only contract parameter that affects the agent's saving decision in period 1. In the context of life-cycle wealth accumulation, this should not seem overly restrictive. Consequently, the optimal interest rate r^* maximises the *perceived* social surplus, while the optimal fee f^* extracts the surplus from the agent. This simplifies the provider's problem substantially. Second, quasi-linearity results in a clear efficiency criterion, which is independent of the transfer between the agent and the provider. Put differently, the resulting efficiency criterion would be the objective function of an agent who has access to the provider's technology of converting savings into future wealth. Third, this simplification is achieved without a significant loss of insight as long as the impact of the fee on the optimal savings remains 'small'. Finally, note that such an approach is standard in the exploitative contracting literature (Heidhues and Kőszegi, 2018; Spiegler, 2011).

For completeness, consider an alternative formulation of the model that relaxes the assumption of quasi-linearity:

$$U_1 = u(c_1) + \beta\delta u(c_2),$$

where the intertemporal budget constraint $c_1 + \frac{c_2}{(1+r)} \leq Y - f$ implies $c_1 = Y - s - f$ and $c_2 = (1+r)s$. Then, the agent's savings $s \equiv s(\beta, r, f)$ that maximise U_1 are affected by the fee f in the same way as by a negative income shock:

$$s(\beta, r, f) = \frac{1}{1+(\beta\delta)^{\frac{-1}{\theta}}(1+r)^{\frac{\theta-1}{\theta}}} \times [Y - f]$$

The agent's valuation of a contract is therefore strictly decreasing in f:

$$V_0(s(\hat{\beta}, r, f), r, f) = u(Y - \hat{s} - f) + \delta u((1+r)\hat{s})$$

$$\frac{d\hat{V}_0}{df} = \underbrace{\frac{\partial u(Y - \hat{s} - f)}{\partial f}}_{<0} + \underbrace{\frac{\partial \hat{V}_0}{\partial \hat{s}}}_{\geq 0}\underbrace{\frac{d\hat{s}}{df}}_{<0} < 0$$

which implies that any interest rate $r > 0$ is associated with a fee $f(r) > 0$ that extracts the entire perceived surplus:

$$V_0(s(\hat{\beta}, r, f(r)), r, f(r)) = \underline{u}$$

Intuitively, because $\frac{d\hat{V}_0}{dr} > 0$, $\frac{df(r)}{dr} > 0$. That means that the monopolistic provider is able to charge the agent a higher fee $f(r)$ when she offers a higher rate of return:

$$dV_0(s(\hat{\beta}, r, f(r)), r, f(r)) = \frac{\partial \hat{V}_0}{\partial \hat{s}} d\hat{s} \; + \; \frac{\partial \hat{V}_0}{\partial r} dr \; + \; \frac{\partial \hat{V}_0}{\partial f} df = 0 \iff$$

$$\iff df/dr = \frac{-1}{\partial \hat{V}_0/\partial f} \underbrace{[\frac{\partial \hat{V}_0}{\partial r} + \frac{\partial \hat{V}_0}{\partial \hat{s}} \frac{d\hat{s}}{dr}]}_{=d\hat{V}_0/dr} > 0$$

Similarly to the proof of Proposition 1, the provider's problem simplifies to:

$$\max_r \; f(r) - K(r, s(\beta, r, f(r)))$$

The exploitative distortion due to naiveté ($\hat{\beta} > \beta$) enters the problem via $f(r)$. Due to monotonicity ($\frac{df(r)}{dr} > 0$) and the fact that the shape of $f(r)$ is determined by $d\hat{V}_0/dr$, the sign of $\frac{d^2\hat{V}_0}{d\hat{\beta}\, dr}$ feeds into the sign of $\frac{d^2 f(r)}{d\hat{\beta}\, dr}$. To see this, note that $\partial \hat{V}_0/\partial f = -u'(Y - f - \hat{s})$ and write:

$$d^2 f/d\hat{\beta}\, dr = \frac{1}{u'(Y-f-\hat{s})}[\frac{d^2\hat{V}_0}{d\hat{\beta}\, dr} + \frac{d\hat{V}_0}{dr} \frac{u''(Y-f-\hat{s})}{u'(Y-f-\hat{s})} \frac{d\hat{s}}{d\hat{\beta}}] =$$

$$= \frac{1}{u'(Y-f-\hat{s})}[\frac{d^2\hat{V}_0}{d\hat{\beta}\, dr} + \underbrace{\frac{d\hat{V}_0}{dr} \frac{-\theta}{(Y - f - \hat{s})} \frac{d\hat{s}}{d\hat{\beta}}}_{<0}]$$

As the second term in the brackets is negative for any parameter values and $\frac{d^2\hat{V}_0}{d\hat{\beta}\, dr} \leq 0$ for $\theta \geq 1$, the entire expression is strictly negative when the income effect dominates. For $\theta < 1$, on the other hand, $\frac{d^2\hat{V}_0}{d\hat{\beta}\, dr} > 0$. This implies that as $\theta \to 0$, $d^2 f/d\hat{\beta}\, dr$ eventually turns positive for $\theta \leq \check{\theta} < 1$.

The solution to the provider's problem is given by:

$$\frac{df(r)}{dr} = \frac{dK(r,s(\beta,r,f(r)))}{dr} = \frac{\partial K}{\partial r} \; + \; \frac{\partial K}{\partial s}[\frac{\partial s}{\partial r} + \frac{\partial s}{\partial f} \frac{df(r)}{dr}]$$

$$\iff \frac{df(r)}{dr}[1 - \frac{\partial K}{\partial s}\frac{\partial s}{\partial f}] = \frac{\partial K}{\partial r} \; + \; \frac{\partial K}{\partial s}\frac{\partial s}{\partial r}$$

where the only term that depends on $\hat{\beta}$ is $f(r)$. Therefore, the conclusion regarding the dependence of the exploitative distortion on the CRRA parameter θ remains valid. The difference from the prediction of the baseline model is that the threshold above which the exploitative savings contract becomes inefficiently cheap is some $\check{\theta} < 1$, rather than 1.

B Endogenous valuation of the outside option

In the baseline simple model, the valuation of the agent's outside option is assumed to be exogenous and type-independent. The individual rationality constraint in the provider's problem thus takes the following form:

$$V_0(s(\hat{\beta}, r), r) - f \geq \underline{u},$$

Then, taking into account the other constraint $s = s(\beta, r)$, the provider's problem can be written concisely as:

$$\max_r \; V_0(s(\hat{\beta}, r), r) - \underline{u} - K(r, s(\beta, r))$$

as in the derivation of Proposition 1.

In contrast, suppose that the agent's outside option involves saving into a costless savings account with a rate of return $\underline{r}$, for example offered by a bank. Then, the valuation of the outside option becomes endogenous:

$$\underline{u} = V_0(s(\hat{\beta}, \underline{r}), \underline{r}) - 0$$

It follows immediately from Lemma 1 that naifs overestimate utility from their outside option, just like any other savings contract. The simplified provider's problem is:

$$\max_r \; V_0(s(\hat{\beta}, r), r) - V_0(s(\hat{\beta}, \underline{r}), \underline{r}) - K(r, s(\beta, r))$$

This formulation implies the following. First, as $\frac{d\hat{V}_0}{dr} > 0$, the provider may charge a positive fee only if she offers $r > \underline{r}$. Second, the provider's choice of the profit-maximising rate of return is independent of the individual rationality constraint. Consequently, the qualitative properties of the exploitative savings contract are the same as under the baseline model. Third, regarding consumer welfare, while the negative efficiency effect prevails under the endogenous valuation of the outside option, the negative distributional effect is mitigated. At the optimum, the utility obtained by the agent is:

$$U_0(P^*) = V_0(s(\beta, r^*), r^*) - f^* = V_0(s(\beta, r^*), r^*) - V_0(s(\hat{\beta}, r^*), r^*) +$$

$$+ V_0(s(\hat{\beta}, \underline{r}), \underline{r})$$

While sophisticated agents obtain their (now type-dependent) reservation utility, the impact of naiveté on welfare can be derived using total differentiation:

$$\frac{dU_0(P^*)}{d\hat{\beta}} =$$

$$
\underbrace{-\frac{\partial V_0(s(\hat{\beta}, r^*), r^*)}{\partial s(\hat{\beta}, r^*)} \frac{d\, s(\hat{\beta}, r^*)}{d\, \hat{\beta}}}_{\text{the 'original' distributional effect}} + \underbrace{\{\frac{\partial V_0(s(\beta, r^*), r^*)}{\partial r^*} - \frac{\partial V_0(s(\hat{\beta}, r^*), r^*)}{\partial r^*}\} \frac{d\, r^*}{d\, \hat{\beta}}}_{\text{the 'original' efficiency effect}} +
$$

$$
+ \underbrace{\frac{\partial V_0(s(\hat{\beta}, \underline{r}), \underline{r})}{\partial s(\hat{\beta}, \underline{r})} \frac{d\, s(\hat{\beta}, \underline{r})}{d\, \hat{\beta}}}_{\text{the mitigating distributional effect due to overestimation of } \underline{u}}
$$

Relative to the model with exogenous valuation of the outside option, the above expression includes an additional term - a *positive* distributional effect, arising from the fact that naifs overvalue their outside option. As it turns out, the total distributional effect of naiveté becomes positive for $\theta > 1$. In this case, naifs no longer overpay for their savings contracts, but in fact require lower fees than sophisticates to satisfy their individual rationality constraint. The intuition behind this result is as follows. Recall that for $\theta > 1$ naifs overvalue cheaper contracts to a greater extent, as captured by $\frac{d^2 \hat{V}_0}{d\hat{\beta}\, dr} < 0$. In that case, the overvaluation of the low-cost outside option $\underline{u} = V_0(s(\hat{\beta}, \underline{r}), \underline{r})$ is of greater magnitude than the overvaluation of the optimal savings contract with $r^* > \underline{r}$. As a consequence, the total distributional effect of naiveté is positive. This mitigates the negative impact of naiveté about the present bias on consumer welfare.

For completeness, note that there is no role for the distributional effect in perfectly competitive environments. Then, naifs are always strictly worse off than sophisticates because of the negative efficiency effect.

C Competition

C.1 Perfect competition

Predictions of the baseline simple model are derived for a special case of monopoly. Consider another extreme example, namely perfect competition. Under perfect competition, many homogeneous providers freely enter the market until each provider's profits are zero at the optimum. Thus in a competitive equilibrium, the offered contract must satisfy an additional zero-profit condition:

$$
\max_{r,f} \;\; \pi = f - K(r, s), \text{ s.t.:}
$$

1. $s = s(\beta, r)$
2. $\hat{V}_0 - f \geq \underline{u}$
3. $\pi = 0$

Notice that conditional on the first constraint, the second constraint (individual rationality) implies $f \leq \hat{V}_0 - \underline{u}$ for any r, which determines an upper bound for the fee f. The third constraint (zero profit) implies $f = K(r, s(\beta, r))$, which is a lower bound for f. Under perfect competition, this lower bound is in fact binding in equilibrium.

Suppose that the monopolistic profits were strictly positive. This implies that under perfect competition the individual rationality constraint is slack, while the zero profit condition is binding. Then, the equilibrium contract solves:

$$\max_r \; V_0(s(\hat{\beta}, r), r) - K(r, s(\beta, r))$$

In contrast to the case of monopoly, this is not an immediate observation. First, the zero profit condition implies that for any r, the fee must equal $K(r, s(\beta, r))$. Then, the above expression gives the agent's *perceived* utility from signing a contract characterised by an interest rate r and a corresponding zero-profit fee. A contract that does not maximise the agent's perceived utility cannot be offered in equilibrium, because a competing provider could offer contract terms that improve the perceived utility while holding the profits constant.

Note that the equilibrium interest rate r^* is not affected by free entry into the market. That is because r^* satisfies the same first-order condition as under monopoly. Under perfect competition, however, the agent pays a strictly lower fee f^*, which makes him better off at the expense of the providers' profits. In sum, competition does away with the negative distributional effect, but not the negative efficiency effect. This can be seen directly from the expression for utility attained by either type in equilibrium:

$$U_0(P^*) = V_0(s(\beta, r^*), r^*) - f^* = V_0(s(\beta, r^*), r^*) - K(r, s(\beta, r^*)) > \underline{u}$$

where the inequality follows from the assumption of strictly positive monopoly profits. The impact of naiveté on welfare is captured by:

$$\frac{dU_0(P^*)}{d\hat{\beta}} = \left[\frac{dV_0}{dr^*} - \frac{dK}{dr^*}\right] \frac{dr^*}{d\hat{\beta}} < 0$$

which captures the negative efficiency effect, but does not include the distributional effect. The fact that this expression is always negative follows from the convexity of K, concavity of V_0, and the sign of $dr^*/d\hat{\beta}$ (see the proof of Proposition 1).

C.2 Imperfect competition

As shown above, the fee is the only contract parameter that changes between monopolistic and perfectly competitive environments, while the optimal interest rate is unaffected by the degree of competition. It is therefore natural to conjecture that a model of imperfect competition would generate the same optimal rate of return combined with an intermediate fee. Following Heidhues and Kőszegi (2010), consider the following Hotelling-style model of imperfect competition. There are two identical providers, denoted A and B, located at the endpoints of a unit interval. The providers simultaneously choose parameters of the offered contract, i.e. the interest rate r^i and the fee f^i, for $i \in \{A, B\}$, in order to maximise expected profits. There is a unit mass of identical agents distributed uniformly along the interval, or, equivalently, a single agent who is *ex ante* equally likely to find himself located at any point along the interval. Agents sign at most one contract. An agent located at $x \in [0, 1]$ evaluates the contract offers according to:

$$\hat{U}_0^A = V_0(s(\hat{\beta}, r^A), r^A) - f^A - \xi x$$

$$\hat{U}_0^B = V_0(s(\hat{\beta}, r^B), r^B) - f^B - \xi(1 - x)$$

where $\xi \geq 0$ is a parameter capturing the agent's 'distance aversion'. In the context of choice of a financial product, ξ is perhaps more naturally interpreted as reflecting the individual's tendency to prefer a 'default' provider. In case the agent rejects both offers, he obtains reservation utility υ.

The intuition that both providers would optimally offer the same rate of return r^* as under monopoly and perfect competition does indeed extend to this imperfectly competitive setting. Taking the perspective of provider A, the agent's individual rationality constraint is:

$$\hat{U}_0^A = V_0(s(\hat{\beta}, r^A), r^A) - f^A - \xi x \geq \underline{u}$$

where $\underline{u} = \max \{\hat{U}_0^B; \upsilon\}$ is endogenous and determined by the competitor's contract offer $P^B = (r^B, f^B)$, the parameters of which provider A takes as given. Let π_x^A denote provider A's profits from attracting the agent located at $x \in [0, 1]$, conditional on the above constraint being satisfied:

$$\pi_x^A = f^A - K(r^A, s(\beta, r^A))$$

Because $\frac{\partial \pi_x^A}{\partial f^A} > 0$, the individual rationality constraint is binding at the optimum. This implies that the optimal contract offer solves:

$$\max_{r^A} \; V_0(s(\hat{\beta}, r^A), r^A) - \xi x - \underline{u} - K(r^A, s(\beta, r^A))$$

which yields the same first-order condition for r^* as under monopoly and perfect competition, independently of the agent's location x and parameter ξ. This is also true for the optimal choice of r^B by provider B. As a result, $V_0(s(\hat{\beta}, r^*), r^*)$ is common across both contract offers.

Depending on the fees charged by the providers, the indifferent agent is located at $x^* \in [0, 1]$, where x^* solves:

$$-f^A - \xi x^* = -f^B - \xi(1 - x^*)$$

or $x^* = \frac{f^B - f^A + \xi}{2\xi}$. Individuals located at $x \in [0, x^*]$ obtain greater *ex post* utility from selecting an offer of provider A and thus the provider attracts a mass x^* of agents when charging f^A. Conditional on f^B, provider A chooses a fee that maximises expected profits:

$$\mathbb{E}\,\pi^A = \{\, f^A - K(r^*, s(\beta, r^*))\,\} \times \frac{f^B - f^A + \xi}{2\xi}$$

The above is maximised by $f^{A*} = \frac{f^B + \xi + K(r^*, s(\beta, r^*))}{2}$. Symmetry of the problem implies that $f^{A*} = f^{B*}$ or, equivalently, $x^* = \frac{1}{2}$. Thus in equilibrium both providers charge:

$$f^* = \begin{cases} \xi + K(r^*, s(\beta, r^*)) & \text{for } \xi \in [0, f^M - K(r^*, s(\beta, r^*))] \\ f^M & \text{for } \xi > f^M - K(r^*, s(\beta, r^*)) \end{cases}$$

where f^M denotes the profit-maximising fee charged by a monopolistic provider. This result has the following natural interpretation. When the agent is not at all 'distance averse' (i.e. $\xi = 0$), the Hotelling model induces Bertrand competition and thus competitive pricing. The larger ξ, the greater the monopolistic powers enjoyed by the providers. The equilibrium fee is therefore monotonically increasing in ξ. For high enough ξ, the providers charge the monopolistic price.

D Variable fees

D.1 Variable fee on savings

Suppose that in addition to the flat-rate fee f, the provider is able to charge a variable fee t proportional to the agent's savings. The fee is calculated based on the agent's actual savings and charged in period 1. The provider's problem is:

$$\max_{r,f,t} \;\; \pi = f + ts - K(r, s), \text{ s.t.:}$$

1. $s = s(\beta, r, t)$

2. $V_0(s(\hat{\beta}, r, t), r, t) - f \geq \underline{u}$

where

$$V_0(s(\hat{\beta}, r, t), r, t) = \frac{[Y - s(\hat{\beta}, r, t) \times (1+t)]^{1-\theta} - 1}{1-\theta} \; + \; \delta \frac{[(1+r) \times s(\hat{\beta}, r, t)]^{1-\theta} - 1}{1-\theta}$$

$$s(\beta, r, t) = \frac{1}{(1+t) + (\beta\delta)^{\frac{-1}{\theta}} (1+r)^{\frac{\theta-1}{\theta}} (1+t)^{\frac{1}{\theta}}} \times Y$$

It is straightforward to note the following:

$$\frac{d\,s(\beta,r,t)}{d\,t} < 0 \quad \text{and} \quad \frac{d\,\hat{V}_0}{d\,t} = \underbrace{\frac{\partial\,\hat{V}_0}{\partial\,t}}_{<0} + \underbrace{\frac{\partial\,\hat{V}_0}{\partial\,\hat{s}}}_{\geq 0}\underbrace{\frac{d\,\hat{s}}{d\,t}}_{<0} < 0$$

The second inequality above shows that the introduction of a variable fee reduces the agent's valuation of a contract not only directly, but also because the variable fee reduces future savings. Importantly, this indirect effect is taken into account only by a (partially) sophisticated agent. A completely naïve agent, in contrast, expects his future self to make a savings decision that maximises the current objective function and therefore $\partial\,\hat{V}_0/\partial\,\hat{s} = 0$ when $\hat{\beta} = 1$.

After substituting the binding constraints into the provider's problem, the optimal interest rate satisfies the following first-order condition:

$$\frac{d\,V_0(s(\hat{\beta},r,t),r,t)}{dr} + t\,\frac{d\,s(\beta,r,t)}{d\,r} = \frac{d\,K(r,s(\beta,r,t))}{dr}$$

For any given level of the variable fee t, the agent's naiveté about the present bias affects the optimal interest rate via $\frac{d\,V_0(s(\hat{\beta},r,t),r,t)}{dr}$ only. That would imply that the direction of the exploitative distortion to the offered rate of return carries over from the baseline model. In addition, from convexity of the provider's cost function, the possibility to charge a variable fee shifts the optimal interest rate r^* in the direction which increases the agent's actual savings, independent of the degree of naiveté:

$$\frac{d\,r^*}{d\,t} \quad \begin{cases} > 0 & \text{for } \theta < 1 \\ = 0 & \text{for } \theta = 1 \\ < 0 & \text{for } \theta > 1 \end{cases}$$

Intuitively, when the provider charges proportionally to the agent's savings, she has an additional incentive to induce higher savings.

The optimal variable fee t^* satisfies:

$$\frac{d\,V_0(s(\hat{\beta},r,t),r,t)}{dt} + s(\beta,r,t) + t\,\frac{d\,s(\beta,r,t)}{d\,t} = \frac{\partial\,K(r,s(\beta,r,t))}{\partial s(\beta,r,t)}\frac{d\,s(\beta,r,t)}{d\,t}$$

This condition indicates that the exploitative distortion to t^* due to the agent's naiveté materialises via $d\,V_0(s(\hat{\beta},r,t),r,t)/dt$, which is the only term that depends on $\hat{\beta}$. Comparing the condition under full naiveté and (partial) sophistication, the magnitude of the strictly negative term $d\,\hat{V}_0/dt$ is greater under partial sophistication due to the indirect effect of the variable fee on savings. Although all the other terms in the above condition are independent of $\hat{\beta}$, additional assumptions need to be imposed on the provider's cost function in order to derive a sharp prediction.

More precisely, using the closed-form solution for $s(\beta, r, t)$ a few lines of algebraic rearrangements yield:

$$s(\beta, r, t) + t\ \frac{d\,s(\beta,r,t)}{d\,t} = s(\beta, r, t)\ \left(1 - \frac{t}{1+t} \frac{1+\frac{1}{\theta}(\beta\delta)^{-1/\theta}(1+r)^{(\theta-1)/\theta}(1+t)^{(1-\theta)/\theta}}{1+(\beta\delta)^{-1/\theta}(1+r)^{(\theta-1)/\theta}(1+t)^{(1-\theta)/\theta}}\right) \xrightarrow{\theta\to 1} s(\beta, r, t)\ \left(\frac{1}{1+t}\right)$$

which is positive and strictly decreasing in t. Then, under an additional assumption that the provider's cost function is relatively insensitive to the level of the agent's savings, i.e. $\partial\, K(r, s(\beta, r, t))/\partial\, s(\beta, r, t)$ is negligibly small, the above first-order condition implies that the provider would optimally charge a naïve agent a higher variable fee t^*.

Finally, taking into account the impact of naiveté on the optimal variable fee, the first-order condition determining the optimal rate of return r^* implies that the introduction of the variable fee amplifies the original exploitative distortion to the contract terms aimed at a naïve agent. That is because $d\, s(\beta, r, t)/d\, r$ has the same sign as $d^2\, V_0(s(\hat{\beta}, r, t), r, t)/d\,\hat{\beta}\, d\, r$. In other words, when the provider can charge a variable fee on the agent's savings, the exploitative contract offered to a naïve agent again promises an inefficiently high rate of return for $\theta < 1$ and an inefficiently low rate of return for $\theta > 1$, but the exploitative distortion is of a larger magnitude, compared to the baseline case with a flat-rate fee only.

D.2 Variable fee on accumulated wealth

Now suppose that the provider is able to charge a variable fee t proportional to the agent's accumulated wealth $(1 + r) \times s$. The fee is charged in period 2 based on the actual size of the agent's pension pot, but, for the sake of notational simplicity, I ignore the discounting of future revenues by the provider. Then, the provider's problem becomes:

$$\max_{r,f,t}\ \pi = f + t(1 + r)s - K(r, s), \text{ s.t.:}$$

1. $s = s(\beta, \check{r})$
2. $V_0(s(\hat{\beta}, \check{r}), \check{r}) - f \geq \underline{u}$

where $\check{r} \equiv (1 + r)(1 - t) - 1$ is the net interest rate that the agent earns on his savings. For realistic values of r and t, $\check{r}$ can be approximated by $r - t$. Note that it is the net interest rate that determines the agent's valuation of the contract, and not particular values of r and t:

$$V_0(s(\hat{\beta}, \check{r}), \check{r}) = \frac{[Y - s(\hat{\beta},\check{r})]^{1-\theta} - 1}{1-\theta}\ +\ \delta \frac{[(1+\check{r}) \times s(\hat{\beta},\check{r})]^{1-\theta} - 1}{1-\theta}$$

Accounting for the optimally binding constraints and rewriting $t(1+r) = r - \check{r} \approx t$, the provider's problem takes a convenient form:

$$\max_{\check{r},t} \;\; \pi = V_0(s(\hat{\beta},\check{r}),\check{r}) - \underline{u} + t\, s(\beta,\check{r}) - K(\check{r}+t, s(\beta,\check{r}))$$

Note that this problem is not equivalent to the problem with a variable fee on savings, because the provider's cost of service is now a function of $\check{r}+t$, rather than just $\check{r}$.

First, the optimal variable fee t, which can be interpreted as a wedge between the gross rate of return that the provider generates and the net rate of return that is promised to the agent, satisfies the following first-order condition:

$$s(\beta,\check{r}) = \frac{\partial\, K(\check{r}+t, s(\beta,\check{r}))}{\partial\,(\check{r}+t)}$$

That is, for a given level of the net interest rate $\check{r}$ promised to the agent, the provider's optimal choice of the variable fee reflects the tradeoff between the marginal revenue proportional to the agent's savings and the marginal cost of generating higher gross returns. Notice that these terms are independent of the degree of the agent's naiveté. From convexity of the provider's cost function, the optimal variable fee is a strictly increasing function of the agent's savings $t^* \equiv t(s(\beta,\check{r}))$.

Second, the optimal net rate of return satisfies:

$$\frac{d\, V_0(s(\hat{\beta},\check{r}),\check{r})}{d\,\check{r}} \;+\; t\,\frac{d\, s(\beta,\check{r})}{d\,\check{r}} = \frac{d\, K(\check{r}+t, s(\beta,\check{r}))}{d\,\check{r}}$$

At first sight, this condition resembles the one derived in the preceding case analysing variable fees on the agent's savings. The difference is that here, the optimal variable fee t^* does not directly depend on the agent's degree of naiveté, but instead is an increasing function of the agent's savings $s(\beta,\check{r})$. What does that imply for the exploitative distortion to the net rate of return offered to a naïve agent? Notice that for a fixed t, the baseline exploitative distortion would carry over to this formulation of the model, in a qualitative sense. However, once the adjustment in t^* is taken into account, the original distortion again becomes amplified. That is because the exploitative contract always induces a naïve agent to save more than his sophisticated counterpart, which creates an incentive for the provider to charge a higher variable fee on the agent's pension wealth. In turn, this amplifies the magnitude of the exploitative distortion to $\check{r}$.

D.3 Alternative business model of the pension industry

Consider the following alternative formulation of the provider's problem:

$$\max_{r^P, r^C, f} \;\; \tilde{\pi} = \delta^P[(1+r^P)f + (r^P - r^C)s] - \tilde{K}(r^P), \text{ s.t.:}$$

1. $s = s(\beta, r^C)$
2. $V_0(s(\hat{\beta}, r^C), r^C) - f \geq \underline{u}$

Changing the variables to $t \equiv r^P - r^C$ and plugging in the binding constraints yields:

$$\max_{r^C, t} \;\; \delta^P[(1 + r^C + t)(V_0(s(\hat{\beta}, r^C), r^C) - \underline{u}) + t\, s(\beta, r^C)] - \tilde{K}(r^C + t)$$

The first-order condition with respect to r^C is:

$$\delta^P\left[V_0(s(\hat{\beta}, r^C), r^C) - \underline{u} + (1 + r^C + t)\tfrac{d\,V_0(s(\hat{\beta}, r^C), r^C)}{d\,r^C} + t\tfrac{d\,s(\beta, r^C)}{d\,r^C}\right] = \tfrac{d\,\tilde{K}(r^C+t)}{d\,(r^C+t)}$$

while the first-order condition with respect to t is:

$$\delta^P\left[V_0(s(\hat{\beta}, r^C), r^C) - \underline{u} + s(\beta, r^C)\right] = \tfrac{d\,\tilde{K}(r^C+t)}{d\,(r^C+t)}$$

where I used the fact that $d\,(r^C\!+\!t)/d\,r^C = d\,(r^C\!+\!t)/d\,t = 1$. Since $V_0(s(\hat{\beta}, r^C), r^C)$ is increasing in $\hat{\beta}$ for every r^C, the convexity of the provider's cost function implies that the provider optimally 'charges' a higher implicit fee t when contracting with a naïve agent, conditional on r^C. That is because the naif's higher willingness to pay for a contract makes it worthwhile for the provider to generate higher total returns on *her own* investment. In what follows, denote the optimal wedge satisfying the condition above by $t^* \equiv t(r^C, \hat{\beta})$.

What about the optimal net interest rate promised to the agent? Accounting for t^*, the relevant first-order condition simplifies to:

$$(1 + r^C + t^*)\,\tfrac{d\,V_0(s(\hat{\beta}, r^C), r^C)}{d\,r^C} \;+\; t^*\,\tfrac{d\,s(\beta, r^C)}{d\,r^C} \;=\; s(\beta, r^C)$$

The exploitative distortion to r^C due to the agent's naiveté can be inferred by comparing the magnitude of the terms on the left-hand side under different values of $\hat{\beta}$, as the right-hand side is invariant to $\hat{\beta}$. First, t^* is higher when the provider is contracting with a naif. Second, $d\,V_0(s(\hat{\beta}, r^C), r^C)/d\,r^C$ is increasing in $\hat{\beta}$ for $\theta < 1$, but decreasing in $\hat{\beta}$ for $\theta > 1$, as demonstrated in the proof of Proposition 1. Third, the sign of $d\,s(\beta, r^C)/d\,r^C$ is positive for $\theta < 1$, but negative for $\theta > 1$. Together, these observations imply that the net interest rate r^C optimally offered to a naif is inefficiently high for $\theta < 1$, as in the baseline model. For $\theta > 1$, in contrast, the conclusion is not as clear-cut. Specifically, the provider finds it optimal to offer an inefficiently low net interest rate to a naïve agent if and only if:

$$(1 + r^C + t(r^C, \hat{\beta}))\, \frac{d\, V_0(s(\hat{\beta}, r^C), r^C)}{d\, r^C} \;+\; t(r^C, \hat{\beta})\, \frac{d\, s(\beta, r^C)}{d\, r^C} \;>$$

$$> \;(1 + r^C + t(r^C, \beta))\, \frac{d\, V_0(s(\beta, r^C), r^C)}{d\, r^C} \;+\; t(r^C, \beta)\, \frac{d\, s(\beta, r^C)}{d\, r^C}$$

for $\hat{\beta} > \beta$. As $(t(r^C, \beta) - t(r^C, \hat{\beta}))) \times d\, s(\beta, r^C)/d\, r^C > 0$, the necessary, but not sufficient, condition for the above to hold is

$$(1 + r^C + t(r^C, \hat{\beta}))\, \frac{d\, V_0(s(\hat{\beta}, r^C), r^C)}{d\, r^C} \;>\; (1 + r^C + t(r^C, \beta))\, \frac{d\, V_0(s(\beta, r^C), r^C)}{d\, r^C}$$

We already know that $d\, V_0(s(\hat{\beta}, r^C), r^C)/d\, r^C$ is decreasing in $\hat{\beta}$ for $\theta > 1$, so the above requires t^* to be 'sufficiently increasing' in $\hat{\beta}$ for the inequality to hold. However, a crucial determinant of the shape of $t(r^C, \hat{\beta})$ is the curvature of the provider's cost function, about which we have only assumed convexity. Thus, although it cannot be ruled out that the provider would again find it optimal to offer an inefficiently low rate of return to a naïve agent, any sharp conclusion would hinge on specific parametric assumptions.

What is the intuition behind this ambiguity? In case when the provider is able to re-invest the collected flat-rate fee f, she has an additional incentive to generate a high *total* return $r^C + t$ when contracting with a naif. That is because the willingness to pay for a contract with given parameters is higher when the agent is naïve rather than sophisticated, as reflected in the size of the monopolistic fee f^*. For $\theta < 1$, the picture is clear. Distorting the net interest rate upwards further amplifies the extent to which a naif will overpay for a savings contract, in line with the mechanism studied in the main body of the paper. Thus this type of distortion achieves the two objectives simultaneously - it increases the overvaluation of a contract by a naif and it results in a higher total return on the provider's funds. For $\theta > 1$, in contrast, the two objectives are in conflict. In particular, generating a higher total return would be consistent with offering a *lower* net interest rate to a naïve agent (as in the baseline model) only if the increase in t^* associated with naiveté is of sufficient magnitude.

E Financial (il)literacy

E.1 'Headline' and 'actual' fees

Consider the following extension accounting for lack of financial sophistication in a stylised way. In contrast to naiveté about the present bias (i.e. one's own behaviour), a financially unsophisticated agent misinterprets the contract terms offered by the provider. First, suppose that the provider posts a 'headline' fee $\underline{f}$ and an 'actual' fee $\bar{f}$. An agent believes that the realised fee will be $\underline{f}$ with probability $\hat{p} \in [0,1]$, and $\bar{f}$ with residual probability $(1-\hat{p})$. Without loss of insight, suppose that the realised fee is $\bar{f}$ with probability 1. Thus $\hat{p}$ can be interpreted as a measure of the agent's financial illiteracy.[1] The provider's problem is:

$$\max_{r,\underline{f},\bar{f}} \pi = \bar{f} - K(r,s) \quad \text{s.t.:}$$

1. $s = s(\beta, r)$
2. $V_0(s(\hat{\beta}, r), r) - \hat{p}\underline{f} - (1-\hat{p})\bar{f} \geq \underline{u}$

Following substitution of the binding constraints, the problem takes the following form:

$$\max_{r,\underline{f}} \quad \tfrac{1}{(1-\hat{p})}\{V_0(s(\hat{\beta}, r), r) - \hat{p}\underline{f} - \underline{u}\} - K(r, s(\beta, r)) \iff$$

$$\iff \max_{r,\underline{f}} \quad \tfrac{V_0(s(\hat{\beta},r),r) - \underline{u}}{(1-\hat{p})} - \tfrac{\hat{p}}{(1-\hat{p})}\underline{f} - K(r, s(\beta, r))$$

As the above expression is strictly decreasing in $\underline{f}$, the headline fee takes the lowest possible value, for instance $\underline{f} = 0$ due to a plausible non-negativity constraint. Intuitively, offering a very attractive headline fee allows the provider to charge an accordingly higher actual fee $\bar{f}$.

The optimal interest rate r^*, which does not depend on $\underline{f}$, satisfies the following first-order condition:

$$\tfrac{1}{(1-\hat{p})} \times \tfrac{d\,V_0(s(\hat{\beta},r),r)}{dr} = \tfrac{d\,K(r,s(\beta,r))}{dr}$$

This implies that under financial illiteracy ($\hat{p} > 0$) the optimal interest rate shifts upwards, irrespective of the degree of naiveté. That is because the provider is able to charge a higher actual fee for any interest rate she offers. Note that such deviations from what would be the optimal contract terms under $\hat{p} = 0$ decrease efficiency further by maximising perceived, rather than actual surplus. Importantly, the exploitative distortion due to the agent's naiveté about the present bias (captured by $\frac{d^2\hat{V}_0}{d\hat{\beta}\,dr}$) not only carries over from the baseline model, but it is also of a greater magnitude because $1/(1-\hat{p}) \geq 1$.

[1]An example of a more refined model of this kind is **?**. See also Heidhues and Kőszegi (2018) for a review of the literature on the optimal design of hidden prices.

E.2 'Headline' and 'actual' rates of return

Consider an alternative model under which an agent is overly optimistic about the expected rate of return on his savings. This could be the case either due to ignorance of implicit fees or due to overoptimistic forecast of the future state of the world.[2] Suppose that the provider posts a 'headline' rate of return $\bar{r}$ and an 'actual' rate of return $\underline{r}$. The agent believes that the realised rate will be $\bar{r}$ with probability $\hat{p} \in [0,1]$, and $\underline{r}$ with residual probability $(1-\hat{p})$. Without loss of insight, suppose that the realised rate of return is $\underline{r}$ with probability 1. Thus $\hat{p}$ can be interpreted as a measure of the agent's financial illiteracy. The provider's problem is:

$$\max_{\bar{r},\underline{r},f} \pi = f - K(\underline{r}, s) \quad \text{s.t.:}$$

1. $s = s(\beta, \underline{r})$
2. $(1-\hat{p})V_0(s(\hat{\beta}, \underline{r}), \underline{r}) + \hat{p}V_0(s(\hat{\beta}, \bar{r}), \bar{r}) - f \geq \underline{u}$

The above formulation makes two implicit assumptions. First, when saving in period 1, the agent takes into account the correct rate of return. As a result, financial illiteracy affects the agent's decisions at the contracting stage only. This simplifies the exposition, but is relaxed below. Second, posting two different rates of return benefits the provider only to the extent that it lowers the cost of service relative to the collected fees. There is no additional revenue, e.g. due to implicit fees.

Following substitution of the binding constraints, the problem becomes:

$$\max_{\bar{r},\underline{r}} \quad (1-\hat{p})V_0(s(\hat{\beta}, \underline{r}), \underline{r}) + \hat{p}V_0(s(\hat{\beta}, \bar{r}), \bar{r}) - \underline{u} - K(\underline{r}, s(\beta, \underline{r}))$$

Because the above expression is strictly increasing in $\bar{r}$, the headline rate of return takes the highest permitted value, allowing the provider to charge a higher fee for any actual rate of return $\underline{r}$. The first-order condition determining the optimal choice of $\underline{r}$ is:

$$(1-\hat{p}) \times \frac{d\,V_0(s(\hat{\beta},\underline{r}),\underline{r})}{d\underline{r}} = \frac{d\,K(\underline{r},s(\beta,\underline{r}))}{d\underline{r}}$$

As can be seen from this condition, which is independent of $\bar{r}$, financial illiteracy ($\hat{p} > 0$) shifts the optimal rate of return downwards, independent of the degree of naiveté. That is because the valuation of the contract is now a function of the actual, but costly, rate $\underline{r}$ and the costless, but fictitious, rate $\bar{r}$. The provider therefore has less of an incentive to generate a high actual interest rate for the agent. Note that the exploitative distortion due to naiveté about the present bias,

[2] The latter is exactly the kind of bias that seems to underpin the design of increasingly popular structured retail products (?).

captured by $\frac{d^2 \hat{V}_0}{d\hat{\beta}\, dr}$, carries over from the baseline model in a qualitative sense. However, because the agent now attaches an irrationally low decision weight to the outcomes corresponding to $\underline{r}$, the exploitative distortion is of a lower magnitude.

For completeness, consider a slight modification of the model under which the agent is equally uncertain about the interest rate in period 1. Then, the provider's problem takes the following form:

$$\max_{\bar{r},\underline{r},f} \pi = f - K(\underline{r}, s) \quad \text{s.t.:}$$

1. $s = s(\beta, \bar{r}, \underline{r}, \hat{p})$
2. $(1-\hat{p})V_0(s(\hat{\beta}, \bar{r}, \underline{r}, \hat{p}), \underline{r}) + \hat{p}V_0(s(\hat{\beta}, \bar{r}, \underline{r}, \hat{p}), \bar{r}) - f \geq \underline{u}$

where $s(\beta, \bar{r}, \underline{r}, \hat{p})$ maximises:

$$\mathbb{E}\, U_1 = \frac{[Y-s]^{1-\theta}-1}{1-\theta} + \beta\delta\{\hat{p}\frac{[(1+\bar{r})s]^{1-\theta}-1}{1-\theta} + (1-\hat{p})\frac{[(1+\underline{r})s]^{1-\theta}-1}{1-\theta}\}$$

Even though there is no closed-form solution for $s(\beta, \bar{r}, \underline{r}, \hat{p})$, the continuity of $\mathbb{E}\, U_1$ implies that the agent's actual savings are continuous in $\hat{p} \in [0, 1]$, ranging from $s(\beta, \underline{r})$ to $s(\beta, \bar{r})$, and analogously for $s(\hat{\beta}, \bar{r}, \underline{r}, \hat{p})$. Importantly, an agent who is naïve about his present bias, still overestimates future saving. That is, $s(\hat{\beta}, \bar{r}, \underline{r}, \hat{p}) > s(\beta, \bar{r}, \underline{r}, \hat{p})$ for $\hat{\beta} > \beta$ and any given contract parameters. Following substitution of the binding constraints, the provider's problem reduces to:

$$\max_{\bar{r},\underline{r}} \quad (1-\hat{p})V_0(s(\hat{\beta}, \bar{r}, \underline{r}, \hat{p}), \underline{r}) + \hat{p}V_0(s(\hat{\beta}, \bar{r}, \underline{r}, \hat{p}), \bar{r}) - \underline{u} - K(\underline{r}, s(\beta, \bar{r}, \underline{r}, \hat{p}))$$

The first-order condition for $\bar{r}^*$ is:

$$(1-\hat{p}) \times \frac{\partial\, V_0(s(\hat{\beta},\bar{r},\underline{r},\hat{p}),\underline{r})}{\partial\, s(\hat{\beta},\bar{r},\underline{r},\hat{p})} \frac{d\, s(\hat{\beta},\bar{r},\underline{r},\hat{p})}{d\bar{r}} + \hat{p} \times \frac{d\, V_0(s(\hat{\beta},\bar{r},\underline{r},\hat{p}),\bar{r})}{d\bar{r}} = \frac{\partial\, K(\underline{r},s(\beta,\bar{r},\underline{r},\hat{p}))}{\partial\, s(\beta,\bar{r},\underline{r},\hat{p})} \frac{d\, s(\beta,\bar{r},\underline{r},\hat{p})}{d\bar{r}}$$

As the headline rate of return now affects the agent's actual and predicted savings, it is no longer true that that the provider's objective is strictly increasing in $\bar{r}$ everywhere. First, the headline rate of return affects the agent's valuation of the contract also in the case when the actual rate of return $\underline{r}$ materialises, due to its impact on the savings decision in period 1. Second, the headline rate of return affects the provider's cost, again due to its impact on the agent's actual savings. These considerations would lead the provider to offer a lower headline rate of return than in the preceding case with no uncertainty about the interest rate in period 1 either when the provider's cost is sufficiently increasing in the actual savings s (which requires $\theta < 1$) or when the agent's valuation of the contract is sufficiently increasing in the predicted savings $\hat{s}$ (which requires $\theta > 1$ as well as $\hat{\beta} < 1$).

Compare this to the first-order condition for the actual rate of return $\underline{r}$:

$$(1-\hat{p}) \times \frac{d\,V_0(s(\hat{\beta},\bar{r},\underline{r},\hat{p}),\underline{r})}{d\underline{r}} + \hat{p} \times \frac{\partial\,V_0(s(\hat{\beta},\bar{r},\underline{r},\hat{p}),\bar{r})}{\partial\,s(\hat{\beta},\bar{r},\underline{r},\hat{p})} \frac{d\,s(\hat{\beta},\bar{r},\underline{r},\hat{p})}{d\underline{r}} = \frac{d\,K(\underline{r},s(\beta,\bar{r},\underline{r},\hat{p}))}{d\underline{r}}$$

How does financial illiteracy ($\hat{p} > 0$) affect the optimal offer of the provider? Unsurprisingly, the two main conclusions from the preceding analysis remain valid. That is, financial illiteracy reduces the optimal rate of return $\underline{r}^*$ irrespective of the degree of naiveté, but the the direction of the exploitative distortion due to naiveté about the present bias carries over from the baseline model. The first point is immediate for either $\hat{\beta} \to 1$ or $\theta \to 1$, when the impact of $\underline{r}$ on the contract valuation in case when $\bar{r}$ realises is negligible. As before, taking into account the agent's financial illiteracy results in a lower optimal rate of return, because the agent attaches an irrationally low decision weight to the associated outcomes. Note that this conclusion remains valid for a wide range of parameter values even if the term on the left-hand side multiplied by $\hat{p}$ is not negligibly small. That is because $\frac{\partial\,\hat{V}_0}{\partial\,\hat{s}}\frac{d\,\hat{s}}{d\,r} < \frac{d\,\hat{V}_0}{d\,r} = \frac{\partial\,\hat{V}_0}{\partial\,r} + \frac{\partial\,\hat{V}_0}{\partial\,\hat{s}}\frac{d\,\hat{s}}{d\,r}$.[3]

Similarly, the second point is immediate for either $\hat{\beta} \to 1$, $\theta \to 1$, or $\hat{p} \to 0$. Then, the exploitative distortion due to naiveté about the present bias carries over from the baseline model in a qualitative sense, but is of a smaller magnitude. Without any of these simplifying, but not unrealistic, assumptions, any sharp conclusion would require further parametric assumptions to determine whether the impact of $\hat{\beta}$ on $\frac{\partial\,V_0(s(\hat{\beta},\bar{r},\underline{r},\hat{p}),\bar{r})}{\partial\,s(\hat{\beta},\bar{r},\underline{r},\hat{p})}\frac{d\,s(\hat{\beta},\bar{r},\underline{r},\hat{p})}{d\underline{r}}$ is of opposite sign and greater magnitude than $\frac{d^2\,V_0(s(\hat{\beta},\bar{r},\underline{r},\hat{p}),\underline{r})}{d\hat{\beta}\;d\underline{r}}$, which would be necessary to overturn the main conclusion of the baseline model.

F Imperfect observability with homogeneous beliefs

Consider a case when the provider serves two indistinguishable types of agents with homogeneous beliefs about their present bias, but with different actual magnitudes of the bias. Without loss of generality, suppose that fraction $\lambda \in [0,1]$ of agents are completely naïve ($\beta < 1$ and $\hat{\beta} = 1$) and that a remaining fraction $(1-\lambda)$ are time-consistent ($\beta = \hat{\beta} = 1$).[4] As the two types share

[3]More precisely, the argument relies on $\left|\frac{\partial\,V_0(s(\hat{\beta},\bar{r},\underline{r},\hat{p}),\underline{r})}{\partial\,s(\hat{\beta},\bar{r},\underline{r},\hat{p})} - \frac{\partial\,V_0(s(\hat{\beta},\bar{r},\underline{r},\hat{p}),\bar{r})}{\partial\,s(\hat{\beta},\bar{r},\underline{r},\hat{p})}\right|$ being negligibly small.

[4]The same qualitative results can be obtained for a more general case of two types characterised by the same $\hat{\beta}$, but different $\beta^L < \beta^H \leq \hat{\beta}$.

a belief $\hat{\beta} = 1$, screening is not possible. That is because whichever contract is preferred by a time-consistent agent is also preferred by a naif. Consequently, such a market is characterised by pooling and no exclusion. The optimal contract solves:

$$\max_{r,f} \ \mathbb{E}[\pi] = \mathbb{E}[f - K(r,s)], \text{ s.t.:}$$

1. $s = s(\beta, r)$ with probability λ; $s = s(1, r)$ with probability $(1 - \lambda)$
2. $V_0(s(1, r), r) - f \geq \underline{u}$

which, following substitution of the binding constraints, simplifies to:

$$\max_{r} \ V_0(s(1, r), r) - \underline{u} - \lambda \times K(r, s(\beta, r)) - (1 - \lambda) \times K(r, s(1, r))$$

The optimal interest rate r^* satisfies the following first-order condition:

$$\frac{d\,V_0(s(1,r),r)}{dr} = \lambda \frac{d\,K(r,s(\beta,r))}{dr} + (1 - \lambda) \frac{d\,K(r,s(1,r))}{dr}$$

The comparison with the baseline case with perfect observability depends on the sign of $\frac{d^2 K(r,s)}{ds\,dr}$, i.e. whether the provider's marginal cost of offering a higher rate of return is increasing or decreasing in the agent's savings. Consider the following possibilities:

- If $\frac{d^2 K(r,s)}{ds\,dr} < 0$, for example due to scale effects or a greater bargaining power of a pension fund with a larger portfolio, then the above first-order condition implies that naifs are offered a higher interest rate than under perfect observability. This counteracts the exploitative distortion and improves efficiency and consumer welfare for $\theta > 1$, but only exacerbates the distortion for $\theta < 1$. Sophisticates are offered a lower interest rate than under perfect observability, which unambiguously diminishes efficiency while preserving consumer welfare. That is because sophisticates correctly value their contracts and always obtain utility of $\underline{u}$.
- If $\frac{d^2 K(r,s)}{ds\,dr} = 0$, then the contract is as (in)efficient as under perfect observability. Welfare of the two types is not affected by pooling.
- If $\frac{d^2 K(r,s)}{ds\,dr} > 0$, for example due to limits to arbitrage or the liquidity risk, then naifs are offered a lower interest rate than under perfect observability, which counteracts the exploitative distortion for $\theta < 1$, but only exacerbates it for $\theta > 1$. Sophisticates are offered a higher interest rate than in isolation, which unambiguously diminishes efficiency while preserving consumer welfare.

To summarise, as a result of pooling, contracts for sophisticated, time-consistent individuals are no longer efficient. But due to their correct valuation of the contract offer, the consumer welfare of sophisticates is preserved. The loss of efficiency in the sophisticates' contract is not necessarily compensated for by an improvement in efficiency of the naifs' contract. Under pooling, naïve present-biased individuals may obtain either a more efficient contract (and be better off) or an even less efficient contract (and be worse off) than under perfect observability, depending on the direction of the original exploitative distortion and the sign of the cross-partial derivative of the provider's cost function. According to the empirical literature on operating costs in the pension industry, neither sign of the relevant cross-partial derivative can be ruled out.[5]

G Imperfect observability with heterogeneous beliefs

Now suppose that the provider is contracting with two indistinguishable types of agents characterised by the same magnitude of their present bias, but different beliefs. Without loss of generality, assume that a population consists of fraction $\lambda \in [0, 1]$ of naïve present-biased agents ($\beta < 1$ and $\hat{\beta} = 1$) and fraction $(1 - \lambda)$ of sophisticated present-biased agents ($\beta = \hat{\beta} < 1$).[6] In contrast to the case with homogeneous beliefs, a monopolistic provider may differentiate the contract offers in order to screen the agents. However, depending on which case maximises the provider's total profit, she might also offer a pooling contract, or exclude sophisticates from the market. These possibilities are analysed in turn.

First, the optimal pooling contract solves the following problem:

$$\max_{r,f} \;\; \pi = f - K(r, s), \text{ s.t.:}$$

1. $s = s(\beta, r)$
2. $V_0(s(1, r), r) - f \geq \underline{u}$
3. $V_0(s(\beta, r), r) - f \geq \underline{u}$

Since $V_0(s(1, r), r) > V_0(s(\beta, r), r)$, constraint 2 (the individual-rationality constraint for naifs) is slack, while constraint 3 (the individual-rationality constraint for sophisticates) binds at the optimum. Thus the problem simplifies to:

[5] For completeness, note that potential concern about non-existence of equilibrium under competition and imperfect observability, as in **?**, does not apply to this setting. With homogeneous beliefs, the two types have identical preferences over contracts. In other words, there is no profitable deviation from the equilibrium pooling contract that would attract just one type.

[6] The same qualitative results can be obtained for a more general case of two types characterised by the same $\beta \leq 1$, but various $\hat{\beta}^H > \hat{\beta}^L \geq \beta$.

$$\max_r \ V_0(s(\beta, r), r) - \underline{u} - K(r, s(\beta, r))$$

The above describes the optimal contract offered to sophisticated agents under perfect observability. Such a contract is efficient by Proposition 1 and in this case pooling improves both efficiency and consumer welfare, eliminating exploitation of naïve agents. Welfare and efficiency of sophisticates' contracts are unaffected by pooling.

Next, consider the optimal design of a screening contract. In addition to the constraints introduced earlier, such a contract must be incentive compatible in the sense that neither of the types has an incentive to mimic the other type. The provider's problem is then:

$$\max_{r^N, r^S, f^N, f^S} \ \mathbb{E}[\pi] = \lambda\{f^N - K(r^N, s^N)\} + (1-\lambda)\{f^S - K(r^S, s^S)\}, \text{ s.t.:}$$

1. $s^N = s(\beta, r^N)$
2. $s^S = s(\beta, r^S)$
3. $V_0(s(1, r^N), r^N) - f^N \geq \underline{u}$
4. $V_0(s(\beta, r^S), r^S) - f^S \geq \underline{u}$
5. $V_0(s(1, r^N), r^N) - f^N \geq V_0(s(1, r^S), r^S) - f^S$
6. $V_0(s(\beta, r^S), r^S) - f^S \geq V_0(s(\beta, r^N), r^N) - f^N$

where superscripts N and S refer to naifs and sophisticates, respectively. While constraints 1-4 are the same as in the baseline problem, constraints 5 and 6 assure that no type prefers a contract designed for the other type ('no mimicking'). Note that the contract terms offered to each type under perfect observability do not solve the above problem. Because $V_0(s(1, r), r) > V_0(s(\beta, r), r)$, constraint 5 would be violated. As sophisticates do not overpay for their savings contracts, naifs would mimic them. Conversely, sophisticates would not want to mimic naifs when constraint 3 binds. These observations reduce the problem to:

$$\max_{r^N, r^S, f^N, f^S} \ \mathbb{E}\,\pi = \lambda\{f^N - K(r^N, s(\beta, r^N))\} + (1-\lambda)\{f^S - K(r^S, s(\beta, r^S)\}, \text{ s.t.:}$$

1. $V_0(s(1, r^N), r^N) - f^N \geq V_0(s(1, r^S), r^S) - f^S$
2. $V_0(s(\beta, r^S), r^S) - f^S \geq \underline{u}$

with both constraints binding at the optimum. The provider leaves as little rent as possible to naifs, subject to them not mimicking sophisticates. Similarly, the provider leaves as little rent as possible to sophisticates, subject to them accepting the contract offer. Thus constraint 2 implies:

$$f^S = V_0(s(\beta, r^S), r^S) - \underline{u}$$

and constraint 1 implies:

$$f^N = V_0(s(1, r^N), r^N) - V_0(s(1, r^S), r^S) + f^S =$$

$$= V_0(s(1, r^N), r^N) - V_0(s(1, r^S), r^S) + V_0(s(\beta, r^S), r^S) - \underline{u}$$

Substitution reduces the problem to:

$$\begin{aligned} \max_{r^N, r^S} \quad & \lambda \left\{ V_0(s(1, r^N), r^N) - \underline{u} - K(r^N, s(\beta, r^N) \right\} + \\ & + \lambda \left\{ V_0(s(\beta, r^S), r^S) - V_0(s(1, r^S), r^S) \right\} + \\ & + (1-\lambda) \left\{ V_0(s(\beta, r^S), r^S) - \underline{u} - K(r^S, s(\beta, r^S) \right\} \end{aligned}$$

This formulation has the following implications. First, despite screening, naifs are offered the same exploitative interest rate as under perfect observability. Due to the no mimicking constraint, however, naifs pay a lower fee, which curbs the negative distributional effect. Second, the interest rate offered to sophisticated agents is distorted away from the first-best in a way that creates a larger wedge between the contact terms aimed at the two types, reducing the rent for naïve agents. Thus the efficiency of sophisticates' contract terms deteriorates, although their welfare is preserved.

Finally, consider the case when the provider excludes sophisticated agents from the market by offering contract terms that violate their participation constraint, but are still accepted by naifs. The optimal of such contracts is the one that is offered to naïve agents under perfect observability. Such a contract maximises profits from interacting with naifs, but it makes sophisticated agents reject the offer. Then, naifs' outcomes are not affected by imperfect observability, but there is an efficiency loss due to the exclusion of sophisticates from the market. Sophisticated agents are nevertheless equally well off.

Pooling, screening, and exclusion result in the following profit levels for the provider:

$$\pi^{\text{pooling}} = \pi^S$$

$$\pi^{\text{screening}} = \lambda \breve{\pi}^N + (1-\lambda)\breve{\pi}^S$$

$$\pi^{\text{exclusion}} = \lambda \pi^N$$

where π^S and π^N denote the profits from operating in an isolated market populated by sophisticates and naifs, respectively. $\breve{\pi}^N$ and $\breve{\pi}^S$ denote the profits from interacting with naifs and sophisticates, respectively, when the provider offers a screening contract. Noting the following:

$$\pi^S \leq \pi^N, \quad \breve{\pi}^S \leq \pi^S, \quad \breve{\pi}^N \leq \pi^N$$

demonstrates that either of the three contract design strategies may be profit-maximising, depending on the parameters of the model (in particular, the population composition λ and the severity of the present bias β).[7]

H Policy interventions

H.1 Ceiling on fees

Regulations that restrict the types and amounts of fees that may be charged by pension providers are observed across many OECD countries (Dobronogov and Murthi, 2005; Tapia and Yermo, 2008). Consider imposing a ceiling on fee, denoted $\overline{f}$. Then, the provider's problem is subject to an additional constraint:

$$\max_{r,f} \ \pi = f - K(r,s), \text{ s.t.:}$$

1. $s = s(\beta, r)$
2. $\hat{V}_0 - f \geq \underline{u}$
3. $f \leq \overline{f}$

The additional constraint makes the optimal contract (weakly) cheaper. That is because the modified problem of the provider is:

$$\max_{r} \ V_0(s(\hat{\beta}, r), r) - \underline{u} - K(r, s(\beta, r)), \text{ s.t.:}$$

1. $V_0(s(\hat{\beta}, r), r) - \underline{u} \leq \overline{f}$

and the following result follows from the fact that $\frac{d\hat{V}_0}{dr} > 0$, see the proof of Proposition 1.

[7] As a final remark, note that for this population, a competitive equilibrium would involve two zero-profit contracts, one maximising sophisticates' perceived utility and another maximising naifs' perceived utility. These are the same contracts as offered under perfect competition and observability. Again, the potential issue of lack of a competitive equilibrium does not apply to this setting. That is because the difference in the agents' beliefs should be interpreted as a difference in 'tastes' rather than 'risks'. In other words, there is no adverse selection under the set of zero-profit contracts.

Lemma A1: *Under an effective ceiling on fees, the optimal interest rate r^* is revised downwards. While efficiency of a sophisticated agent's contract declines, consumer welfare is preserved. By contrast, efficiency and welfare attained by a naïve agent improve for $\theta < 1$, but decline for $\theta > 1$. In all cases, the provider's profits decrease.*

An effective ceiling not only mechanically reduces the transfer from the agent to the provider, but it also shifts the profit-maximising interest rate downwards. For a sophisticated agent, such a change implies no change in consumer welfare, but it strictly lowers efficiency by forcing contract parameters away from the first-best. For a naïve agent, by contrast, the impact on both welfare and efficiency depends on the size of the CRRA parameter and the direction of the exploitative distortion. In particular, the policy has a detrimental effect when $\theta > 1$ and the original exploitative contract is already inefficiently cheap.[8] Finally, the presence of the additional constraint lowers the provider's profits at the optimum.

H.2 Minimum savings requirement

The government can also impose a lower bound on private savings by making private pension contributions compulsory and specifying the minimum contribution rate.[9] For example, minimum default contributions into private pension schemes are regulated in Australia and the UK.

Suppose that the government introduces a minimum savings requirement $\underline{s}$. Then, the agent's actual saving is given by $s = \max\{s(\beta, r), \underline{s}\}$, while his expected saving is given by $\hat{s} = \max\{s(\hat{\beta}, r),\ \underline{s}\}$. Note that this still implies $\hat{s} \geq s$. Moreover, there exist values for r such that $s = \underline{s}$, while $\hat{s} > \underline{s}$. In such cases, the minimum savings requirement is binding, but a naïve agent does not realise this and treats the commitment provided by $\underline{s}$ as inconsequential. The provider's problem becomes:

$$\max_{r,f} \ \pi = f - K(r, s), \text{ s.t.:}$$

1. $s = \max\{s(\beta, r), \underline{s}\}$
2. $\hat{V}_0 - f \geq \underline{u}$

[8] If a policymaker suspects the savings contract provided by the private market to be inefficiently cheap, a regulation 'opposite' to the ceiling on fee would be desirable, such as enforcing return guarantees. In practice, this may well be prohibitively costly due to financial risks involved. Furthermore, such guarantees are not easily implementable due to various administrative and institutional barriers (Antolín et al., 2011).

[9] A theoretical argument in favour of this policy is provided by Amador et al. (2006) who study optimal commitment contracts in a setting where taste shocks create a trade-off between flexibility and commitment. They show that the optimal commitment always features a restriction resembling the minimum savings requirement.

Whether or not the requirement is binding at the optimum, and whether the agent realises this, depends on the interest rate. So how does the imposition of $\underline{s}$ affect the solution to the provider's problem? First, note that for $\theta < 1$ and for each r, the provider's marginal cost, which is given by:

$$\frac{dK}{dr} = \frac{\partial K}{\partial r} + \frac{\partial K}{\partial s}\frac{ds}{dr}$$

is bounded from above by the original marginal cost function. That is because $\frac{ds}{dr} > 0$, but the second component of the above expression disappears whenever the requirement binds. The underlying additional assumption is that the provider's cost is weakly increasing in the agent's savings, $\frac{\partial K}{\partial s} \geq 0$, as indicated by the empirical evidence. By similar reasoning, the provider's marginal cost is bounded from below by the original marginal cost function for $\theta > 1$.

Second, the provider's marginal benefit is:

$$\frac{d\hat{V}_0}{dr} = \frac{\partial \hat{V}_0}{\partial r} + \frac{\partial \hat{V}_0}{\partial \hat{s}}\frac{d\hat{s}}{dr}$$

and the above is bounded from above (below) by the original marginal benefit curve for $\theta < 1$ ($\theta > 1$). These counteracting shifts in the marginal cost and marginal benefit curves imply that a minimum savings requirement has an ambiguous impact on the optimal interest rate r^*.[10]

Consider the impact on welfare attained by a sophisticated agent. Given that $\frac{\partial V_0}{\partial s} \geq 0$, a minimum savings requirement provides commitment that (weakly) increases the agent's utility net of fees $V_0(s, r)$ for every r. Due to the provider's monopolistic power, however, the agent still obtains utility of $\underline{u}$, while the provider benefits from the increase in generated surplus.

A more specific prediction regarding the optimal interest rate r^* can be made for a case of a completely naïve agent. As $\hat{\beta} \to 1$, $\frac{\partial \hat{V}_0}{\partial \hat{s}} \to 0$, because a naif believes that his future self will discount exponentially and mistakenly applies the envelope theorem. Then, imposing a minimum savings requirement does not shift the provider's marginal benefit curve, but only the marginal cost curve.[11] Consequently, the optimal interest rate is (weakly) higher under a minimum savings requirement for $\theta < 1$ and (weakly) lower for $\theta > 1$, amplifying the original exploitative distortion. At the same time, for any interest rate r the minimum savings requirement (weakly) reduces welfare loss from naiveté. To see why, recall that $V_0(s, r)$ (weakly) increases under a minimum savings requirement due to the value of commitment. A completely naïve agent is nonetheless unwilling to pay for commitment that he perceives to be inconsequential. Overall, the fee that the provider is able to charge remains unaffected, but the actual welfare attained by a naif improves for any r. These observations are summarised below.

[10] Note that the efficient interest rate is also a function of $\underline{s}$.

[11] That is true provided that the policy does not force the agent to save 'too much', i.e. more than an exponential discounter would.

Lemma A2: *The impact of a minimum savings requirement on the optimal interest rate r^* is ambiguous in general. Although for $\hat{\beta} \to 1$ the minimum savings requirement (weakly) exacerbates the original exploitative distortion, welfare loss due to naiveté might diminish.*

Intuitively, when binding, the minimum savings requirement affects both revenues (due to the value of commitment) and costs of the provider. Moreover, by fine-tuning the offered interest rate, the provider effectively determines whether the minimum requirement is binding or not, and whether the agent realises this. In general, either contract design strategy could be profit-maximising, depending on parameters of the model. For example, a binding minimum savings requirement that is correctly acknowledged by the agent eliminates profits due to naiveté, but at the same time it increases the agent's willingness to pay for a contract, if he is at least partially sophisticated. Consequently, the impact of the minimum savings requirement on the optimal rate of return remains ambiguous.

The model yields a more specific prediction for the case of (nearly) complete naiveté. Then, the minimum savings requirement results in an offer of a (weakly) higher interest rate for $\theta < 1$ and a (weakly) lower interest rate for $\theta > 1$, which exacerbates the original exploitative distortion. The unambiguity of this prediction arises from the observation that under complete naiveté, the minimum savings requirement affects only the cost curve of the provider. Since a completely naïve agent perceives the requirement to be irrelevant no matter what the rate of return, his willingness to pay for a savings contract is unchanged. What is perhaps interesting is the possibility that a naïve agent may obtain commitment for which he is otherwise unwilling to pay, thus reducing the welfare loss from naiveté. By contrast, any gains from commitment provided to a sophisticated agent would be extracted by the provider.

I Calibration - Benchmark targets

As a benchmark case, the numerical model is calibrated to match the cost-to-assets ratio of 0.005, the share of administrative costs of 0.50, the markup of 0.20, and the equilibrium interest rate of 5%. Whenever applicable, numerical values are calculated at the means for computational convenience. Parameter γ_1, the elasticity of administrative costs, is set equal to 0.5. This section discusses the choice of these values in turn.

Operational costs expressed as a fraction of assets under management (the cost-to-assets ratio) is a standard metric used by the empirical literature on operation of pension funds. The observed cost-to-assets ratio tends to vary considerably across providers and countries, but its values lie typically between 0.1% and 1.2%, with the US pension funds incurring costs towards a lower end of the range (Bateman and Mitchell, 2004; Bauer et al., 2010; Bikker and De Dreu, 2009; Bikker et al., 2012; OECD, 2017).

A paper by Bikker and De Dreu (2009) is the only one that allows to compare the relative importance of administrative and investment cost components in the total cost of pension funds. On average, administrative costs are 50% greater than investment costs (implying a share of 0.60), but there is much more variation in the distribution of administrative costs. Moreover, investment costs tend to be under-reported. Bikker et al. (2012) remark only that administrative costs constitute "a very large proportion" of the total cost.

There is surprisingly little data on markups of pension funds, although a few papers stress the need to differentiate between costs and charges in the pension industry due to imperfect price competition (e.g., Bikker et al., 2012; OECD, 2017). According to the report by the Financial Conduct Authority (2017), the average profit margin in the asset management industry in the UK is as high as 36%. I interpret this number as an upper bound for markups in the pension industry.

Annual rates of return generated by pension funds typically lie within a range of 3% to 7% in real terms, although there is considerable variation across time and across countries (Bauer et al., 2010; OECD, 2017). In addition, setting $R^{Z*} = 5\%$ as a benchmark target reflects the consumption flow from the illiquid asset assumed by Laibson et al. (2018). Note that Laibson et al. (2018) consider rates of return of 3.38% and 6.59% as two extreme alternatives to their benchmark assumption.

Focusing on pension funds operating in Australia, Bateman and Mitchell (2004) find that the elasticity of administrative costs with respect to assets under management is 0.46, and 0.87 for a subset of retail funds. Bikker et al. (2012), who use a less inclusive definition of administrative costs, estimate the elasticity of 0.19 for a sample of Australian, Canadian, Dutch, and American pension funds. For a subset of the US pension funds, the estimate equals 0.23.

J Calibration - Sensitivity analysis

Table J.1 presents the sensitivity of the numerical results to changes in the target values and the assumed elasticity of administrative costs γ_1.

Table A1: Sensitivity analysis

Target	**Calibrated parameters**	**Pareto-efficient** R^Z	**CE welfare loss**	**Retirement wealth**

Benchmark	$k_1 = 1.5083$ $\gamma_1 = 0.5$ $k_2 = 2.650 \times 10^{12}$ $\gamma_2 = 5.75$ $\xi = 0.481$	5.25%	0.23%	−8%
$\gamma_1 = 0.35$	$k_1 = 10.0620$ $\gamma_1 = 0.35$ $k_2 = 3.206 \times 10^{12}$ $\gamma_2 = 5.80$ $\xi = 0.385$	5.25%	0.23%	−8%
$\gamma_1 = 0.65$	$k_1 = 0.2234$ $\gamma_1 = 0.65$ $k_2 = 2.190 \times 10^{12}$ $\gamma_2 = 5.70$ $\xi = 0.481$	5.25%	0.23%	−8%
Cost-to-assets ratio 0.0035	$k_1 = 1.0558$ $\gamma_1 = 0.5$ $k_2 = 1.213 \times 10^{15}$ $\gamma_2 = 7.45$ $\xi = 0.308$	5.25%	0.26%	−8%
Cost-to-assets ratio 0.0065	$k_1 = 1.9607$ $\gamma_1 = 0.5$ $k_2 = 3.544 \times 10^{10}$ $\gamma_2 = 4.55$ $\xi = 0.602$	5.25%	0.26%	−8%
Admin share 0.40	$k_1 = 1.2066$ $\gamma_1 = 0.5$ $k_2 = 2.203 \times 10^{11}$ $\gamma_2 = 5.05$ $\xi = 0.385$	5.25%	0.22%	−8%

Admin share 0.60	$k_1 = 1.8099$ $\gamma_1 = 0.5$ $k_2 = 7.938 \times 10^{13}$ $\gamma_2 = 6.70$ $\xi = 0.481$	5.25%	0.23%	−8%
Markup 0.15	$k_1 = 1.5083$ $\gamma_1 = 0.5$ $k_2 = 2.649 \times 10^{12}$ $\gamma_2 = 5.75$ $\xi = 0.385$	5.25%	0.23%	−8%
Markup 0.25	$k_1 = 1.5083$ $\gamma_1 = 0.5$ $k_2 = 2.649 \times 10^{12}$ $\gamma_2 = 5.75$ $\xi = 0.481$	5.25%	0.23%	−8%
$R^{Z*} = 4.75\%$	$k_1 = 1.4079$ $\gamma_1 = 0.5$ $k_2 = 2.554 \times 10^{12}$ $\gamma_2 = 5.60$ $\xi = 0.385$	5.00%	0.31%	−11%
$R^{Z*} = 5.25\%$	$k_1 = 1.5936$ $\gamma_1 = 0.5$ $k_2 = 2.306 \times 10^{12}$ $\gamma_2 = 5.85$ $\xi = 0.481$	5.50%	0.26%	−6%

The sensitivity analysis demonstrates that the quantitative results are robust to alternative target values. The calibrated parameters are not substantially affected and the quantitative impact of exploitative contracting remains almost exactly the same, except when targeting a different rate of return. When the model is calibrated so that the providers offer $R^Z = 4.75\%$ in equilibrium, the welfare loss associated with exploitative contracting increases to 0.31% and the negative impact on retirement wealth is also greater at 11%. When the model is calibrated so that the providers offer $R^Z = 5.25\%$ in equilibrium, the welfare loss equals 0.26% of annual consumption, but the impact on retirement wealth is slightly lower at 6%. In either case, the equilibrium contract is found to be inefficiently cheap.

K Calibration - Robustness checks

Table K.2 shows the results of recalibrating the model when the CRRA parameter is restricted to take the value of either 2.0 or 0.5. While imposing $\theta = 2.0$ substantially worsens the fit of the model to the data, imposing $\theta = 0.5$ improves it minimally. That is due to the fact that the re-calibration procedure iterates over a very fine multidimensional grid to find a parameter constellation that maximises the fit locally, as explained in the main text.

Table A2: Household-side moments and goodness of fit

	Data	**Recalibrated pref. params**	**Recalibrated under $\theta = 2$**	**Recalibrated under $\theta = 0.5$**
Calibrated β	-	0.5030	0.4319	0.7473
Calibrated δ	-	0.9880	0.9941	0.9733
Calibrated θ	-	1.1051	2.0	0.5
Frac. borrowing, 21-30	0.815	0.608	0.537	0.576
Frac. borrowing, 31-40	0.782	0.714	0.603	0.661
Frac. borrowing, 41-50	0.749	0.838	0.715	0.792
Frac. borrowing, 51-60	0.659	0.853	0.656	0.829
Avg. debt to income, 21-30	0.199	0.241	0.192	0.207
Avg. debt to income, 31-40	0.187	0.272	0.228	0.240
Avg. debt to income, 41-50	0.261	0.319	0.253	0.285
Avg. debt to income, 51-60	0.276	0.330	0.219	0.316
Avg. wealth to income, 21-30	1.23	0.999	0.860	1.075
Avg. wealth to income, 31-40	1.86	1.471	1.458	1.534
Avg. wealth to income, 41-50	3.24	3.091	3.207	2.970
Avg. wealth to income, 51-60	5.34	6.271	6.683	5.517
Goodness-of-fit	-	221.34	311.06	217.03

Table K.3 presents the numerical results of a battery of robustness checks.

Table A3: Robustness checks

Check	**Calibrated parameters**	**Pareto-efficient R^Z**	**CE welfare loss**	**Retirement savings**
Benchmark	$k_1 = 1.5083$ $\gamma_1 = 0.5$ $k_2 = 2.650 \times 10^{12}$ $\gamma_2 = 5.75$ $\xi = 0.481$	5.25%	0.23%	−8%
$K(R^Z, Z_t) = k_1 + k_2(R^Z - R^X)^{\gamma_2}$	$k_1 = 793.36$ $k_2 = 4.6947 \times 10^{12}$ $\gamma_2 = 5.90$ $\xi = 0.2465$	5.25%	0.23%	−8%
$K(R^Z, Z_t) = k_1 + k_2[(R^Z - R^X)Z(t)]^{\gamma_2}$	$k_1 = 793.36$ $k_2 = 1.3966 \times 10^{-19}$ $\gamma_2 = 5.45$ $\xi = 0.7523$	5.25%	0.23%	−8%
Alt. business model	$k_1 = 1.5083$ $\gamma_1 = 0.5$ $k_2 = 4.695 \times 10^{12}$ $\gamma_2 = 5.90$ $\xi = 0.3852$	5.25%	0.17%	−8%
Competition (markup of 0)	$k_1 = 1.5083$ $\gamma_1 = 0.5$ $k_2 = 2.1895 \times 10^{12}$ $\gamma_2 = 5.70$ $\xi = 0.1262$	5.25%	0.26%	−8%

Monopoly (no markup target)	$k_1 = 1.5083$ $\gamma_1 = 0.5$ $k_2 = 5.7629 \times 10^{11}$ $\gamma_2 = 5.35$ $\xi = 4.4842$	6%	1.14%	-26%
Variable fee 0.005 (markup of 0.30)	$k_1 = 1.2709$ $\gamma_1 = 0.5$ $k_2 = 3.822 \times 10^{15}$ $\gamma_2 = 7.75$ $\xi = 0.0414$	5.25%	0.26%	-13%
$\theta = 2.0$ ($\beta = 0.4319$, $\delta = 0.9941$)	$k_1 = 1.8281$ $\gamma_1 = 0.5$ $k_2 = 4.633 \times 10^{8}$ $\gamma_2 = 3.40$ $\xi = 0.00013$	5.50%	0.42%	-5%
$\theta = 2.0$ ($\beta = 0.4319$, $\delta = 0.9941$); monopoly	$k_1 = 1.8281$ $\gamma_1 = 0.5$ $k_2 = 1.008 \times 10^{8}$ $\gamma_2 = 3.00$ $\xi = 0.0012$	6.50%	2.61%	-14%
$\theta = 0.5$ ($\beta = 0.7473$, $\delta = 0.9733$)	$k_1 = 1.1199$ $\gamma_1 = 0.5$ $k_2 = 8.426 \times 10^{15}$ $\gamma_2 = 8.05$ $\xi = 26.728$	5.25%	0.13%	-57%
$\theta = 0.5$ ($\beta = 0.7473$, $\delta = 0.9733$); monopoly	$k_1 = 1.1199$ $\gamma_1 = 0.5$ $k_2 = 5.754 \times 10^{15}$ $\gamma_2 = 7.95$ $\xi = 311.151$	3.50%	1.37%	$+100\%$

Partial naiveté ($\hat{\beta} = 0.87575$)	$k_1 = 1.5109$ $\gamma_1 = 0.5$ $k_2 = 8.114 \times 10^{14}$ $\gamma_2 = 7.25$ $\xi = 0.4815$	5.00%	-	-
Partial naiveté ($\hat{\beta} = 0.7515$)	$k_1 = 1.5280$ $\gamma_1 = 0.5$ $k_2 = 5.677 \times 10^{14}$ $\gamma_2 = 7.15$ $\xi = 0.3852$	5.00%	-	-

Notes: Unless stated otherwise, the calibrations target the benchmark moment values.

Some of the above results require additional explanation. The calibration with an alternative objective function for the provider follows the theoretical extension modelling an alternative business model of the pension industry. That is, while the investment cost component of the provider's cost function includes the cost of generating returns exceeding the rate of return R^Z promised to the agent, the provider's revenues consist of the re-invested fee as well as any capital gains over and above R^Z:

$$\tilde{\pi}_t = (1/R^X)\big[(1 + R^Z + \tau)f + \tau Z_t\big] - K(R^Z + \tau, Z_t)$$

where $\tau \geq 0$ denotes the wedge between the total return generated by the provider and the return promised to the agent. As it turns out, for parameter constellations that satisfy the cost and markup targets, the optimal (profit-maximising) wedge is zero. This is likely due to the fact that when the cost function of the provider is steep enough, so that a naïve agent is not willing to compensate her for rates of return exceeding 5%, the provider herself also finds the cost of generating a positive wedge prohibitively high.

In the case of monopoly, the contract offered in market equilibrium is simply the one that maximises the provider's profits while extracting the entire perceived surplus from the agent. An agent would actually be better off paying a monopolistic fee for a contract with $R^Z = 6\%$, but the shift from the target rate of return of 5% to the more expensive contract can no longer be interpreted as a Pareto improvement. That is because the monopolist makes a strictly higher profit when offering the cheaper contract.

When accounting for variable fees, I assume that every contract imposes a variable fee equal to 0.5%, in addition to the fixed fee. The model is calibrated to match a higher average markup in order to guarantee that the discounted present value of the provider's profits is non-negative. As the provider's revenue now consists of the variable and the fixed fee, the calibrated fixed fee is less than 20% of the fixed fee in the benchmark calibration. The original conclusions not only remain valid, but the presence of variable fees makes the quantitative effects of inefficient contracting slightly more pronounced.

The equilibrium savings contract turns out to be inefficiently cheap also for the household parametrisation with $\theta = 0.5$. That the case is because in a dynamic setting the timing effects can distort the textbook relationship between the interest rate and optimal savings. More precisely, even though $\theta < 1$, the household starts saving for retirement earlier on in the life cycle as the interest rate increases, resulting in the total amount saved being *decreasing* in R^Z for values around 5%. Then, the providers optimally offer an inefficiently cheap savings contract, taking advantage of overestimation of future savings at the contracting stage. In contrast, a monopolist interacting with a household parametrised by $\theta = 0.5$ offers an inefficiently expensive contract, consistent with the prediction of the simple model. Because of a very high intertemporal elasticity of substitution, the small differences in contract terms have a huge impact on the agent's savings. On the other hand, the parametrisation with $\theta = 2.0$ indicates the prevalence of inefficiently cheap savings contracts, both under a competitive environment and under a monopoly. In this case, the impact of exploitative contracting on consumer welfare is much larger than under the benchmark calibration.

To examine how the assumed degree of naiveté affects the numerical results, I repeat the analysis for two cases with $\beta < \hat{\beta} < 1$, keeping the level of present bias constant at $\beta = 0.503$. Then, the agent's actual choices are given by a solution to a problem of a household with present bias β and beliefs $\hat{\beta}$, while the expected choices solve a problem of a sophisticated household with present bias $\hat{\beta}$. In particular, the beliefs at the contract evaluation stage and throughout the life cycle coincide. The two values for $\hat{\beta}$ that I consider gradually reduce the severity of the agent's naiveté.[12]

Apparently, departing from complete naiveté eliminates Pareto-inefficiency of the equilibrium contracts. The intuition behind this result is not immediate, however. Partial sophistication generates two separate effects, relative to the case of complete naiveté. First, the actual behaviour of a household characterised by present bias β changes, as different predictions of behaviour of the future selves are taken into account and potentially give rise to complicated strategic considerations (Harris and Laibson, 2001).[13] Second, the expected future behaviour now corresponds

[12]Note that $0.75 \times 1 + 0.25 \times 0.503 = 0.87575$ and $0.5 \times 1 + 0.5 \times 0.503 = 0.7515$.

[13]The analysis of a three-period model by Salanié and Treich (2006) reveals that whether sophisticates or naifs save more for any given magnitude of the present bias depends on the curvature of the instantaneous utility function. Under a CRRA formulation with $\theta > 1$, the 'precautionary' motive of sophistication dominates the 'discouragement effect' and it is the sophisticates who accumulate more savings. The reverse is true for $\theta < 1$. With a logarithmic utility function ($\theta = 1$), savings are independent of the degree of naiveté.

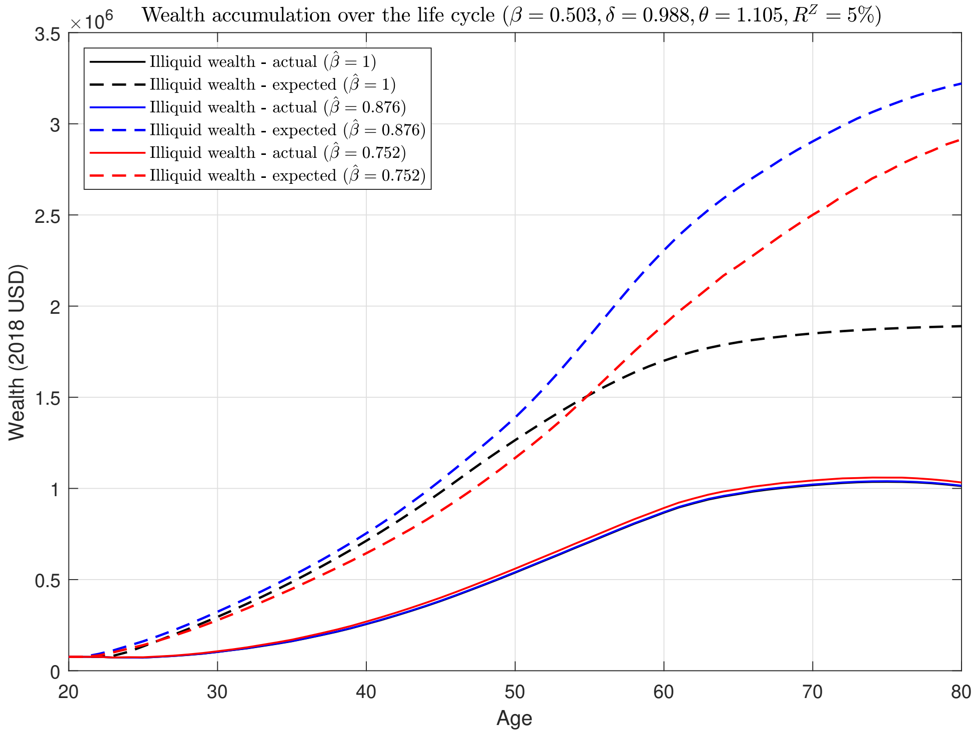

Figure A1: Wealth accumulation paths (expected and actual)
Source: Author's calculations

to that of a sophisticated agent with present bias $\hat{\beta}$, which can depart from the behaviour of an exponentially-discounting agent in a non-trivial way. These effects are illustrated in Figure K.1.

The graph illustrates that the impact of partial naiveté on actual wealth holdings is negligible. In other words, holding β constant, there is little impact of different beliefs $\hat{\beta}$ on the agent's actual accumulation of pension wealth. This is consistent with remarks made by Laibson et al. (2018) who conduct their benchmark calibration under the assumption of complete naiveté but argue that this has little impact on generated behaviour, while providing greater numerical stability relative to the case of (partial) sophistication with possibly discontinuous policy functions. In contrast, the impact of partial naiveté on expected wealth holdings is substantial and non-monotonic in $\hat{\beta}$. First, changing $\hat{\beta}$ from 1 to 0.876 *increases* the expected holdings of pension wealth and thus exacerbates the overestimation of future usage at the contracting stage. That is because a sophisticated agent with a moderate present bias compensates for his future self-control problem by accumulating a larger proportion of his wealth in the form of an illiquid asset than an agent who discounts exponentially. Second, lowering $\hat{\beta}$ further to 0.752 *reduces* the expected holdings of pension wealth because the impact of contemporaneous present bias starts playing a bigger role.

Thus it seems that partial naiveté alleviates concerns related to Pareto-inefficiency of equilibrium outcomes. Note, however, that for the three values $\hat{\beta} \in \{1, 0.876, 0.752\}$, the present-biased agent obtains very similar levels of utility from signing a private pension contract. Thus distributional concerns might still call for a regulatory intervention. This point is made clearer by the fact that in both instances of partial naiveté, the agent's willingness to pay for a contract with a given rate of return is substantially higher than under complete naiveté. At $R^Z = 5\%$, the willingness to pay is as much as 30% higher under partial naiveté.

Finally, for computational feasibility, the numerical model assumes that the agent's consumption-saving decisions are independent of the fee charged by the provider. I test whether it is an overly restrictive assumption by simulating behaviour of a household that re-optimises their consumption and wealth accumulation paths, taking as given the level of fees from the benchmark calibration. Relative to the model that allows the household to re-optimise, the benchmark model overestimates the actual pension wealth accumulation by 3% (at $R^Z = 5\%$), as the agent does not take into account a decrease in his lifetime resources due to fees. The difference between the expected pension wealth holdings, which drive the contractual choice, is much smaller at only 1%. Moreover, the agent's life-cycle consumption path, which ultimately determines the perceived and actual utility from accepting a contract, is hardly affected by the simplifying assumption of no re-optimisation. Average *absolute* deviation of the actual simplified consumption path from the re-optimised one equals 0.28%. In case of the perceived consumption path, the average absolute deviation is larger at 1.86%, but any meaningful differences materialise only between the ages of 50 and 60.

After accounting for household's re-optimisation, a contract with $R^Z = 5\%$ still maximises the agent's perceived utility, while his actual utility would be maximised by the choice of $R^Z = 5.25\%$, thus supporting the conclusion regarding Pareto-inefficiency of the market equilibrium. Under this specification, choosing an inefficient contract generates a CE loss of consumer welfare of 0.08% and lowers the agent's savings at retirement by 7%. The magnitudes of the quantitative effects are smaller, because the fees implied by the benchmark calibration are 'too high' for contracts with $R^Z > 5\%$, but 'too low' for contracts with $R^Z < 5\%$, once the household is able to re-optimise. Nonetheless, the existence of Pareto-improving contract terms in the direction predicted by the simple model is confirmed.