Pricing and Incentives in Defined Contribution Retirement Systems

Sam Flanders* Melati Nungsari† Marcela Parada-Contzen‡

July 27, 2018

Abstract

This paper studies the effects of different pricing schemes on overall surplus in a privately managed retirement system with multiple service providers and switching costs. We develop a theoretical model based on the Chilean retirement system and consider a repeated auction for monopoly rights over new enrollees by comparing three different pricing schemes: 1) fees over contributions, 2) fees over returns, and 3) a two-part tariff implemented by first conducting an auction over a guaranteed rate of return, and then allowing the firm to keep a portion of returns generated above this guaranteed rate. We prove that the third pricing scheme maximizes market efficiency. We also show that allowing price discrimination can improve consumer surplus, as firms charging a single price may choose not to serve low value customers.

JEL Classification: J26, J32, J38

Keywords: mandatory pension systems, management fee, auctions.

*Asia School of Business and MIT Sloan School of Management. Sasana Kijang, 2 Jalan Dato' Onn, 50480, Kuala Lumpur, Malaysia. Phone: (60)13-207-3225. E-mail: sflander@mit.edu.

†Asia School of Business and MIT Sloan School of Management. Sasana Kijang, 2 Jalan Dato' Onn, 50480, Kuala Lumpur, Malaysia. Phone: (601)13-290-3225. E-mail: melati@mit.edu.

‡Instituto de Economía - Pontificia Universidad Católica de Chile. Address: Vicuña Mackenna 4860, Macul, Santiago, Chile. Phone: (56-2)-23547115. E-mail: marcela.parada@uc.cl. Marcela thanks the funding granted by Fondecyt, Project No. 3180155.

1 Introduction

In privately managed, defined contribution (DC) retirement system, savings are usually managed by companies known as pension fund administrators (PFAs). For its services, the PFAs charge a fee set as a share of contributions, benefits, assets managed, or as an annual fixed amount (Creighton and Piggott, 2006). In this paper, we theoretically study the surplus-maximizing, and hence efficiency-maximizing, pricing scheme in a privately managed DC system. We consider the main institutional settings of the iconic Chilean system, which has served has as an archetype for the implementation of private retirement systems around the world (Orszag and Stiglitz, 2001). We compare total market surplus under three pricing schemes: 1) the current structure with fees over contributions, 2) an alternative scheme with fees over returns, and finally, 3) a two-part tariff (i.e., fee over contributions and returns) where where firms bid on a guaranteed rate of return.

Because of its effect on pension levels, fee levels has been of continuous debate in the literature (Tuesta, 2014). Previous evidence show that fees substantially reduce replacement rates (Modigliani and Muralidhar, 2004), pension returns (Bateman et.al., 2001), and may have undesirable distributional impacts (Baily and Kirkegaard, 2008). Despite the evidence on the effect of fees on pension outcomes, its effects on the efficacy of retirement system have received limited attention (Creighton and Piggott, 2006).

The architecture of DC systems vary across countries. Specifically, the fee structure depends on the maturity of the financial sector and the PFAs market upon introduction. For example, in Latin America, because there was a concern that entry and competition was going to be insufficient (Dobronogov and Murthi, 2005), because financial markets where still undeveloped (Tuesta, 2014), and because it was not clear whether the industry would have the initial capacity to charge fees on assets (Lasaga and Pollner, 2003), an up-front fee over contributions was implemented. In Eastern Europe, where well-established Western European financial companies are interested in operating in the market, there is a bigger presence of fees on returns (Dobronogov and Murthi, 2005).

This paper contributes to the literature on pricing and incentives in retirement markets. Due to market differences and because counterfactual evidence on running DC systems is rare, one issue that remains unanswered is the effect of alternative pricing schemes on total welfare. To the best of our knowledge, this is the first paper to study welfare effects of different pricing schemes in a privately-managed DC retirement market. Our main result is the following: one pricing scheme that provides the proper incentives to induce a first-best outcome is one where firms bid on a guaranteed rate of return for consumers. Under this pricing scheme, firms charge a fixed fee over contributions for a promised return to consumers and, if they are able to obtain to generate a higher return, keep a proportion of the additional return above the promised rate to themselves.

A novelty of this paper is that we explicitly model the auction dynamics for monopoly rights to new enrollees. This policy reform was introduced in Chile in 2008. Since 2010, new enrollees are automatically assigned to the firm with the lowest bid for 24 months. Auctions to assign monopoly rights over enrollees have been recommended for increasing competition and decreasing fees (Fisher et.al., 2006; Kurach et.al., 2017). After the Chilean experience, auctions were also implemented in Peru and have received the attention of policy makers in Mexico, Australia, New Zealand, and Poland (Mesa-Lago, 2016; Chomik, 2015; Heuser et.al., 2015; Kurach et.al., 2017).

1.1 Relevant Literature

As mentioned, there is a wide range of fee structures in retirement markets. In most Latin American and Eastern European countries, fees are charged as a share of contributions or as a combination between contributions and assets managed (Tapia and Yermo, 2008). Some, like Mexico and Peru, have transitioned from fees over contributions to fees over assets managed (Tuesta, 2014). Others like Costa Rica, Kazakhstan, Croatia, and Poland, PFAs have charged management fees over returns (Dobronogov and Murthi, 2005). Fees over assets managed or as a function of return tend to occur in voluntary pension plans, where political tensions with respect to levels and forms of fees has been less severe (Tuesta, 2014).

The idea of linking firms' revenues with their performance seeks to incentivize PFAs to maximize the value of assets or returns (Whitehouse, 2001). This argument replicates the functioning structure of the mutual fund management industry, which might be one of the closets industry to the PFAs market. Even though the mutual fund industry tends to charge higher fees than PFAs (Tuesta, 2014), the effect of pricing incentives for achieving better performance has been well documented both theoretically and empirically (Warner and Wu, 2011; Basak and Pavlova, 2013; Koijen, 2014).

Linking performance to fees also incentivizes product differentiation as firms have more incentive to perform well (i.e., providing higher returns) and therefore increase competition. When fees are charged over contributions, firms have not incentive to differentiate. In fact, the evidence show that PFAs tend to herd, resulting in firms buying or selling at the same time (Raddatz and Schmukler, 2013).

Another area of controversy has been the low competition in the PFAs market. Generally, there is evidence that concentration and high fee levels occur in systems with savings plans that are not exclusive to a group workers in a firm, industry, or occupation (Holzmann et.al., 2005). The low competition is explained by market characteristics, such as high exit barriers and non-homogeneity (Kurach et.al., 2017). There also demand-side factors such as little price response. There is evidence that consumers do not pay attention to fees but to other factors, such as firm characteristics (Kurach et.al., 2017; Hastings et.al., 2017). The little price response is also explained by financial illiteracy (Mitchell et.al., 2007), behavioral responses with respect to how information is presented (Hastings et.al., 2011), or switching costs (Illanes, 2016; Luco, 2016).

For increasing competition and consumers' responses to prices, Chile implemented biennial auctions for monopoly rights to new enrollees (Berstein et.al., 2009)). The auction-winning firm is the one that offers to lowest fee for the next 24 months. The design is such that the winning bid must be below the current winning fee. Since the auctions between the firms in

Chile are conducted every two years, this paper relates to the literature on repeated auctions.

A commonly studied problems in repeated auctions is one of collusion (McAfee and McMillan, 1992). Bidders in repeated auctions can collude through the adjustment of continuation payoffs in a way that partially compensates for the lack of monetary transfer (Aoyagi, 2003). This collusion can occur even without explicit communication and limited monitoring (Skrzypacz and Hopenhay, 2004). In this paper, we sidestep the issue of collusion since 1) the Chilean retirement market is heavily monitored and regulated by the government, and 2) there have only been 4 auctions conducted in the history of the market. These two facts provide sufficient proof to suggest that side payments and transfers are not possible.

2 Chilean Retirement System: Background

2.1 Participants and Auction Mechanism

Since the model presented in this paper is based on demand and structural characteristics of the Chilean system, we first present a condensed background on this system. First of all, participants in the Chilean retirement system are either enrollees, PFAs, or the government. The role of the government is to supervise the PFAs market through a special division called the Superintendence of Pensions (SP).

Enrollment is mandatory and automatic for employees. They are required to save a 10 percent of their wage if working (contributors). Self-employed workers and non-employed individuals may voluntary enroll. Since voluntary participation has been historically low, in this paper we focus on mandatory savings.[1]

The PFAs are for-profit private firms whose objective is to manage individuals savings and pensions. Details about quantities and shares are presented in Table B1. Firm can participate in the auctions by bidding to monopoly rights for new enrolles for 24 months.

[1] As of December 2016, mandatory contributors represented a 97 percent of market participation (see Table B1 for details).

The auction is assigned to the firm that offers the lowest fee for 2 years. Individuals enrolled before July of 2010 are not subject of being auctioned, although they can freely chose to enroll in any firm, including the auction-winning firm. After 24 months, auctioned individuals can chose whether to stay or to transfer to another firm.

The regulation considers special circumstance under which new enrollees can transfer to another firm during the auctioned period. For instance, transferring to another firm is possible under the following circumstances: 1) if the auction-winning-firm does not charge the lowest fee for 2 consecutive months as some other firm decreased its fees, 2) if the return provided by the firm is under the minimum required return in the market, and 3) if the lowest fee does not compensate the financial gain that the individual could have earned in a different firm.

2.2 Fee structure

The PFAs charge a fee over individual income which is paid by contributors out of their monthly mandatory contribution rate. We refer to this fee as "fee over contribution". The fee is firm-specific as it does not depend on the investment fund (product) that the individual chooses. Fees are the main source of revenue for firms. As of June 2016, fees represent more than 90 percent of total revenue.[2]

PFAs define the fee within the legal framework: for increasing fees the firm must give a 90 days notice and for decreasing fees a 30 days notice. Currently, the fee ranges between 0.41 and 1.54 percent, with a mean fee of 1.15 percent.

The evolution of market fees between 1988-2016 is presented in Figure A1 and fees per firm for the years 2010-2016 in Figure A2. Starting in 2008, the average market fee has decreased substantially, partly due to the implementation of the auction mechanism in 2010. Overall, empirical observations suggest that firms do not raise fees and maintain the same fee for

[2]Before 2008, there was also a lump-sum fee. It was eliminated in 2008 as it was not an important component on firms revenue and had an important effect on individuals savings balances, specially for low-income individuals (see Castro (2005)).

long periods of time. Auction-losing firms tend to maintain their fees, while auction-winning firms' fees tend to drop over time.[3] Consistent with this empirical observation, our model assumes that the auction winner is the lowest-priced provider for the period in which it holds a monopoly.

2.3 Product Description

The PFAs offer five investment funds defined by the system (known as Accounts A, B, C, D, and E). These accounts vary in their level of financial risk, which is defined within an interval (see Table B2). Investments portfolios are publicly available by the SP. Monthly returns have ranged within -22 to 10 percent, with standard deviations from one to four percent (see Table B2). Long-run average yearly returns (annualized rate) range between three and six percent (see Table B3 for 12-months, 36 months, and total returns on investment funds per account).

PFAs are responsible to achieve a fund-specific return above a minimum level.[4] In case that the minimum level is not achieved, firms need to compensate individuals. The objective of this policy, rather than incentivize high returns, is to protect enrollees' assets. Figure A3 show the evolution of returns and the bounds. The two lines in the bottom represent the bounds. It is also easy to observe that firms' returns do not substantially deviate from each other and follow the same pattern. We consider this as empirical evidence that firms have similar cost structures, and hence similar portfolios – an assumption that is reflected in our model. In order to increase retirement wealth for enrollees, it is crucial to regulate proper pricing schemes to incentivize firms to actively to care about the returns they are generating. In this paper we explore how different pricing schemes can help incentivize firms' performances in this market.

[3]After the reform, consumers turnover has increased, but the change is not substantial. For example, between 2010 and 2016, consumers turnover increased from three to six percent. Therefore, for the model, we consider that consumer response to fees remains small.

[4]This minimum level is defined by the lowest return between: (1) the mean annual return for the last 36 months for all the firms in the market minus four percent for Accounts A and B, and minus two percent for Accounts C, D and E; (2) the mean annual return for the last 36 months for all the firms in the market minus the (absolute value) 150 percent of such return.

3 Model

In this section, we present the setup of the model by describing the environment – i.e. consumers, firms, and the timing of for the auction. We assume that all firms participate in a first-price auction for monopoly rights over new customers.[5] Additionally, in the baseline model, the effort that the firm exerts in generating returns for customers is not modeled. To begin, we also assume that the PFAs charge fees over contributions. In Sections 6 and 7, we extend the baseline model to include alternative pricing schemes such as fees over returns and a two-part tariff. In Section 8, we further extend the model and allow for consumer heterogeneity in income.

3.1 Consumers

Consumers live for three periods – young, middle-aged, and old. When consumers are young, they are automatically assigned to the firm with the lowest bid in the auction for monopoly rights from the previous period. Consumers work when young and middle-aged and obtain the same amount of income in each period.[6] Consumers then save a proportion of their monthly income in their retirement account. When consumers are middle-aged, they have the option of switching to any other firm in the market but incur a switching cost s of doing so. When old, consumers retire, face no decisions, and consume the entirety of their savings.[7] To simplify the analysis, assume that there is a mass of size $\frac{1}{2}$ of young consumers and a mass of size $\frac{1}{2}$ of middle-aged consumers. Consumers discount their expected lifetime utility at rate δ.

[5] We explored a second-price auction and arrived to the same conclusions described in Section 4 with a first-price auction but with a higher likelihood of collusion. Therefore, the first-price auction is a superior design. The proofs are available from the authors.

[6] The model where income increases over a consumer's life-cycle yields the same results as in this paper and can be provided to the reader upon request.

[7] Consumption is not modeled in this paper for periods when the consumer is young and middle-aged – workers simply work to obtain a wage and save a proportion of it.

3.2 Firms

There are 3 firms who compete for consumers.[8] The firms participate in a sealed bid, first-price auction for monopoly rights to all the young consumers in the subsequent period. Each firm faces the provision cost of c per dollar-invested for consumer. We assume that the production technology is the same for all firms and therefore all firms face the same cost function

3.3 Timing

In each period t, the auction to determine monopoly rights over young consumers in period $t+1$ is held. All firms participate by submitting a bid to the regulator. If a firm wins the auction by submitting the lowest bid, it will have monopoly rights for one period starting at time $t+1$. The winning firm is not allowed to change its fees but the firms who lose the auction are free to set any fee above the minimum bid and do not need to implement their bid. In each and every period, a firm can be in one of two possible states with respect to the auction – it was either a winner from the previous period's auction, in which case we call it an **incumbent**[9] in this period, or it lost the auction from the previous period, in which case we call it an **entrant** in this period. Incumbents enter a period with monopoly rights over all young consumers in the period. Since incumbents have monopoly power, they have a higher expected profit and thus, higher valuations. This being, since firms compete Bertrand-style and are forced to bid at their valuation, in a steady-state equilibrium it must be that incumbents necessarily transition to being entrants in the next period.

[8]The model is constructed using a finite number of firms, echoing the Chilean retirement market, where there are currently 6 providers. 3 firms is the minimum amount of firms that is required to allow us to study competition between the two losers of the auction.

[9]The incumbency is with respect to the auction, not with respect to entry into the market

4 Fees over contributions, effort not modeled

We first consider the steady-state equilibrium in pure strategies. Since the three firms compete Bertrand-style, competition forces the firms to choose fees such that their expected utility over the lifetime of the customer is zero. This being, a firm may end up charging below marginal cost in one period to generate positive surplus in the next period. It is important to note that winning the auction in one period increases a firm's expected utility since it now has monopoly rights over new enrollees. Since Bertrand competition forces firms to bid at their valuations in equilibrium, and since these valuations are determined by a firm's expected utility, a firm that has won an auction in period t (i.e. the incumbent in period $t+1$) necessarily transitions to becoming an entrant in period $t+1$. This is because the incumbent in $t+1$ will have a higher valuation due to monopoly power in period t, which means that it will place a higher bid and lose in the auction held in period $t+1$.

In equilibrium, there are two prices of relevance – the first being p_a, the lowest bid offered in the auction in a period t. This bid was necessarily put forth by an entrant in period t. The second price os p_I, which is the price charged by the incumbent in period t (i.e. the winner from period $t-1$). Since a firm can be in one of two states in each period (entrant or incumbent), and due to the fact that an incumbent in a period t cannot win the auction again in period t, the competition to win the auction in period t is between two entrants. Thus, the probability of winning the auction for any entrant is $\frac{1}{2}$. In the case where the entrant wins the auction in period t, it charges p_a minus c, its provision cost, to all of the $\frac{1}{2}$ young consumers in period t, and necessarily becomes an incumbent in period $t+1$. If the entrant does not win the auction in period t, it remains an entrant in period $t+1$. Furthermore, an incumbent enters a period t with a mass of middle-aged consumers attached to it. Given the switching cost s, to avoid its customers from switching, the incumbent charges $p_I = p_a + s$ to the mass $\frac{1}{2}$ of middle-aged consumers it has monopoly power over in period t, and proceeds to period $t+1$ as an entrant. Given these transitions and parameters, we describe the value functions below.

Definition 1. *The value functions for the incumbent and entrants are given by*

$$V_E = \frac{1}{2}(\frac{1}{2}(p_a - c) + \delta V_I) + \frac{1}{2}\delta V_E \tag{1}$$

$$V_I = \frac{1}{2}(p_a + s - c) + \delta V_E \tag{2}$$

Solving the system of two equations given by 1 and 2 yields that

$$V_I = \frac{c - p_a - s + \frac{\delta s}{2}}{\delta(1+\delta) - 2} \tag{3}$$

$$V_E = \frac{p_a - c + \delta(p_a - c + \delta)}{2 - \delta(1+\delta)} \tag{4}$$

Since the firms are indifferent between winning and losing, we are able to solve for the optimal p_a from the indifference condition, given below. The indifference condition provides a relationship between the value function for the entrant, V_E, and the value function for the incumbent, V_I. It states that in a steady-state equilibrium, an entrant is indifferent between winning the auction (and proceeding to the next period as an incumbent), and losing the auction (and staying an entrant in the next period).

Definition 2. *The indifference condition for the equilibrium in pure strategies is given by*

$$p_a - c + 2\delta V_I = \delta V_E \tag{5}$$

Thus, solving 3, 4, and 5 yields that

$$p_a = \frac{2c(1+\delta) - \delta(p_a + 2s)}{2+\delta} \text{ and } p_a = c - s\frac{\delta}{1+\delta} \tag{6}$$

Proposition 1. *The fees set by the incumbent and entrants in the pure-strategy Nash equilibrium are given by*

$$p_I = p_a + s = \frac{s}{1+\delta} + c \tag{7}$$

with the corresponding equilibrium value functions of

$$V_I > 0 \text{ and } V_E = 0 \ \forall \text{ parameter values} \tag{8}$$

The results from Proposition 1 can be interpreted very intuitively. The degree to which the incumbent can set high fees depends on the consumer switching cost, s. As s increases, consumers are less likely to switch away the incumbent after the monopoly period and thus, the incumbent charges high fees to capture this surplus. The auction price then can be seen as the inverse of the incumbent price – the auction winner must price competitively with respect to the switching costs. If the switching costs are higher, the auction winner must price lower in order to induce the consumers to subscribe to their services. In equilibrium, the entrant is not able to capture any value in expectation due to the fact that they do not have any market power over any segment of consumers, unlike the incumbent, who enjoys monopoly rights for a period. Thus, the incumbent's value function is positive but that of the entrant is zero.

4.1 Equilibrium in mixed strategies

Unlike the pure strategy setup where an incumbent in one round necessarily becomes an entrant in the next period, an incumbent in the mixed strategy equilibrium can still win in the next round but with a lesser likelihood. This is true since the bid distribution in equilibrium is unbounded – no matter how high a bidder places a bid, there is still a chance that the bid will win.

Let the bid strategy be p with a probability density function $f(.)$ and cumulative density function $F(.)$. In the following discussion, we will use F_E to be the cumulative density function for the entrant and F_I as the cumulative density function for the incumbent. Let p_L denote the minimum possible bid and p_W be the winning bid. Recall that the marginal cost of provision of services is c. There are then two cases–

1. $p_L \geq c$.

2. $p_L < c$.

For tractability, we restrict ourselves to the case of symmetric minimum bids, which allows us to disregard case 2. This is because if the winning bid was below cost, this must mean that the expected utility in that period is negative – thus, if any firm plays a bid such that $p < c$, they will make negative profits in that round and will leave the market.

For any auction in the mixed strategy equilibrium, there are two possibilities. The first is that the incumbent wins again with bid $p_W - s$, retains his young consumers from the last period (now middle-aged), and receives monopoly rights over the young consumers from this period. In this case, both of the entrants are left to complete Bertrand-style, and thus will both place bids equal to c.

The second possibility is that an entrant wins. Denote the winning entrant E_W and the loser as E_L. In this case, after the auction, E_W will play the bid p_W and E_L and the incumbent compete Bertrand-style for middle-aged customers. E_L and the incumbent will then be induced, by competition, to both bid at cost, playing c. However, since consumers have a switching cost s, the incumbent will win over E_L and keep its consumers from last period, who are now middle-aged.

Consider the two Bellman equations for the incumbent and entrant. Note that the Bellman equations are written assuming the bid placed is the minimum bid – hence the transition in the next period is necessarily to being an incumbent.

Definition 3. *The value functions for the entrants and incumbent are given by*

$$V_E = \frac{1}{2}(p_L - c + 2\delta V_I) \tag{9}$$

$$V_I = p_L - c + 2\delta V_I \tag{10}$$

The indifference condition must mean that all bids give an equally good payoff. Playing the minimum bid ensures winning, so the cumulative density function of the minimum bid for the incumbent is $1 - (1 - F_E)^2$. The indifference condition is then

$$p_L - c + \delta V_I = (p_L - c + \delta V_I)(1 - F_E)^2 + (s + \delta V_E)(1 - (1 - F_E)^2) \tag{11}$$

Proposition 2. *In the mixed-strategy Nash equilibrium, the cumulative density functions of the entrants and incumbent are given by*

$$F_E(p) = 1 - \frac{\sqrt{-2s + p_L(2+\delta) - c(2+\delta)}}{\sqrt{2p - 2s + p_L\delta - c(2+\delta)}} \tag{12}$$

$$F_I(p) = \frac{(-p\alpha - c(1+\delta)(-\alpha+\beta) + p_L(-\delta\alpha + \beta + \delta\beta))}{(-p + c - p_L\delta + c\delta)\sqrt{-2s + p_L(2+\delta) - c(2+\delta)}} \tag{13}$$

where

$$\alpha = \sqrt{-2(c+s) - c\delta + p_L(2+\delta)}$$

$$\beta = \sqrt{2p - 2s + p_L\delta - c(2+\delta)}$$

Proof. To solve for Equation 12, solve a system of three equations given by 9, 10, and 11.

Then, let's solve the entrant's problem in order to solve for F_I. If the entrant loses, there are two cases. The first is that the other entrant wins. The second is that the incumbent wins again, which means that both of the entrants compete for middle-aged consumers. If the incumbent bids below c, the entrant will not compete. If the winning bid is above c, the entrant competes Bertrand with the other losing entrant and earns zero profit – this case is ignored. This give the entrant's indifference equation as

$$(\frac{p_L}{2} - \frac{c}{2} + \delta V_I) = (\frac{p}{2} - \frac{c}{2} + \delta V_I)(1 - F_I)(1 - F_E) + (F_E + F_I - F_E F_I)(\delta V_E) \tag{14}$$

Substituting the previous equations into Equation 14 allows us to solve for the cumulative density function for the incumbent.

□

5 Fees over contributions with unobservable effort

Now, we will model the level of effort exerted by the firm in obtaining returns for its customers. In this section, the firm can choose a level of effort which is not observable by the customers. As the reader will soon note, even with this strict assumption on observability, the first-best outcome can still be achieved. Hence, in this paper, we did not consider observable effort.[10] Additionally, given the evidence from the literature on consumer financial knowledge from Section 1.1, assuming unobservable effort is a better representation of reality compared to the assumption of perfectly observable effort.

Each firm can invest effort e per consumer in their portfolios in each period. Consumers

[10]The analysis for observable effort can be obtained from the authors.

can never of observe this effort level but they do observe returns. Given e, consumers face an expected return of $r(e)$, where $r(e)$ is strictly increasing, positive, concave, and $r^{'}(0) > 1$. Bids are placed before effort is determined and we assume that consumers know the set of possible firm strategies. We use the following notation – when bidding, the bid level is p_a with a corresponding effort level e_a and the incumbent (i.e. the auction winner from the previous period) has a bid p_I with a corresponding effort e_I. In the following definition, we define the consumer utility function and value functions both the entrants and incumbent.

Definition 4. *Let p be the fee that consumers pay to the firms. Let the indicator variable I_{switch} be such that I_{switch} is 0 if there is no switching and 1 if consumer decides to switch, s is the switching cost, and the consumer utility in expectation is*

$$U = r(e) - p - I_{switch}s$$

The value functions for the entrant and incumbent are

$$V_E = \frac{1}{4}(p_a - e_a + 2\delta V_I) + \frac{1}{2}\delta V_E \tag{15}$$

$$V_I = \frac{1}{2}(p_I - e_I) + \delta V_E \tag{16}$$

with the indifference condition $\frac{(p_a - e_a + 2\delta V_I)}{2} = \delta V_E$.

Proceeding as we did in Section 4, we use Equations 15, 16, and the indifference condition to first solve for the equilibrium p_a, which is given by

$$p_a = e_a - e_I\delta - p_I\delta \tag{17}$$

Substituting Equation 17 into the value functions in Equations 15 and 16 enables us to solve for the equilibrium levels of effort, given below.

Proposition 3. *The equilibrium level of effort in the model with unobservable effort is a corner solution at zero, i.e.,*

$$e_I^* = 0 \text{ and } e_E^* = 0$$

The intuition for the proof is simple: Since effort is only observed by consumer after a firm wins the auction, it is strictly dominant to exert less effort. Thus, effort will be zero for entrants who win the auction, as well as for the incumbent who has no incentive to exert a nonzero amount of effort. Now, consider the firm profit function, given by

$$(p - e)q$$

where p is the price that the firm charges, e is the effort level it exerts, and q is the mass of consumers that it serves. Since effort is unobservable, e is independent of q. Thus, we can rewrite the profit function as

$$p\,q - e$$

This being, the profit function is clearly maximized when effort is minimized. Denote the lowest amount of effort as e_0.

Proposition 4. *The equilibrium prices for the winner of the auction (p_a^*) and the incumbent (p_I^*) are given by*

$$p_a^* = c + \frac{\delta s}{1 - \delta}$$

$$p_I^* = p_a^* + s = c + \frac{s}{1 - \delta}$$

Proof. Substituting in e_0 into the Equation 17 gives

$$p_a = e_0 - e_0\delta - p_I\delta = (1-\delta)e_0 - \delta p_I$$

Recalling from the baseline model with pure strategies that $p_I = p_a + s$, substituting into the equation for p_a gives

$$p_a = c - \frac{\delta}{1-\delta}s$$

Thus, if e_0 represents the firm's marginal cost, the unobservable effort case reduces to the baseline model with no effort. In summary, note that that the equilibrium outcome for unobservable effort with fees over contributions is the same as in Section 4, where effort was not modeled. □

6 Fees over returns with unobservable effort

Now, consider a similar setup from the previous subsection but assume that pricing is based on returns. This is, firms set a price p that is a percentage of the returns that it gets to keep for itself. Denote p_a as the price set through the auction and p_I as the price set by the incumbent. Firms can invest effort e per customer in their portfolios in each period. Given e, the generated return is $r(e)$, where the function r behaves as in Section 5.

Definition 5. *The value functions and indifference condition are:*

$$V_E = \frac{1}{4}(p_a(r(e_a) - e_a + 2\delta V_I) + \frac{1}{2}\delta V_E \tag{18}$$

$$V_I = \frac{1}{2}(p_I(r(e_I) - e_I) + \delta V_E \tag{19}$$

$$\textit{Indifference Condition: } \frac{(p_a r(e_a) - e_a + 2\delta V_I)}{2} = \delta V_E \tag{20}$$

Solving Equations 18, 19, and 20 yields that

$$p_a = \frac{p_a \delta r(e_a) + 2(e_a + e_I \delta - p_I \delta r(e_I))}{(2+\delta) r(e_a)} \tag{21}$$

Thus, the optimal value of p_a can only be solved given an explicit form for the function r. This being said, even without an explicit form for r, we can still draw some conclusions about the optimal range for p_a. In particular, we have the following proposition.

Proposition 5. *The effort level induced under this pricing scheme is inefficient.*

Proof. Suppose $p_a \geq 1$. Then a competitor can bid slightly below p_a with $p = 1 - \psi$ for a small $\psi > 0$ and win the auction with certainty. For a zero level of effort, we have $r(e_0) - e_0 > \epsilon$. Thus, for $\psi < \epsilon$, the competitor can make profit by undercutting, implying that this cannot be an equilibrium. Thus, p_a is necessarily less than 1, which means that the effort level must be below that which is optimal. □

In conclusion, if we allow the firms to charge fees over returns, by modeling effort as being unobservable, we see that the effort level induced under this pricing scheme does not implement the first-best outcome. Thus, this pricing scheme is inefficient and undesirable. This agrees with the results from Section 5, where zero effort was exerted in equilibrium.

7 Two-part tariffs with unobservable effort

Now, let's consider a two-part tariff, whereby consumers are charged a fee over contributions (i.e. the fixed fee portion of the two-part tariff) and a fee over returns (i.e. the variable fee portion). Suppose that in each period, firms can choose to invest effort e in each portfolio. Given e, the expected return is $r(e)$, where $r(e)$ is increasing and concave, with $r(e_0) - e_0 > \epsilon$ for some $\epsilon > 0$, where e_0 is the minimum level of effort that is possible.

The auction is held before effort is determined. Suppose that consumers know firm strategies, but we restrict ourselves to non-reputational equilibria where consumers expect firms to only put in the effort that is individually optimal without reputation effects, and therefore firms do not invest effort to maintain reputation. Suppose also that there are two parts to the pricing – an initial price f per dollar invested and a return-share p.

The incumbent wins the auction with prices f_I and p_I and effort level e_I if the consumer gains from buying from it than a competitor with prices f, p and effort e,

$$r(e_I)(1 - p_I) - f_I \geq r(e)(1 - p) - f \tag{22}$$

Define $u_I = r(e)(1 - p) - f$, the minimum level of utility that the incumbent must provide in order to win the auction. Thus, given p_I, Equation 22 gives that $f_I = r(e_I)(1 - p_I) + u_I$. Consequently, we need only determine p_I and the value for f_I will be determined correspondingly.

The incumbent then chooses p_I to maximize profit, which is given by

$$\Pi_I = r(e_I)p_I + f_I - e_I \tag{23}$$

Equation 23 can be condensed by substituting in for the values of f_I:

$$\Pi_I = r(e_I)p_I + r(e_I)(1 - p_I) + u_I - e_I = r(e_I) - e_I + u_I$$

Since u_I is independent of the incumbent's optimizing decision, it will choose e_I^* such that

$$r^{'}(e_I^*) = 1 \implies p_I^* = 1$$

Thus, under this pricing scheme, incumbents also put in a first-best amount of effort.

Proceeding as we did in the previous sections, we now determine the prices that win the auction and the corresponding effort levels. A firm wins the auction if

$$f(e_a)(1-p_a) - f_a \geq f(e)(1-p) - f$$

Define $u_a = f(e)(1-p) - f$. Then, given by p_a, who f_a must be $f(e_a)(1-p_a) + u_a$ Given that a firm wins the auction, the expected utility for future periods is unconditional on prices. Thus, maximizing period utility is sufficient. The firm will choose p_a to maximize

$$\Pi_a = f(e_a)p_a + f_a - e_a = f(e_a)p_a + f(e_a)(1-p_a) + u_a - e_a = f(e_a) - e_a + u_a$$

Which, as the case for the incumbent, gives that $p_a^* = 1$ with a corresponding first-best effort level of e_a^*. Given e_a^*, e_I^*, p_a^*, and p_I^*, we have that $u_a^* = -f_a^*$ and $u_I^* = -f_I^*$.

The intuition for this result follows. If we allow the firm to change a per-return price, it will choose to keep every dollar of returns it gains for itself, giving the consumer nothing. To compensate the consumers for this, the firm will provide a negative pricing for each dollar invested, which is effectively a provision of a fixed return over the amount of the investments that the consumer has made.

Definition 6. *The value functions for the entrant and incumbent in this model are given by*

$$V_E = \frac{1}{4}(r(e) - f_a - e + 2\delta V_I) + \frac{1}{2}\delta V_E \tag{24}$$

$$V_I = \frac{1}{2}(r(e) - f_I - e) + \delta V_E \tag{25}$$

The indifference condition is given by

$$\frac{r(e) - f_a - e + 2\delta V_I}{2} = \delta V_E \tag{26}$$

Solving Equations 24 and 25 yields

$$V_I = \frac{2e + 2f_I + f_a\delta - f_I\delta - 2r(e)}{2(\delta^2 + \delta - 2)}$$

$$V_E = \frac{e + f_a + e\delta + f_I\delta - (1+\delta)r(e)}{2(\delta^2 + \delta - 2)}$$

Solving the system of three equations yields that

$$f_a = \frac{(f_a - 2f_I)\delta - 2e(1+\delta) + 2(1+\delta)r(e)}{2+\delta} \implies f_a^* = -\delta f_I - e(1+\delta) + (1+\delta)r(e)$$

Given that you are competing with a new incumbent, the outside option is $f_a - s$, and old incumbent utility is f_I. Thus, the consumer is indifferent when

$$f_a - s = f_I \implies f_I = f_a - s$$

This gives that

$$f_a^* = r(e) - e + \frac{s\delta}{1+\delta}$$

and

$$f_I^* = r(e) - e + \frac{s}{1+\delta}$$

In summary, we have proven the following proposition, which states that the two-part tariff described in this section can achieve the first-best outcome and is, consequently, one example of a pricing scheme that maximizes efficiency.

Proposition 6. *The first-best level of effort by both the incumbent (e_I^*) and auction winner (e_a^*) can be achieved under this two-part tariff scheme with the corresponding equilibrium prices $p_a^* = p_I^* = 1$ and $f_a^* = r(e_a^*) - e_a^* + \frac{s\delta}{1+\delta}$ and $f_I^* = r(e_I^*) - e_I^* + \frac{s}{1+\delta}$.*

8 Heterogeneous Customers

Until now, we have assumed that all customers are identical, both across individuals and over the lifecycle. We'll now relax that assumption with regard to wages. Allowing for time-varying wages over the lifecycle has little effect on the model other than changing the exact apportionment of surplus and prices over the lifecycle, so we'll treat the more interesting case of heterogeneity across individuals, modeled by high wage (w_H) and low wage (w_L) customers.

We now consider two cases: one where firms can price discriminate, and one where they must offer a single contract. Note that, in this Bertrand competition setting, firms cannot extract additional surplus from customers by discriminating. Instead, firms will effectively serve high and low income customers separately, with all the results carrying over from the one-type cases above. Thus, price discrimination will admit first best outcomes. By contrast, when firms are limited to a single price for all customers there are several possible sources of inefficiency.

The key factor in this model will be the relationship between wages on the one hand, and switching costs and provision costs on the other. If switching costs and costs of provision are proportional to wages, then, since prices are also proportional to wages, price discrimination is unnecessary and all the results from the one-type case carry over. Instead, we consider cost functions $s(I)$ and $c(I)$ where either

1. Switching costs are decreasing relative to wages:

$$\frac{\partial s(I)}{\partial I} < 0 \text{ while } c(I) = cI \text{ for some constant c, or}$$

2. The cost of provision is a decreasing fraction of wages:

$$\frac{\partial c(I)}{\partial I} < 0\text{while } s(I) = sI \text{ for some constant s}$$

Define the switching cost ratio as

$$s_r(I) = \frac{s(I)}{I}$$

and the provision cost ratio

$$c_r(I) = \frac{c(I)}{I}$$

In our two type setting, the above conditions reduce to $s_{rH} < s_{rL}$ and $c_{rH} < c_{rL}$ for I_H and I_L.

We're now ready to set up the two-type model for firms offering single contracts. We'll normalize the mass of L agents to $\frac{1}{2}$, and I_L to 1. Then $m_H > 0$ and $I_H > 1$ are free variables. The complication here relative to previous extensions is that, because they can only offer one contract, but with heterogeneous customers, the firms may want to offer a contract that only one type of customer will accept. Thus, there are two cases for both value functions. For model one (nonlinear switching costs), given the outside option of taking the auction price of p_A, old customers are willing to pay $p_A + s_r$ per dollar invested. Thus, high wage customers are willing to pay a smaller markup per dollar over the auction price. Then, the two cases to consider are

1. The former incumbent sets a price of $p_A + s_{rL}$ and high wage customers switch to the new auction winner in their second period, or
2. The former incumbent sells to all customers at a lower price of $p_A + s_{rH}$

For case 1), we then have[11]

$$V_A = \frac{1}{4}(\frac{1}{2} + 2I_H m_H)(p_A - c) + 2\delta V_L) + \frac{1}{2}\delta V_A \tag{27}$$

and

[11] As in previous sections, we don't differentiate between savings and wages. Allowing savings for middle-aged customers to be derived both from a wage and previous period returns will not change the results, so we abstract away from this issue.

$$V_L = \frac{1}{2}\frac{1}{2}(p_A + s_{rL} - c) + \delta V_A \tag{28}$$

The indifference condition for auction participants requires that

$$\frac{1}{2}(\frac{1}{2} + 2I_H m_H)(p_A - c) + 2\delta V_L) = \delta V_A \tag{29}$$

Putting Equations 27, 28, and 29 together, we have

$$p_A = c - s_{rL}\frac{\delta}{1 + 4m_H I_H + \delta}$$

and

$$p_L = c + s_{rL}\frac{1 + 4m_H I_H}{1 + 4m_H I_H + \delta}$$

The continuation values are then $V_A = 0$ and $V_L = s_{rL}\frac{1+4m_H I_H}{4(1+4m_H I_H+\delta)}$. Lifetime consumer surplus is $CS_I = I(r - p_A + \delta(r - p_L - \mathbb{I}_I(p_A - p_L + s_{rI}))$, where $\mathbb{I}_I$ is an indicator for switching. Then low wage consumers have lifetime consumer surplus of

$$CS_L = (r - c)(1 + \delta) - s_{rL}\delta\frac{4m_H I_H}{1 + 4m_H I_H + \delta} - \tag{30}$$

and high wage consumers have a lifetime consumer surplus of

$$CS_H = I(r - c)(1 + \delta) + s_{rL}\delta\frac{(1 + \delta)}{1 + 4m_H I_H + \delta} - \delta s_{rH} \tag{31}$$

Additionally, the total surplus in the market is $CS = m_L CS_L + m_H CS_H$. Defining the $CS^* = (r - c)(1 + \delta)(m_L I_L + m_H I_H)$, we have that the consumer surplus generated in Case 1) is

$$CS_1 = CS^* - \delta m_H I_H s_{rH} - s_{rL}\delta\frac{(1 - \delta)m_H I_H}{1 + 4m_H I_H + \delta} \tag{32}$$

We can interpret Equation 32 as follows: CS^* represents the optimal consumer surplus – consumers receive the equivalent of returns minus costs of provision for two periods, weighted by wage and mass of consumers. However, the actual consumer surplus with switching is not optimal. This is not because firms extract surplus for themselves–the present value of producer surplus in the bidding period is still zero. Instead, surplus is lowered in two ways:

1. Directly by switching costs since the high wage consumers switch in the second period, and

2. Due to a wedge between consumers and producers. Winning the auction means charging the auction price p_A for $\frac{1}{2} + 2m_H I_H$ worth of assets. This buys the winner the ability to charge p_L for $\frac{1}{2}$ worth of assets next period, discounted by δ. Then, for the Bertrand zero profit condition to hold, the ratio between the auction subsidy and the switching cost premium must be $\frac{p_A - c}{p_L - c} = \frac{\delta}{1+4m_H I_H}$. Implicit in this is that the second period subsidy to the high wage customers – a payoff discounted by δ for consumers – is undiscounted for bidders. In previous cases of the model, payoff timing was the same for consumers and bidders. Auction participants valued young payoffs at 1 and middle aged payoffs at δ, as did consumers themselves. Now, high wage consumers value the second period subsidy less than producers do, and its benefits are strictly less than the costs it imposes on non-switching customers. Specifically, in Equation 32, high wage customers get the auction subsidy twice, since they buy from the auction winner both times. Thus, they are effectively subsidized by the market, and, if switching costs for high wage customers are low enough, they may actually get more surplus than in a market with homogeneous customers. However, low wage customers must pay for this double subsidy, dramatically decreasing their benefit from the pension system, and because of asymmetric discounting this cost exceeds the high wage customers' benefit.

By contrast, Case 2) is largely identical to the homogeneous consumer case. The value functions are

$$V_A = \frac{1}{4}(\frac{1}{2} + I_H m_H)(p_A - c) + 2\delta V_L) + \frac{1}{2}\delta V_A \tag{33}$$

and

$$V_L = \frac{1}{2}(\frac{1}{2} + I_H m_H)(p_A + s_{rH} - c) + \delta V_A \tag{34}$$

The indifference condition for auction participants requires that

$$\frac{1}{2}(\frac{1}{2} + I_H m_H)(p_A - c) + 2\delta V_L = \delta V_A \tag{35}$$

Putting these conditions together, we have

$$p_A = c - s_{rL}\frac{\delta}{1+\delta} \tag{36}$$

and

$$p_L = c + s_{rL}\frac{1}{1+\delta} \tag{37}$$

The continuation values are then $V_A = 0$ and $V_L = s_{rH}\frac{1+2m_H I_H}{4(1+\delta)}$. The low wage consumers have lifetime consumer surplus of

$$CS_L = (r - c)(1 + \delta)$$

and high wage consumers have a lifetime consumer surplus of

$$CS_H = I(r - c)(1 + \delta)$$

which gives that the consumer surplus in Case 2) is

$$CS_2 = CS^*$$

Thus, the potential inefficiencies of a single price arise only when the firms choose to serve only the low wage customers in middle age. When will they choose to do this? We can check for deviations for Case 1) and Case 2). In Case 1), an incumbent may choose to deviate by lowering prices and serving all middle aged customers. They will chose to do this when $(\frac{1}{2}+I_H m_H)(p_A+s_{rH}-c) > \frac{1}{2}(p_A+s_{rL}-c)$, which obtains when $\frac{s_{rH}}{s_{rL}} > \frac{1+\delta+2I_H m_H(2+\delta)}{(1+2I_H m_H)(1+4m_H I_H+\delta)}$. That is, trading off lower price for higher quantity is more attractive when the price differential is smaller. Because auction prices differ between Cases 1) and 2), the cutoffs for deviations will also differ. Under Case 2), firms will deviate to serving only the high wage workers when $(\frac{1}{2}+I_H m_H)(p_A+s_{rH}-c) > 1/2(p_A+s_{rL}-c)$, or when $\frac{s_{rH}}{s_{rL}} > \frac{1+\delta}{1+2I_H m_H+\delta}$. Since $\frac{1+\delta}{1+2I_H m_H+\delta} > \frac{1+\delta+2I_H m_H(2+\delta)}{(1+2I_H m_H)(1+4m_H I_H+\delta)}$, we can find cases where a pure strategy equilibrium does not exist.

8.1 Heterogenous Customers with Endogenous Savings

Finally, we'll consider the case with heterogeneous customers where the firm extracts all returns from a customer's savings and pays consumers a fixed return on investment. In previous sections, we've elided this issue by treating savings as a non-durable good: each period, customers earn wages, invest, and consume wages for a payoff. In reality, customers save in both periods, and consumption of savings occurs in old age. These differences are unimportant if firms charge prices based on wages – the models are isomorphic so long as time discounting corresponds to the interest rate – but if firms charge prices based on total savings then prices in the first period will determine how much firms can extract in the second period, something not captured in the non-durable savings case. Thus, pricing over contributions and pricing over savings will not be precisely equivalent.

Generally, the results are the same as the non-durable savings model – firms subsidize in the first period and extract in the second period. The only salient difference occurs when firms price high wage customers out of the market. As before, this leads to inefficiency, but with endogenous savings the magnitude of that inefficiency is slightly lower for pricing over savings than for pricing over contributions because the subsidies are slightly smaller. That is, consumer surplus is slightly higher for pricing over savings. Because this is the only case of interest, we'll only consider Case 1) as defined in the previous section.

We'll now introduce an explicit model of savings. Defining output as $O = m_L w_L + m_H w_H$, the total output in for a given generation is O at the beginning of their first period. This output is invested at a proportional cost c, yielding a return of $r - c$ the next period. This version of the model is much more algebraically complex, so we'll make some simplifying assumptions: $m_H = \frac{1}{2}$ and $c = 0$. Then, in the second period, output entering the period is $O(1 + r)$ – consumers earn more, and the savings from last period gain a return. This output is again invested, yielding $O(1 + r)r$. This must all be discounted by δ^2, and we'll assume $\delta = \frac{1}{r}$ to avoid having two different rates of discounting, yielding $O\frac{1+r}{r}$. Switching costs will remain proportional to wages. Therefore, for pricing over savings, we have the following continuation values:

$$V_A = \frac{1}{4}(1/2 + (2 + f_A)\frac{w_H}{2})(r - f_A) + \frac{2V_L}{r}) + \frac{1}{2r}V_A \tag{38}$$

$$V_L = \frac{1}{2}\frac{1}{2}((1 + f_A)(r - f_A) + s_{rL}) + \frac{V_A}{r} \tag{39}$$

Then low wage consumers have lifetime consumer surplus of

$$CS_L^f = (1 + f_A)f_A - s_{rL} \tag{40}$$

and high wage consumers have a lifetime consumer surplus of

$$CS_H^f = w_H((1 + f_A)f_A - s_{rH}) \tag{41}$$

Pricing over contributions is equivalent to the model studied in Section 5, but with the new model of savings the expressions are slightly different. The continuation values are as follows:

$$V_A = \frac{1}{4}(\frac{1}{2} + w_H)p_A + 2V_L/r) + \frac{1}{2r}V_A \tag{42}$$

$$V_L = \frac{1}{2}\frac{1}{2}(p_A + s_{rL}) + \frac{V_A}{r} \tag{43}$$

Then, low wage consumers have lifetime consumer surplus of

$$CS_L^p = r(1 + r - p_A) - p_A - s_{rL} \tag{44}$$

and high wage consumers have a lifetime consumer surplus of

$$CS_H^p = w_H(r(1 + r - p_A) - p_A - s_{rH}) \tag{45}$$

Our goal is to determine which pricing scheme generates higher consumer surplus, which is proven in the proposition below.

Proposition 7. *The total surplus in the case for pricing over savings is higher than in the case for priving over contributions, i.e.*

$$CS^f \equiv CS_H^f + CS_L^f > CS^p \equiv CS_H^p + CS_L^p$$

Proof. Appendix. □

Relative to the non-durable savings model, the endogenous savings model puts more weight on older customers. In the second period, agents have greater savings, making the marginal cost of offering additional subsidies in the pricing over savings case greater since it decreases profits in the second period much faster. Therefore, subsidies are lower, and the time-inconsistency inefficiency identified in Section 5, which is proportional to the size of the subsidy, is attenuated. It is also worthwhile to note that even though $CS^f > CS^p$, it still features some inefficiency relative to price discrimination and the efficient level of consumer surplus, CS^*.[12]

9 Conclusion

In summary, this paper studies the private, defined contribution retirement system in Chile by explicitly modeling a repeated auction with effort in order to derive the surplus-maximizing, and hence efficiency-maximizing, pricing scheme. We first characterize the Nash equilibria for the first price auction with no effort. We then study different pricing schemes by explicitly modeling effort. We show that if effort is unobservable, firms will invest no effort in obtaining high returns for their portfolios. We consider three pricing schemes that can be characterized by their degree of variable pricing relative to fixed pricing – a scheme that involves only charging a fee over contributions (as is currently practice in the market), a scheme that involves only charging a price that varies with returns, and a two-part tariff scheme that involves a combination of the two previous schemes. We find that the two-part tariff scheme maximizes total economic surplus – that is, efficiency is maximized when firms compete in the auction on the fixed fee component, and then are incentivized to invest more effort in their portfolios in order to capture a higher revenue. It's important to note that the two-part tariff pricing scheme is not the unique pricing scheme that implements the first-best level of effort. Finally, we show that by modelling heterogeneity in income, allowing price discrimination can improve consumer surplus.

[12] In this setting, the efficient level of consumer surplus is $CS^* = (1+r)r\theta$ and $CS^f_\theta = \theta(f_A(1+f_A) - s_r)$. A bit of algebra reveals that $CS^* = (1+r)r\theta > CS^f_\theta = \theta(f_A(1+f_A) - s_r)$ if and only if $\frac{4s_r(rw+1)\left(rw\left(r^2+r+s_r-2\right)+3(r-1)r+s_r\right)}{(r-1)^2} + 8r\left(r^2w(2w+1)+3rw+r+1\right) > 0$, which must hold given that $r > 1$ and $min\{w, s_r\}$.

References

Aoyagi, M. (2003). Bid rotation and collusion in repeated auctions. *Journal of Economic Theory, 112(1), 79-105.*

Baily, M. N., & Kirkegaard, J. F. (2008). US pension reform: Lessons from other countries. *Columbia University Press.*

Basak, S., and Pavlova, A. (2013). Asset prices and institutional investors. *The American Economic Review*, 103(5), 1728-1758.

Bateman, H., Kingston, G., & Piggott, J. (2001). Forced saving: Mandating private retirement incomes. Cambridge University Press.

Berstein, S., Castañeda, P., Fajnzylber, E., and Reyes, G. (2009). Chile 2008: A Second-Generation Pension Reform. *Superintendencia de Pensiones de Chile.*

Castro, R. (2005). Efectos de largo plazo de la comisión fija en el sistema chileno de AFP. *Documento de trabajo n9, Superintendencia de Administradoras de Fondos de Pensiones*

Chomik, R and Piggott, J. (2015). Superannuation and retirement incomes. *JASSA*, (1), 43.

Creighton, A., and Piggott, J. (2006). The structure and performance of mandated pensions. *The Oxford Handbook of Pensions and Retirement Income, 13, 241.*

Dobronogov, A., & Murthi, M. (2005). Administrative fees and costs of mandatory private pensions in transition economies. Journal of Pension Economics & Finance, 4(1), 31-55.

Fischer, R., González, P., and Serra, P. (2006). Does competition in privatized social services work? The Chilean experience. *World Development, 34(4), 647–664.*

Hastings, J., Mitchell, O. S., and Chyn, E. (2011). Fees, Framing, and Financial Literacy in the Choice of Pension Manager. *Financial Literacy: Implications for Retirement Security and the Financial Marketplace, 101.*

Hastings, J. S., Hortaçsu, A., and Syverson, C. (2017). Sales Force and Competition in Financial Product Markets: The Case of Mexico's Social Security Privatization. *Econometrica, 85(6), 1723-1761.*

Heuser, A., Kwok, J., Snethlage, D., and Watts, D. (2015). Review of the KiwiSaver Fund Manager Market Dynamics and Allocation of Assets. *The NZ Treasury.*

Holzmann, R., Hinz, R. P., Von Gersdorff, H., Intermit, G., Impavido, G., Musalem, A. R., ... and Schwarz, A. (2005). Old-age income support in the twenty-first century: an international perspective on pension systems and reform. *Washington, DC: World Bank.*

Illanes, G. (2016). Switching Costs in Pension Plan Choice. *Working Paper.*

Koijen, R. S. (2014). The Cross-Section of Managerial Ability, Incentives, and Risk Preferences. The Journal of Finance, 69(3), 1051-1098.

Kurach, R., Kusmierczyk, P., and Papla, D. (2017). Can auctions help reduce mandatory pension fund fees? *Journal of Pension Economics & Finance.*

Luco, F. (2016). Switching Costs and Competition in Retirement Investment. *Working Paper.*

McAfee, R. P., and McMillan, J. (1992). Bidding rings. *The American Economic Review, 579-599.*

Mitchell, O., Todd, P., and Bravo, D., (2007). Learning from the Chilean Experience: The Determinants of Pension Switching. *Working paper, University of Chile, Department of Economics.*

Mesa-Lago, C. (2016). Suggestions for Pension Re-Reform in Peru. *Revista Apuntes*, 43(78), 41-60.

Modigliani, F., & Muralidhar, A. (2004). Rethinking pension reform. *Cambridge University Press*

Lasaga, M., & Pollner, J. (2003). Peru's Private Pension System: An Analysis of Its Evolution, Current Market Structure and Recommendations for Reform. *World Bank, Washington, DC.*

Orszag, P. R., and Stiglitz, J. E. (2001). Rethinking pension reform: Ten myths about social security systems, in R. Holzmann and J. Stiglitz (eds.). *New Ideas About Old Age Security: Toward Sustainable Pension Systems in the 21st Century. Washington, D.C.: World Bank Publications, 17-56.*

Raddatz, C., and Schmukler, S. L. (2013). Deconstructing herding: Evidence from pension fund investment behavior. *Journal of Financial Services Research, 43(1), 99-126.*

Skrzypacz, A., and Hopenhayn, H. (2004). Tacit collusion in repeated auctions. *Journal of Economic Theory, 114(1), 153-169.*

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

Tuesta, D. (2014). Factors behind the administrative fees of private pension systems: an international analysis. *Journal of Pension Economics & Finance, 13(1), 88-111.*

Warner, J. B., and Wu, J. S. (2011). Why do mutual fund advisory contracts change? Performance, growth, and spillover effects. *The Journal of Finance, 66(1), 271-306.*

Whitehouse, E. (2001). Administrative charges for funded pensions: comparison and assessment of 13 countries. *Private pension systems: Administrative costs and reforms*, 85-154.

Appendix

Proofs:

Proposition 7:

Proof. Because of the quadratic pricing over savings, direct comparison of the surpluses is difficult. Instead, we'll compare f_A to $f_A^{p\theta}$, the fixed fee that generates CS_θ^p. We'll show $f_A^{p\theta} > f_A$, meaning pricing over savings yields lower costs and higher consumer surplus. Direct inspection of Equations 40, 41, 44, and 45 shows that the ratio of surpluses only differs between H and L only by the value of $s_{r\theta}$, and we'll show the inequality holds independent of that value, so we'll condense notation to f_A^p, s_r, and CS^p without loss. We can recover f_A using the same derivation as before, yielding

$$f_A = \frac{\sqrt{4(rw+1)\left(r^2(2w+1)+r+s_r\right)+\left(r^2(-w)+2rw+1\right)^2}+r^2w-2rw-1}{2rw+2} \tag{46}$$

To find f_A^p, we equate CS^f to CS^p:

$$(1+f_A)f_A - s_r = r(1+r-p_A) - p_A - s_{rL} \tag{47}$$

Solving for f_A in terms of p_A and plugging in $p_A = -\frac{s_r}{2rw+r+1}$, we get

$$f_A^p = \frac{1}{2}\left(\sqrt{\frac{4(r+1)s_r}{2rw+r+1}+(2r+1)^2}-1\right) \tag{48}$$

It remains to show that Equation 46 is greater than 48. Subtracting 48 from 46, we have

$$-A+\sqrt{B}-\sqrt{C} \tag{49}$$

where $A=\frac{(r-1)rw}{2rw+2}$, $B=\frac{(r+1)s_r}{2rw+r+1}+\frac{1}{4}(2r+1)^2$, and $C=\frac{4(rw+1)(r(2rw+r+1)+s_r)+((r-2)rw-1)^2}{(2rw+2)^2}$. Direct inspection shows all three terms are non-negative. Then it suffices to show that $\sqrt{B}-\sqrt{C}>A$, which is true, since

$$B+C-2\sqrt{BC}>A^2 \implies$$

$$B+C-A^2>2\sqrt{BC} \implies$$

$$(B+C-A^2)^2>4BC \implies$$

$$A^4-2A^2(B+C)+(B-C)^2>0$$

In fact, $A^4-2A^2(B+C)+(B-C)^2=\frac{(r-1)^2r^2s_rw^2\left(r^2(2w+1)+r+s_r\right)}{(rw+1)^2(2rw+r+1)^2}>0$. Thus, $CS^f>CS^p$. □

A Figures

Figure A1: Evolution of market fee over wage (1988-2016)

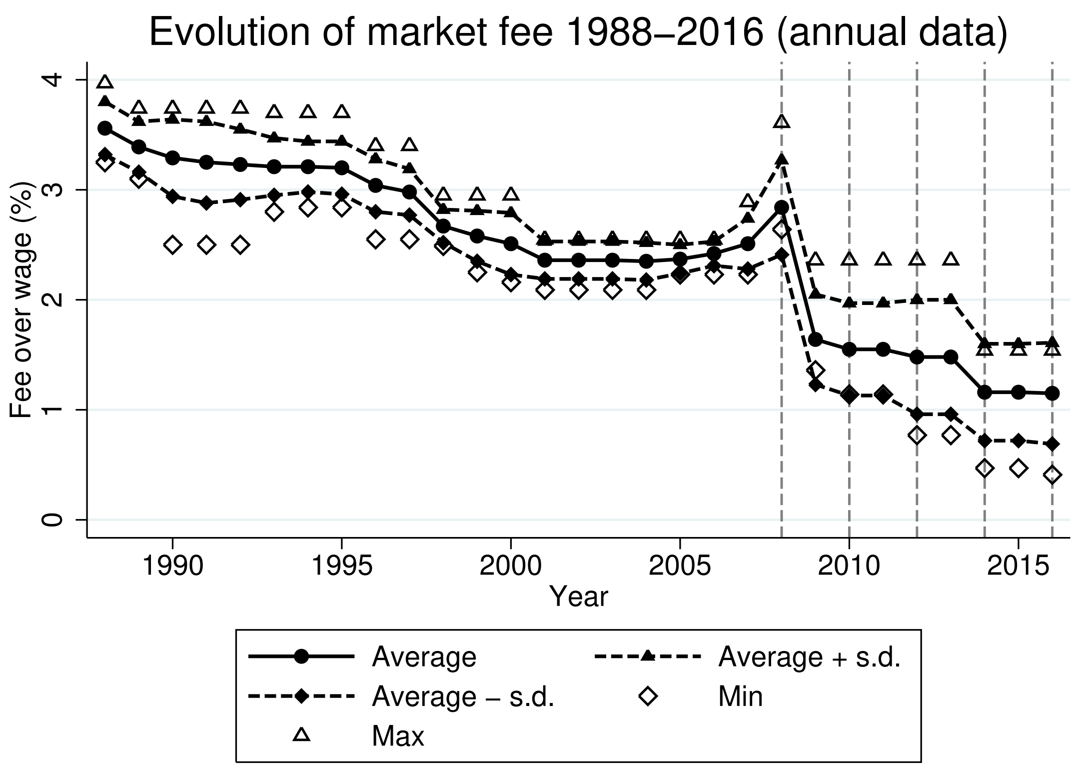


Figure A2: Fee over wage per firm (2010-2016)

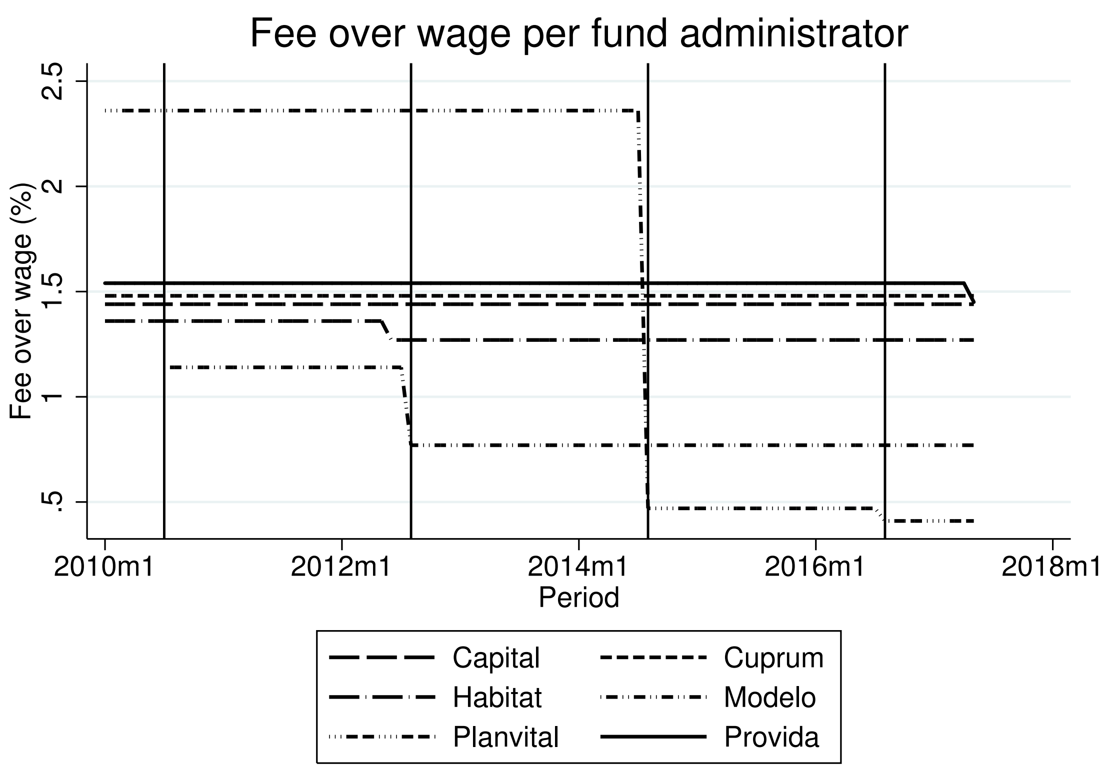
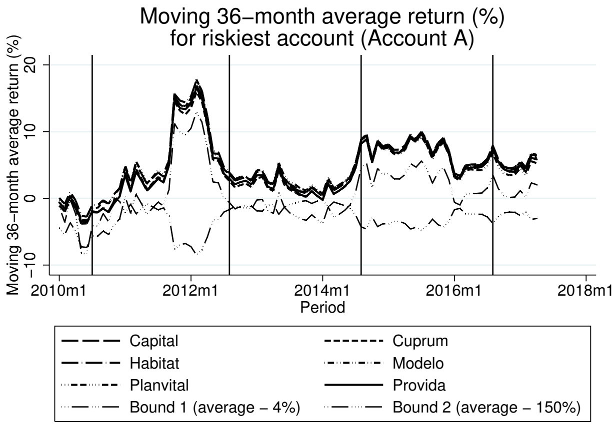
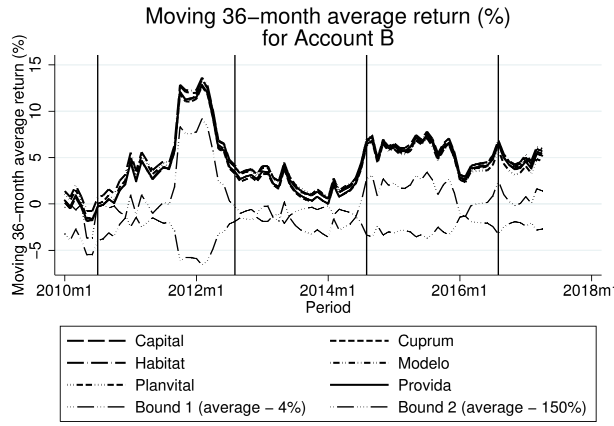


Figure A3: Return per account per firm and minimum bounds (36-months moving average)

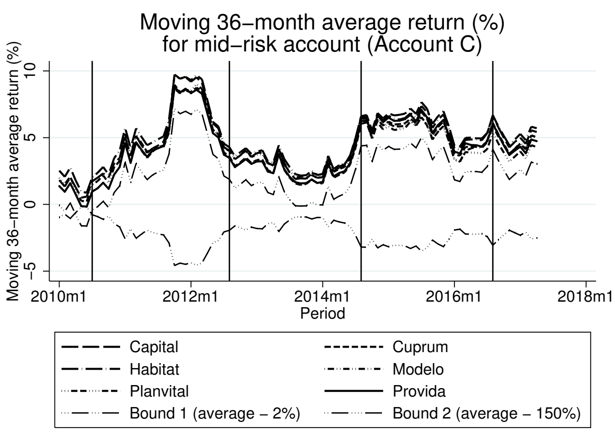
(a) Account A



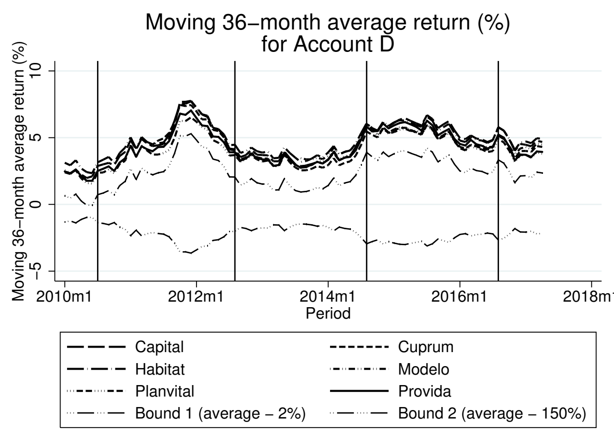
(b) Account B



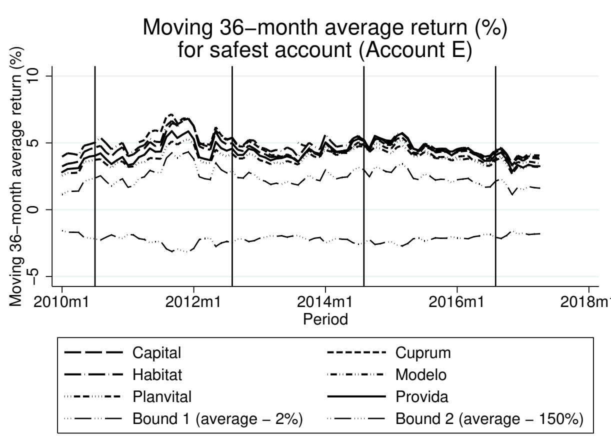
(c) Account C



(d) Account D



(e) Account E



B Tables

Table B1: Market equilibrium quantities and shares

Firm	Active Enrollees	Contributors			
		Total	Mandatory	Voluntary	
				Independent	Non-employed
Market equilibrium quantities (number of individuals)					
Capital	1,712,740	910,142	883,881	26,158	103
Cuprum	622,563	428,743	418,964	9,614	165
Habitat	2,038,030	1,142,507	1,120,887	21,299	321
Modelo	1,501,765	684,090	657,370	26,269	451
Planvital	1,161,224	548,591	530,003	17,091	1,497
Provida	3,142,115	1,570,264	1,528,289	41,954	21
Total	10,178,437	5.284,337	5,139,394	142,385	2,558
Market equilibrium shares (%)					
Capital	16.83	17.22	17.20	18.37	4.03
Cuprum	6.12	8.11	8.15	6.75	6.45
Habitat	20.02	21.62	21.81	14.96	12.55
Modelo	14.75	12.95	12.79	18.45	17.63
Planvital	11.41	10.38	10.31	12.00	58.52
Provida	30.87	29.72	29.74	29.47	0.82
Total	100.00	100.00	100.00	100.00	100.00

Note: (a) Based on data by the Superintendence of Pensions of Chile. (b) Information at December 31 of 2016.

Table B2: Market average monthly return per account (%)

Fund	Share on equities	Oct 2002–Dec 2016	Aug 2010–Dec 2016
Account A	40–80%	0.59	0.28
		(3.51)	(2.82)
		[-22.44–10.23]	[-6.48–6.65]
Account B	25-60%	0.48	0.26
		(2.54)	(2.05)
		[-15.16–6.84]	[-4.21–4.44]
Account C	15–40%	0.42	0.28
		(1.71)	(1.38)
		[-8.84–4.38]	[-2.55–3.36]
Account D	5–20%	0.38	0.31
		(1.09)	(0.87)
		[-4.85–3.13]	[-2.49–3.02]
Account E	0–5%	0.32	0.32
		(0.87)	(0.85)
		[-3.17–3.84]	[-2.54–3.84]

Note: (a) Average monthly real returns per account. Standard deviation in round parentheses, minimum and maximum returns in square parentheses. (b) Accounts B, C, D, and E are mandatory to offer, while Account A is voluntary to offer. (c) Own elaboration based on statistics from the Superintendencia de Pensiones.

Table B3: Real returns on investment Funds per firm (annualized return, %)

Firm	Account A			Account B		
	12 months	36 months	2002-2016	12 months	36 months	2002-2016
Capital	-1.10	3.69	6.08	0.94	3.74	5.24
Cuprum	-0.81	4.01	6.14	1.12	4.04	5.38
Habitat	-0.22	4.14	6.25	1.68	4.24	5.40
Modelo	0.26	3.89	–	2.14	3.90	–
Planvital	-1.62	3.10	5.73	-0.03	3.09	5.10
Provida	-1.59	3.51	6.04	0.35	3.60	4.97
Total	-0.87	3.85	6.13	1.07	3.88	5.23
Firm	**Account C**			**Account D**		
	12 months	36 months	2002-2016	12 months	36 months	2002-2016
Capital	1.63	4.02	4.67	2.69	3.91	4.50
Cuprum	1.95	4.53	5.19	2.82	4.45	4.80
Habitat	2.48	4.71	5.22	3.49	4.55	4.85
Modelo	2.76	4.03	–	3.76	4.13	–
Planvital	1.20	3.43	4.80	2.30	3.62	4.23
Provida	1.01	3.89	4.59	1.73	3.65	4.24
Total	1.74	4.24	4.90	2.60	4.07	4.54
Firm	**Account E**					
	12 months	36 months	2002-2016			
Capital	4.23	3.79	4.03			
Cuprum	3.95	3.85	3.98			
Habitat	4.32	4.00	4.13			
Modelo	4.43	3.55	–			
Planvital	3.55	3.18	3.29			
Provida	2.77	3.28	3.50			
Total	3.89	3.74	3.92			

Note: (a) Annualized rate of return. Constructed based on data by the Superintendence of Pensions of Chile. (b) Information at December 31 of 2016. (c) 2002-2016: September of 2002 to December of 2016.