

Switching Costs in Pension Plan Choice

Gastón Illanes*

April 14, 2017

PLEASE SEE LATEST VERSION AT:

<http://sites.northwestern.edu/gil208/>

Abstract

How well do market mechanisms for retirement savings function when there are switching costs? This work answers this question by estimating a dynamic demand model with switching costs for pension fund administrator choice in Chile's privatized pension market. This market exhibits significant price dispersion and very low switching rates, and switching costs are often mentioned as a likely driver of this outcome. If this is the case, then regulatory intervention to lower switching costs may increase welfare. This is not only important for the functioning of the Chilean pension market, but also more generally for other settings where governments mandate consumer participation and set the default as continuing in the same firm as last period. A key challenge in dynamic demand models is the fact that consumers form expectations about the future evolution of product characteristics and base their choices on them. Using a new methodology, based on a combination of revealed preference inequalities and latent variable integration, this work takes these expectations into account without having to model them explicitly, while using exclusion restrictions to separate switching costs from unobserved preference heterogeneity. I find

*I thank Victor Chernozhukov, Nancy Rose, and Paulo Somaini for their invaluable advice. I would also like to thank Nikhil Agarwal, Isaiah Andrews, Panle Jia Barwick, Glenn Ellison, Sara Fisher Ellison, Anna Mikusheva, James Poterba, Che-Lin Su and Michael Whinston for their helpful comments and suggestions. This paper has benefited from conversations with Carolina Alfero, Vivek Bhattacharya, Loreto Cox, Manasi Deshpande, Peter Diamond, Kathleen Easterbrook, Dong Jae Eun, Alfred Galichon, Jon Gruber, Sally Hudson, Augusto Iglesias, Bradley Larsen, Jing Li, Elena Manresa, Jorge Mastrangelo, Sarah Moshary, Yusuke Narita, Carlos Noton, Ana Maria Orellana, Manisha Padi, Parag Pathak, Maria Polyakova Breiter, Frank Schilbach, Bradley Shapiro, Ashish Shenoy, Daan Struyven, Daniel Waldinger and Melanie Wasserman; as well as participants in the MIT Industrial Organization seminar and field lunch, the Econometrics and Public Finance field lunches, and the Structural Econometrics Reading Group. I would also like to thank seminar participants at the Centro de Economía Aplicada - Universidad de Chile and at the Universidad de Santiago de Chile. I thank Augusto Iglesias and Solange Bernstein for their crucial help procuring data, as well as Olga Fuentes, Eugenio Salvo and Ximena Quintanilla at the Superintendencia de Pensiones. I gratefully acknowledge support from the Lynde and Harry Bradley Foundation, the George P. and Obie B. Shultz Dissertation Grant, and from Conicyt Chile.

evidence for a lower bound on switching costs of \$1,200 dollars, a number significantly higher than that found in previous work. Furthermore, I find evidence that consumers over-value returns differences across pension fund administrators relative to price differences. Observed prices are, on average, roughly twice as high as in a no switching cost counterfactual, suggesting that policy interventions to lower switching costs would be beneficial.

1 Introduction

Economists often argue that, absent market failures, competition drives markets to the efficient outcome. The driving force behind this result is the idea that consumers will re-optimize and switch suppliers when more attractive options appear, creating incentives for providers to set prices such that the efficient level of output is realized. However, one often-cited impediment to such effective competition is switching costs. While a first intuition may suggest that said costs would raise prices, theory indicates that this need not be the case. As argued by Farrell and Klemperer (2007) (among others), firms face a harvesting motive and an investment motive in these settings. The investment motive drives firms to lower prices in order to build a larger consumer base, while the harvesting motive leads firms to raise prices in order to extract rents from locked-in consumers. Which effect dominates, and thus the effect of switching costs on pricing and on efficiency, is an empirical matter. As a result, there is a thriving literature that estimates switching costs and tries to determine their market effects, in markets as varied as pension choice (Luco (2014)), drug insurance programs (Polyakova (2014)), health insurance markets (Handel (2013)), managed care plans (Nosal (2012)) and even in goods where logistical switching costs may not be present, but consumer inertia could still be relevant, such as orange juice and margarine (Dubé et al. (2010)). This work aims to determine the effect of switching costs on pricing in Chile's mandatory and privatized pension market, while proposing a new methodological strategy that is less restrictive than current methods and that may be useful in other settings.

The Chilean pension fund administration market is an interesting example of a setting where switching costs may be driving markets to an inefficient outcome. Chile established a market-based private account pension system in 1980, where formal sector workers are mandated to save 10% of their wages in a Pension Fund Administrator (PFA). Individuals can switch PFAs freely, but they cannot withdraw any money from their account until retirement. Commissions in this system are charged as a monthly percentage of wages. Loads, defined as the ratio between commission rates paid to the

PFA and the amount of money remitted to the system¹, ranged between 10% and 20% for the period between 2002 and 2011. Despite these significant differences in loads across companies, switching is low: on average, only 0.31% of customers change their PFA in a given month. From a theoretical perspective, both firm differentiation and switching costs could be drivers behind this price dispersion and persistence. This paper models this consumer inertia as a switching cost, estimates a dynamic demand model that incorporates both said cost and preference heterogeneity, and compares observed prices to a no switching cost counterfactual.

Characterizing consumer behavior in pension plan markets is important, not only because it gives insight for policy-makers in countries with pension systems that have a private component (or that are considering implementing one), but also because it sheds light on consumer behavior in other markets where participation is mandated and consumers may face similar informational, logistical or behavioral constraints that lead to demand inertia. This is particularly evident for health care markets. As a result, this paper contributes to both the pension choice literature (Duarte and Hastings (2012), Hastings et al. (2013), Hastings and Tejeda-Ashton (2008), Krasnokutskaya and Todd (2009), Luco (2014)) and to the broader literature on switching costs in mandated markets (Handel (2013), Luco (2014), Nosal (2012), Polyakova (2014)).

As the previous literature has recognized, there are several complications that make estimating switching costs difficult. First, as originally argued by Heckman (1981), separately identifying unobserved and persistent preference heterogeneity from state dependence is challenging. Second, the presence of switching costs implies that rational consumers should be forward looking, choosing goods by taking into account not only their current price and characteristics but also the expected evolution of these variables over time. And third, because firms also should be forward looking, maximizing the present discounted value of profits by counterbalancing the investment and the harvesting motives. To deal with these challenges, researchers have either imposed that consumers are myopic and only consider current period characteristics and prices when making their choices (Handel (2013), Luco (2014), Polyakova (2014)), or have been forced to explicitly model consumers' beliefs regarding the future evolution of characteristics, including prices (Nosal (2012)).

This paper builds on this literature by estimating switching costs in pension plan choice as well as their impact on pricing, while developing a methodology that takes into account the aforementioned challenges and that is broadly applicable to settings where researchers are interested in dynamic

¹ $Load = \frac{Commission\ Paid}{Commission\ Paid + Mandatory\ Contribution}$

demand models. This methodology relies on revealed preference inequalities to simplify the dynamic problem, following Bajari et al. (2007) and Pakes et al. (2014), while using latent variable integration methods (Galichon and Henry (2011); Schennach (2014)) to deal with selection in unobserved and possibly persistent preference heterogeneity. The model is identified through exclusion restrictions, which in this setting will take the form of an independence restriction between an unobservable and a set of instruments, changes in wages and lagged returns. Since commissions in this market are quoted as a percentage of income, changes in wages expose individuals to different prices, helping to trace out the trade-off between switching costs and prices. As for lagged returns, they both affect past choices and current account balances, helping identify the trade-off between switching costs and higher expected returns. Crucially, this estimation strategy neither assumes that consumers are myopic nor requires the econometrician to model beliefs about the evolution of future characteristics of the good. Furthermore, no assumptions beyond the aforementioned exclusion restriction and a conservative support restriction are needed regarding the distribution of unobserved preference heterogeneity. Relative to traditional demand estimation frameworks, such as BLP (Berry et al. (1995)) or maximum simulated likelihood, this model relaxes constraints that are imposed on the distribution of the unobservables at the cost of a more stringent exclusion restriction and set identification. Furthermore, relative to recent work in dynamic demand estimation (Handel (2013) and Gowrisankaran and Rysman (2012)), this method requires fewer constraints on the distribution of the unobservable and does not require a model for the evolution of consumer beliefs regarding future characteristics of goods, again at the expense of a more stringent exclusion restriction and set identification.

Using this methodology and a parametric utility model that incorporates both pricing and the impact that returns differences across firms will have on savings accounts at the time of retirement yields a switching cost estimate of at least \$1,200 dollars, which is in line with PDV differences in commissions paid across firms for reasonable discount rates. It would be impossible to recover such a high parameter estimate when assuming myopic consumers, however, as observed commission differences across firms in a single month are never more than roughly \$75. Parameter estimates also show that consumers behave as if they expect account balances to grow at a monthly rate between 0.9% and 1.3%, which compounds to yearly returns between 11.4% and 16.8%. As a comparison, yearly returns have averaged 8.73% since the inception of the system and 5.03% since September 2002², so we can conclude that participants in the system are choosing firms as if returns differences will have

²For a particular fund (C), the only existing fund between the creation of the system and 2002. The relationship between funds and PFAs will be explained in the following section.

a greater impact in their retirement account balance than what is actually the case. What explains this overvaluation? One possibility is that this result is driven by differences in saliency between commissions and returns, since the model is estimated using dollars of commissions as the numeraire. This would imply that individuals value an extra dollar of returns more heavily than a dollar of commissions saved. This behavior is in line with results from other papers looking at consumer choice that have also found differences in the value that consumers assign to money from different sources. Among others, Abaluck and Gruber (2011) find that participants in Medicare Part D value premiums an order of magnitude more than they value out of pocket expenditures, Chetty et al. (2009) find that individuals underreact to taxes that are not salient, and Ellison and Ellison (2009) find that shoppers of computer memory modules are more sensitive to differences in prices than to differences in taxes. The previous argument also has implications for the interpretation of the switching cost parameter, as the lower bound is \$1,200 dollars of commissions, not \$1,200 dollars in cash. That is, this parameter should not be interpreted as predicting that the mean participant in this system would not switch for \$1,200 dollars cash, but that they would not switch for savings of \$1,200 dollars of commissions in PDV terms.

Despite having obtained a high switching cost parameter estimate, one cannot determine the effect of switching costs on pricing without solving for the counterfactual equilibrium when there are no switching costs. This is because the effect of switching costs on prices depends on whether the investment motive or the harvesting motive dominates (Dubé et al. (2010), Cabral (2012))³. In this setting, prices drop by 46% on average in a no switching cost counterfactual simulation, with a range across firms between 33% and 75%. As a result, the market is in the range where switching costs raise prices, and any policy intervention to lower these costs will result in lower prices. Even after eliminating switching costs, the over-valuation of returns described in the previous paragraph also leads to higher prices, as counterfactual simulations where individuals compound balances using historical returns lead to a further price drop. This suggests that policy interventions to increase the saliency of commissions relative to returns realizations would help reduce prices. However, this effect is small relative to the effect of switching costs.

As for the welfare effects of switching costs, note that the mandatory nature of contributions in this market implies that for formal sector workers total market demand is perfectly inelastic, and prices are only a transfer between consumers and firms. As a result, switching costs do not have effects

³The prediction from this literature is that “high” switching costs raise prices, while “low” switching costs lower prices. However, what is a high switching cost and what is a low switching cost is market-specific and not determined by theory.

on welfare through quantity withholding, unless one specifies either an elasticity of contributions for informal sector workers or a different weight in the social welfare function for consumers and firms. Estimating this elasticity is left for future research, but provided it is non-zero, lowering switching costs would increase welfare.

The remainder of this paper is organized as follows. Section 2 gives an overview of Chile’s privatized pension system, gives details regarding the data that is being used and provides descriptive evidence. Section 3 specifies a dynamic discrete choice model, while Section 4 discusses alternative methodological approaches to estimation and proposes a combination of revealed preference inequalities and latent variable integration methods as a way to deal with some key methodological issues. Section 5 discusses implementation details of this procedure, Section 6 presents results, and Section 7 obtains predicted prices under a no switching cost counterfactual, and compares them to observed pricing. Section 8 concludes.

2 Data and Descriptive Evidence

2.1 Description of the System

Chile has a regulated private and mandatory pension system, where formal sector workers must choose one of several pension fund administrators (PFAs). They contribute 10% of their monthly income to a PFA account, up to a cap that is linked to the consumer price index and that in July 2013 was at about \$320 dollars. Informal sector workers contribute voluntarily, and often do not contribute at all⁴. On top of this contribution, consumers pay a percentage of their income as commission. That is, a worker that chooses a PFA that charges a 3% variable commission rate has 13% of her income automatically transferred to that company each month. Commission levels have never been regulated, and charging any other type of commission is not allowed⁵, so for example commissions linked to the amount of savings in the account or commissions for switching providers are not observed.

In order to map commission rates to a more familiar scale, one can define the load charged by each PFA as the ratio between the amount paid in commissions and the total amount transferred to the system:

$$Load = \frac{Monthly\ Commission\ Paid}{Monthly\ Commission\ Paid + Monthly\ Mandatory\ Contribution} \quad (2.1)$$

⁴In August 2012, there were roughly 10 million accounts in the system, and 96.1% of them corresponded to formal sector workers.

⁵Since September 2008. Before that, PFAs were allowed to charge a monthly fixed fee as well.

Figure 1 plots loads from 2002 through 2012, while Figure 2 plots commission rates for the same period. The main stylized fact to be obtained from these plots is that commission rates are dispersed, with loads ranging from 10% to 20% of the total amount contributed to the system. Are they also high, relative to international standards? This is a difficult question, as the structure of fees varies significantly across countries. The OECD (OECD (2005, 2011)) and Tuesta (2013) perform an international comparison of fees by taking the ratio of the aggregate flow of commissions paid in a particular year to the total assets administered (See Figure 3). Under this metric, the Chilean system's fees are around the median internationally. Appendix A performs the exercise of mapping commission rates in the Chilean system to commissions quoted as expense ratios. The main conclusion of this exercise is that commissions in this system seem high relative to what one could purchase in the US, particularly considering that the inexpensive options that are available in the US were not available in Chile during this period. However, it is hard to draw a conclusive result regarding the competitiveness of commissions in Chile from these international comparisons, for several reasons. First, because any comparison between systems with different commission structures is sensitive to assumptions about discount rates, contribution rates, and wage profiles, among other variables. Second, because markets in other countries need not be competitive to begin with. And third, because our view regarding prices in this market will depend heavily on what factors are driving optimal pricing. That is, we are likely to view pricing differences in a different light if they are due to quality differences across firms than if they are mostly driven by switching costs.

Ultimately, determining whether prices in Chile are high relative to international standards is not the focus of this paper. Instead, the goal is to determine what is driving price levels and price dispersion across firms in this market, as well as the low observed switching rates. Many factors could play a role: quality differences, firm differentiation, and demand inertia could all be affecting optimal pricing. This paper builds a demand model that takes these factors into account, and then uses the results from that model to quantify their effects on pricing. Before discussing the model, however, it will be important to discuss some regulatory features of the system and some reduced form evidence that will underpin the modelling assumptions.

To begin, there are several possible dimensions for differentiation: deposit safety (theft risk), investment ability, quality, advertising, sales-forces, etc. Some of these dimensions are not relevant, while others could be significant. There is no vertical differentiation in theft risk, as there are several regulations in place to protect workers' funds. The most relevant is that PFAs must keep savings

separate from their own cash flows, and cannot use them to lend to themselves or as capital. This implies that even if an PFA goes bankrupt, the value of the pension funds that it manages should be unaffected. In fact, during the 1982-83 crisis, several PFAs went bankrupt, and workers' savings were not affected (Diamond and Valdés (1993)). Furthermore, PFAs can merge, be bought and sold, and there can be entry and exit into the market, and during all these transactions individuals' savings must be untouched. Figure 4 plots the monthly number of firms from January 1988 to December 2011, showing that the market was relatively dynamic during the early 90's, and has stabilized since. Overall, there should be no differential risk across companies of the money being misappropriated.

Second, PFAs are not free to invest as they wish or to offer products of their choosing, and because of this they are constrained in their ability to differentiate on returns. Instead, they must offer five different funds to invest in, each with caps for exposure to different asset classes. Funds are labeled from A to E, with A having the largest proportion of variable income securities and E the smallest. Table 1 shows a summary of investment rules for each fund (OECD (2012)). Workers are free to choose between funds⁶, and if they do not choose a fund, they are placed by default in one according to their age. Appendix B discusses other regulations that make differentiation on returns difficult, as well as evidence showing that returns across firms are very highly correlated and that price differences across firms are not correlated with returns differences. This Appendix also presents evidence from competition amongst PFAs during the retirement phase of an individual's lifetime. Overall, the evidence presented in this appendix confirms the notion that pricing differences across companies are not due to differences in returns.

However, returns are not the only quality dimension over which firms can differentiate. Service quality, number and location of branches, salesforces, advertisements, among others, are all possible sources of differentiation across firms. The effect of these variables on consumer choice will vary across individuals and time, and will be subsumed into the unobservable in the structural model.

Another factor that could explain observed pricing differences is switching costs. Finding conclusive evidence of switching costs in the reduced form is difficult, as low switching rates can also be explained by persistent preference heterogeneity. Ultimately, we need a model to separate switching costs from persistent preference heterogeneity. Nevertheless, some suggestive reduced form evidence of switching costs can be found by looking at mergers. As shown in Figure 4, the PFA market had over 20 active firms during the 90's. Some of these firms ceased to exist by merging with other companies,

⁶Except for males over 56 years old, and females over 51, who cannot choose Fund A

creating new PFAs, while others were absorbed by existing PFAs, maintaining the marketing and market positioning of the absorbing company. Whenever a merger takes place and one of the merging parties ceases to exist, while the other continues with the same branding, I will call the disappearing firm the “absorbed” firm and the continuing firm the “absorbing” firm. There are 6 mergers that fit this criterion⁷, taking place between 1993 and 1999. After the merger, workers who had chosen the absorbed firm become customers of the absorbing firm, unless they decide to switch. If there were no switching costs, one would expect that after some time passes the distribution of choices of the absorbed firm’s customers would look like the distribution of choices in the population. Column 1 of Table 5 compares the probability that an individual chooses the absorbing firm in January 2007 if they were a customer of an absorbed firm and if they weren’t. It shows that being a customer of an absorbed firm increases the probability of choosing the absorbing firm roughly twenty years later by 26.7 percentage points. This is not conclusive evidence for the presence of switching costs, as merger partners aren’t chosen at random, and one could argue that absorbed firms are chosen as merger partners because of their customers’ strong preference for the absorbing firm. Column 2 of Table 5 studies this issue by comparing the probability that a customer of the absorbed firm chooses the absorbing firm in January 2007 with the probability that a customer of the absorbing firm at the time of merger chooses the absorbing firm some twenty years later. If absorbed firms were selected due to their customers’ preference for the absorbing firm, one would expect these probabilities to be similar, but this is not the case: customers of the absorbed firm are significantly less likely to choose the absorbing firm in January 2007. Further suggestive evidence of the presence of switching costs in this market, stemming from a reform introduced in 2010 to periodically auction off the right to serve first-time workers for their first two years in the system, is presented in Appendix C.

To summarize the arguments from this subsection, prices in this market are disperse, and while both switching costs and firm differentiation could explain this phenomenon, it is unlikely that differentiation by itself can create some of the observed features of the data. The next logical step is to build a model of product choice in this market that considers switching costs as well as vertical differentiation and to estimate it. Before doing so, the next subsection will introduce the data used for estimation and present further descriptive evidence that will guide the later modelling choices.

⁷Planvital and Invierta (1993), Provida and El Libertador (1995), Santa Maria and Banguardia (1995), Planvital and Concordia (1996), Provida and Union (1998), and Provida and Proteccion (1999). Planvital and Magister (2004) also meets the criterion, but due to data limitations I cannot identify Magister’s customers, so this merger is not considered in the analysis.

2.2 Data and Descriptive Evidence

This paper works with an administrative database made available by the Chilean Pension Superintendency⁸, which consists of a representative sample of accounts in the Chilean pension system since the system was created. The data is structured as a monthly panel, with data from 1980 to 2011. For observations before January 2007, this database contains information on age, gender, date person joined the pension system, retirement date, monthly wages, and monthly commission rates paid. For observations after January 2007, this database also has information on which PFA an individual chooses each month, and how their account balance is distributed between the 5 funds offered by each PFA. To my knowledge, this is the first paper that uses administrative data on which company each individual chooses to study switching costs in this market, as previous papers (Luco (2014)) have had to infer individuals' choices. The availability of actual choice information is the reason why this paper focuses on the sample period between January 2007 and December 2011.

Restricting the sample to individuals who are younger than 65 and who have not retired, the data used in this analysis consists of between 19,855 and 20,367 individuals per month, with the fluctuation being explained by retirement and entry of new customers. The total number of observations in the sample is 1,169,489. Table 2 shows some key indicators for the sample. An important feature of the data is the large proportion of observations who have zero wages (45.9%). This is because individuals who have been formal sector workers in the past but are currently unemployed or informally employed, and as a result need not be contributing to the system, will still be sampled into the data, as it is a representative sample of all accounts. This explains the discrepancy between mean wages (\$492) and mean non-zero wages (\$908) in Table 2. Over the entire sample, 21.5% of individuals always have zero wages. These individuals do not have a price motive for switching companies, as prices are a percentage of income, but could potentially switch due to the effect of other product characteristics. The average account balance is \$14,211 dollars, and this rises to \$20,238 if employed. Table 3 shows percentiles of the account balance distribution, both for the entire sample and for individuals who always have zero wages in the sample. Note that individuals in the latter group have significantly lower savings, but also that some have managed to save a significant amount. Therefore, if they have differential beliefs about returns across firms they could still find it optimal to switch.

Table 4 shows summary statistics on switching behavior. The monthly switching rate is 0.31%, while the yearly switching rate is 3.18%. For the sample period 89.3% of individuals never switch, while

⁸Historia Previsional Administrativa de Afiliados Activos, Pensionados y Fallecidos, available at <http://www.spensiones.cl/portal/informes/581/w3-propertyvalue-6480.html>

7.3% switch once. Switching more than once is an uncommon occurrence, only happening 1.69% of the time. Switchers have significantly higher wages than average (\$1,237 versus \$492), are younger (38.2 versus 41.5), and have more money saved (\$26,006 versus \$14,211). To further explore the correlation between switching and observable characteristics, Figure 6 present predicted switching probabilities obtained from the following linear probability model:

$$1 [Switch_{it}] = b_{age}(age_{it}) + b_{wage}(wage_{it}) + b_{bal}(balance_{it}) + \epsilon_{it} \quad (2.2)$$

where $b_{var}(\cdot)$ is a cubic b-spline of the relevant variable. That is, each plot presents the fits of the model for a particular variable after partialling out the others. Note that each plot is scaled so that its minimum switching probability is zero. The first plot shows that wages are positively correlated with switching for all but high amounts, at which point they become negatively correlated. There are several factors contributing to these results. Since price is a percentage of wages, individuals with higher wages save more when switching to a lower priced firm, and should be expected to switch more. At the same time, salesforces are more likely to target high earners, inducing them to switch. Finally, there is a strong correlation between wages and education, and so one would expect that as wages increase the probability that an individual is more informed about the working of the system rises. The fact that switching is negatively correlated with wages for high earners could be due to a higher opportunity cost of time, to more effective customer retention strategies by firms, or to stronger firm differentiation for these groups due to targeted advertising. The second plot shows that account balances are positively correlated with switching for all but high balance individuals. Since balances represent the accumulation of wages over time, similar explanations apply. Finally, the third plot shows that older individuals are less likely to switch, even after controlling for wage and account balance. This is an interesting result, as the theoretical prediction is ambiguous. On the one hand, younger individuals have more time to “pay off” the investment of switching, and as a result should be more willing to switch. On the other, older individuals should be more willing to switch to firms that are cheap today but that are expected to be more expensive in the future, as by the time the firm raises prices they will have retired. The fact that older individuals have lower switching rates suggests that the former effect dominates. If there were no switching costs, however, it would be difficult to understand why younger individuals are more likely to switch even after controlling for the effects of wages and balances.

The fact that prices are quoted as a percentage of income implies that wage increases exacerbate

pricing differences across firms, creating incentives to switch. As long as this effect exists, and wage changes are orthogonal to unobserved preference heterogeneity in this market, they could be provide a valid instrument to identify the model. Table 5 shows that in fact wage changes affect switching behavior, even after controlling for individual fixed effects. The column marked “Relative Switching Probability” presents the relative switching probability between an individual whose wage doesn’t change and one with wage increases of \$1,000, \$100, and \$10 dollars, and a wage drop of \$100 dollars. For example, a person whose wage increased by \$100 is 28% more likely to switch than one whose wage didn’t change. As for the exclusion restriction, there are three possible channels through which one could find dependence between wage changes and unobserved preference heterogeneity. First, it is possible that some firms are better at identifying individuals whose wages are increasing, and target their sales and marketing efforts accordingly. Second, it is possible that individuals who are more likely to have wage changes differentially prefer certain PFAs, even after controlling for pricing. And third, individual’s preferences could change when their wages change, particularly after coming back to work. The first concern is dampened by the high frequency of the data, as it seems improbable that salesforces would be able to immediately identify and target individuals whose wage increased in a particular month. The second and third channels are potentially more problematic, but I am unaware of any evidence supporting them. Thus, the assumption that changes in wages are independent of the unobservable will be the first exclusion restriction used to identify the model.

The second exclusion restriction will be between the unobservable and lagged returns. Lagged returns affect previous choices, and therefore the identity of that the firm that an individual is locked in to in the current period. They also shift current account balances. At first glance, the fact that returns differences across firms are not systematic would favor the assumption that they are independent of unobserved preference heterogeneity. However, if sales or marketing efforts respond to spurious differences in returns across firms, then this assumption would be violated. The question then is what is the speed of accommodation: do firms respond to monthly differences in returns realizations with changes in their sales and marketing efforts? If firms take more than a month to respond, the assumption is valid. To test this hypothesis, I collected quarterly data on number of salesforce workers hired by each PFA from the Chilean Pension Superintendency’s website, and purchased monthly advertising expenditures estimates from Megatime, a company that tracks advertising on different platforms. Table 6 presents results of the regressions of number of salesforce workers hired and advertising expenditures on last month’s and semester’s returns, by fund. Note that for all funds lagged monthly

returns have a statistically insignificant relationship with the number of salesforce workers hired and advertising expenditures. The magnitudes are also small. For example, the largest effects imply that a 1% increase in monthly returns leads to six more salesforce workers and \$128 more dollars spent in advertising. Semester returns have a statistically significant relationship with the number of salesforce workers hired, but with an unexpected sign for all funds but E, and again these are also economically small. As for advertising expenditures, this variable also has an statistically insignificant relationship with semester returns for all funds but E, and as before the effects are small. Overall, these results validate the assumption that differences in last months' returns do not affect this month's advertising or salesforce efforts.

The preceding discussion has served to show some salient features of switching behavior in this market, as well as to introduce the exclusion restrictions that will identify the model. The data also allows us to get a sense of the money that individuals are leaving on the table from not switching to the cheapest firm. To do so, Table 7 reports the results of comparing the PDV of commissions paid during each worker's lifetime to what they would have paid if they had instead chosen the cheapest firm each period⁹. The row titled "PDV Commission Savings" reports the results obtained using a 5% yearly discount rate. Since many observations have zero wages for every month in the sample period, naturally the amount these individuals would save is zero, explaining the first percentiles. The 50th percentile of the PDV savings is \$176, and the 90th percentile is \$1,475. That is, 10 percent of the sample would save at least \$1,475 in PDV terms from switching to the cheapest path. Of course, this calculation ignores differences across firms in other dimensions, particularly returns. However, the second row of this table shows that adding returns to the calculation¹⁰ actually exacerbates the difference between the observed choice path and the cheapest choice path, as during this period the cheapest firms happened to have higher returns realizations. Panel B repeats this analysis, but dropping from the sample individuals who always have a zero wage during the sample period. In this case, the 50th percentile of PDV savings is \$373 when ignoring returns differences, and \$630 when including them. These numbers give a sense of the magnitude of switching costs one should expect to find in this market. However, these calculations ignore the possibility that individuals value firms for reasons other than returns and prices. In order to incorporate this effect, we need to build a model. That is

⁹With a few assumptions made to get around data limitations. Since choice data is only available for the period between January 2007 and December 2011, wages and commissions paid after this date are unknown. This calculation assumes that wages are fixed at their December 2011 level for each individual, and uses observed commission rates for the years 2012 to 2014. After that, commissions are assumed to be fixed at their December 2014 levels. For the cheapest firm this is not an unreasonable assumption, as they are the auction winners and are locked in to their December 2014 price until mid 2016.

¹⁰By incorporating the difference in account balances in December 2011 or at retirement, whichever comes earliest.

the subject of the following section.

3 Model

This section introduces a model of pension fund administration choice. As justified in the previous section, this model will consider pricing, switching costs, returns differences across firms, and other sources of differentiation across companies that are unobservable to the econometrician but observed by consumers. It will model individuals as making choices across PFAs each month by maximizing the sum of the PDV of flow utilities over time and the expected PDV of their retirement balance. As a convenient simplification, flow utilities will be assumed to be linear and additively separable with different time periods and with expected retirement balances. This imposes risk neutrality, which is reasonable, as this model is not about the allocation of money across funds, but rather fund administrator choice, and the latter is by far the greatest source of risk in this setting. Furthermore, I will assume that there is no sensitivity of contributions to differences in returns realizations across firms. Since mandatory contributions are a fixed percentage of wages, this assumption could only be violated if voluntary contributions vary with past returns realizations or if individuals' labor force participation decision changes with these realizations. The latter is unreasonable, as realized returns differences across firms are often small, while the former requires more scrutiny. There are two ways this assumption could be violated: first, if voluntary savers change their savings behavior, and second, if individuals who are voluntarily contributing more than the mandatory amount change their voluntary savings with these realizations. As of January 2012, voluntary savers account for roughly 0.02% of active accounts and 1% of all accounts. As their impact on the market is negligible, they are dropped from the analysis. Voluntary contributions beyond the mandatory amount are possible, but PFAs compete in this market with a large number of other companies, and comprehensive data on this sector is unavailable. Furthermore, even if one chooses a PFA for the administration of the voluntary savings account, it need not be the same PFA that is administering the mandatory account. As a result, the voluntary market will be assumed separate from the mandatory, and will be ignored in this analysis. Finally, in 2010 a reform was enacted that auctions off the right to serve first-time workers for their first two years in the system. These workers cannot choose PFAs and are not considered in the analysis.

Suppose that each period individuals in the system are choosing which PFA to keep their investments in. Let i denote the individual, t the time period, and j the firm. Let d_{it} denote i 's PFA choice

in period t , B_{it} their PFA savings balance, X_{ijt} a set of individual-firm-time level observable characteristics, ϵ_{ijt} a set of individual-firm-time level characteristics that is observed by individuals but not by the econometrician, and Ω_{it} i 's information set in period t . Finally, let T_i denote i 's retirement date, which is assumed to be fixed at 60 for females and 65 for males, and β denote the discount rate. Assume the timing of the game is as follows:

1. Period t begins. Individuals observe $(X_{ijt}, \epsilon_{ijt})$, the choice set \mathcal{J}_t , and their account balance B_{it} . They update their beliefs, forming Ω_{it} .
2. Individuals choose a firm for period t , perceive the flow utility benefits and costs according to $u(d_{it}, d_{i,t-1}, X_{ijt}, \epsilon_{ijt})$, and contribute $c_{i,t+1}$.
3. Returns are realized, period t ends.
4. Period $t+1$ begins. Individuals observe $(X_{ij,t+1}, \epsilon_{ij,t+1})$, the choice set \mathcal{J}_{t+1} , and their account balance $B_{i,t+1} = B_{it}(1 + r_{d_{it}}) + c_{i,t+1}$. They update their beliefs and form $\Omega_{i,t+1}$.

At stage 1, each person is solving the following problem:

$$\begin{aligned}
E[V_{it}(d_{i,t-1}, B_{it}) | \Omega_{it}] &\equiv \max_{\{j_\tau \in \mathcal{J}_\tau\}_{\tau=t}^{T_i}} \left\{ u(j_t, d_{i,t-1}, X_{ijt}, \epsilon_{ijt}) \right. \\
&\quad \left. + \sum_{\tau=t+1}^{T_i} \beta^{\tau-t} \cdot E[u(j_\tau, j_{\tau-1}, X_{ij\tau}, \epsilon_{ij\tau}) | \Omega_{it}] \right. \\
&\quad \left. + \beta^{T_i-t} \cdot E[B_{iT_i}(\{j_\tau\}_{\tau=t}^{T_i}, B_{it}) | \Omega_{it}] \right\} \\
&= \max_{j_t} u(j_t, d_{i,t-1}, X_{ijt}, \epsilon_{ijt}) + \beta \cdot E[V_{i,t+1}(j_t, B_{i,t+1}) | \Omega_{it}]
\end{aligned} \tag{3.1}$$

where $E[V_{it}(d_{i,t-1}, B_{it}) | \Omega_{it}]$ denotes individual i 's expected value of being locked in to firm $d_{i,t-1}$ and of having an account balance of B_{it} in period t , given their information set Ω_{it} , and $E[B_{iT_i}(\{j_\tau\}_{\tau=t}^{T_i}, B_{it}) | \Omega_{it}]$ is a function that takes as inputs a stream of choices between the current period and the retirement period, and an account balance, and returns the expected account balance at the time of retirement given an information set.

For the sake of expositional clarity, this notation ignores the possibility that consumers have uncertainty over the realizations of the characteristics of the goods that enter flow utility in the current period. Under the aforementioned timing assumptions, there is uncertainty when choosing a firm regarding returns realizations for the period, but in the utility parameterization introduced later returns do not enter flow utility, so it is not necessary to introduce the extra notation. It is important to note, however, that this approach can handle settings where there is uncertainty over the realizations of characteristics that enter flow utility in the current period.

Papers that assume myopic consumers (Handel (2013); Luco (2014); Polyakova (2014)) set the discount rate $\beta = 0$, and typically use a distributional assumption on ϵ_{ijt} to construct choice probabilities. For example, assuming $\epsilon_{ijt} \equiv \xi_{jt} + \eta_{ijt}$ and imposing that η_{ijt} is a logit error leads to a standard multinomial logit model. One can parameterize utility and estimate such a model via Maximum Simulated Likelihood or Simulated Method of Moments. Some papers incorporate a control function approach to deal with price endogeneity, while others assume that fixed effects and interactions of random coefficients with individual characteristics are enough to deal with the correlation between unobserved preference heterogeneity η_{ijt} and endogenous characteristics, such as prices. The identification assumption in these cases is that unobserved preference heterogeneity η_{ijt} is uncorrelated with any characteristic after controlling for the included variables. Note that any forward-looking behavior will be loaded onto unobserved preference heterogeneity in such a model, so to obtain consistent estimates one also requires that consumers are in fact myopic, or that the error when assuming myopia is uncorrelated with the included characteristics. If instead they are forward looking, parameter estimates will be biased, and the direction of the bias is unclear. Relative to myopic consumers, forward-looking individuals may be more or less price sensitive, depending on their beliefs regarding the evolution of prices over time. If they anticipate that firms that charge lower prices today are more likely to charge higher prices tomorrow, they will be less price sensitive, while if they anticipate that the price difference will be persistent they will be more price sensitive.

Alternatively, some papers work with similar distributional assumptions, but incorporate forward looking consumers. Nosal (2012) presents a direct application of forward-looking behavior to demand estimation with switching costs. A closely related literature is that of experience goods (Akerberg (2003); Erdem and Keane (1996) among others), as uncertainty over characteristics creates a switching cost between goods that have been tried before and have been found better than the expected value of goods that have not been tried before. More generally, Gowrisankaran and Rysman (2012) and Hendel and Nevo (2013) have estimated dynamic demand models in other settings. The key challenge with dynamic demand estimation is the fact that the econometrician must make assumptions regarding individuals' expectations of the evolution of future characteristics, including prices. If these assumptions are incorrect, parameter estimates will also be biased.

A recent literature (Pakes et al. (2014); Morales et al. (2014)) suggests the use of moment inequalities as a way to incorporate forward looking behavior without explicitly modeling beliefs, and this will be the strategy used in this paper. The argument is based on using revealed preferences and one-period

deviations from observed behavior to control for dynamic considerations, as argued in Bajari et al. (2007). Assume that we observe an individual choosing the same firm for two consecutive periods ($d_{i,t}^* = d_{i,t-1}^* = j$). Then:

$$\begin{aligned} E[V_{it}(j, B_{it}) | \Omega_{it}] &= u(j, j, X_{ijt}, \epsilon_{ijt}) + \beta \cdot E[V_{i,t+1}(j, B_{i,t+1}) | \Omega_{it}] \\ &\geq u(j', j, X_{ij't}, \epsilon_{ij't}) + \beta \cdot E[V_{i,t+1}(j', B_{i,t+1}) | \Omega_{it}] \end{aligned} \quad (3.2)$$

for any alternative $j' \in \mathcal{J}_\tau$. Rearranging,

$$\begin{aligned} &u(j, j, X_{ijt}, \epsilon_{ijt}) - u(j', j, X_{ij't}, \epsilon_{ij't}) \\ &\geq \beta \cdot (E[V_{i,t+1}(j', B_{i,t+1}) | \Omega_{it}] - E[V_{i,t+1}(j, B_{i,t+1}) | \Omega_{it}]) \end{aligned} \quad (3.3)$$

One can use revealed preference arguments to find a lower bound of the terms in the right hand side of equation 3.3. To do so, it is useful to define $\{j_\tau^*\}_{\tau=t+1}^{T_i}$ as the sequence of choices that attains $E[V_{i,t+1}(j, B_{i,t+1}) | \Omega_{it}]$:

$$\begin{aligned} \{j_\tau^*\}_{\tau=t+1}^{T_i} \equiv & \arg \max_{\{j_\tau \in \mathcal{J}_\tau\}_{\tau=t+1}^{T_i}} \left\{ E[u(j_{t+1}, j, X_{ij,t+1}, \epsilon_{ij,t+1}) | \Omega_{it}] \right. \\ & + \sum_{\tau=t+2}^{T_i} \beta^{\tau-t-1} \cdot E[u(j_\tau, j_{\tau-1}, X_{ij\tau}, \epsilon_{ij\tau}) | \Omega_{it}] \\ & \left. + \beta^{T_i-t-1} \cdot E[B_{iT_i}(\{j_\tau\}_{\tau=t+1}^T, B_{it}) | \Omega_{it}] \right\} \end{aligned} \quad (3.4)$$

Note that this is an object that is not observed in the data: it is the stochastic sequence of choices that attains the maximum expected value of the problem from $t+1$ onwards, conditional on the information set in period t , and given that j is chosen in t and that the account balance remains B_{it} . It is useful, however, because one can derive that:

$$\begin{aligned} E[V_{i,t+1}(j', B_{i,t+1}) | \Omega_{it}] &\geq E[u(j_{t+1}^*, j', X_{ij,t+1}, \epsilon_{ij,t+1}) | \Omega_{it}] \\ &+ \sum_{\tau=t+2}^{T_i} \beta^{\tau-t-1} \cdot E[u(j_\tau^*, j_{\tau-1}^*, X_{ij\tau}, \epsilon_{ij\tau}) | \Omega_{it}] \\ &+ \beta^{T_i-t-1} \cdot E[B_{iT_i}(\{j_\tau^*\}_{\tau=t+1}^{T_i}, B_{i,t+1}(j')) | \Omega_{it}] \end{aligned} \quad (3.5)$$

This weak inequality holds because $\{j_\tau^*\}_{\tau=t+1}^{T_i}$ is a possible choice sequence, but not necessarily the utility maximizing sequence. Note also that by definition $\{j_\tau^*\}_{\tau=t+1}^{T_i}$ attains $E[V_{i,t+1}(j, B_{it}) | \Omega_{it}]$. This

allows us to bound the difference in continuation values in the right hand side of equation 3.3 by:

$$\begin{aligned}
& E [V_{i,t+1} (j', B_{i,t+1}) | \Omega_{it}] - E [V_{i,t+1} (j, B_{i,t+1}) | \Omega_{it}] \geq \\
& E [u (j_{t+1}^*, j', X_{ij,t+1}, \epsilon_{ij,t+1}) - u (j_{t+1}^*, j, X_{ij,t+1}, \epsilon_{ij,t+1}) | \Omega_{it}] \\
& + \beta^{T_i-t-1} \cdot E \left[B_{iT} \left(\{j_\tau^*\}_{\tau=t+1}^{T_i}, B_{i,t+1} (j') \right) - B_{iT} \left(\{j_\tau^*\}_{\tau=t+1}^{T_i}, B_{i,t+1} (j) \right) | \Omega_{it} \right]
\end{aligned} \tag{3.6}$$

This condition implies that the expected difference in continuation values at time t must be weakly greater than the expected difference obtained when assuming that, regardless of the firm picked in t , from period $t+1$ onwards the individual will follow the sequence of choices that maximizes expected continuation value when picking firm j in period t . This difference simplifies to a flow utility difference in period $t+1$ and an account balance difference at the time of retirement. All flow utilities from $t+2$ to retirement cancel out, as both choices and lagged choices are the same from that point onwards.

Parameterizing utility and the balance generating function allows for a further simplification of this expression. Note that the first term in the right-hand side of 3.6 is the difference in flow utilities from choosing j_{t+1}^* when one is locked in to j' and when one is locked in to j . Assuming $u (j_t, d_{t-1}, X_{ijt}, \epsilon_{ijt}) = -\delta \cdot 1 [j_t \neq d_{t-1}] + \alpha w_{it} p_{jt} + \epsilon_{ijt}$ and $\delta \geq 0$, where p_{jt} is firm j 's price in period t and w_{it} is individual i 's salary the same period, this difference is no lower than $-\delta$, and we have that:

$$\begin{aligned}
& E [V_{i,t+1} (j', B_{i,t+1}) | \Omega_{it}] - E [V_{i,t+1} (j, B_{i,t+1}) | \Omega_{it}] \geq \\
& -\delta + \beta^{T_i-t-1} \cdot E \left[B_{iT} \left(\{j_\tau^*\}_{\tau=t+1}^{T_i}, B_{i,t+1} (j') \right) - B_{iT} \left(\{j_\tau^*\}_{\tau=t+1}^{T_i}, B_{i,t+1} (j) \right) | \Omega_{it} \right]
\end{aligned} \tag{3.7}$$

We can simplify this expression further by defining the expected balance generating function as:

$$E \left[B_{iT} \left(\{j_\tau^*\}_{\tau=t+1}^{T_i}, B_{i,t+1} \right) | \Omega_{it} \right] \equiv E \left[\prod_{\tau=t+1}^T (1 + r_{j\tau}) B_{i,t+1} + \sum_{k=t+2}^T \prod_{\tau=k}^T (1 + r_{j\tau}) c_{ik} | \Omega_{it} \right] \tag{3.8}$$

which is simply the compounding of the initial balance plus the compounding of future contributions c_{ik} . If the future path of contributions does not change when facing an account balance of $B_{i,t+1} (j') = B_{it} (1 + r_{j't}) + c_{i,t+1}$ or $B_{i,t+1} (j) = B_{it} (1 + r_{jt}) + c_{i,t+1}$, we have that:

$$\begin{aligned}
& E [V_{i,t+1} (j', B_{i,t+1}) | \Omega_{it}] - E [V_{i,t+1} (j, B_{i,t+1}) | \Omega_{it}] \\
& \geq -\delta + \beta^{T_i-t-1} \cdot E \left[(B_{i,t+1} (j') - B_{i,t+1} (j)) \prod_{\tau=t+1}^{T_i} (1 + r_{j^*\tau}) | \Omega_{it} \right] \\
& = -\delta + \beta^{T_i-t-1} \cdot B_{it} \cdot E \left[(r_{j't} - r_{jt}) \prod_{\tau=t+1}^{T_i} (1 + r_{j^*\tau}) | \Omega_{it} \right]
\end{aligned} \tag{3.9}$$

That is, the expected disutility of being attached to a different firm is bounded below by δ , the cost

of switching back, and the expected difference in returns. We can substitute this lower bound back into equation 3.3, which gives us that:

$$\begin{aligned} & \alpha w_{it} (p_{jt} - p_{j't}) + \delta (1 + \beta) \\ & + \beta^{T_i-t} \cdot B_{it} \cdot E \left[(r_{jt} - r_{j't}) \prod_{\tau=t+1}^{T_i} (1 + r_{j^*\tau}) \mid \Omega_{it} \right] \\ & \geq \epsilon_{ij't} - \epsilon_{ijt} \end{aligned} \quad (3.10)$$

Recall that this derivation assumed that the individual made the same choice in two consecutive periods, $d_{i,t}^* = d_{i,t-1}^* = j$. This is not the only possible case: individuals can switch, retire, or can be making a choice for the first time. The term accompanying the switching cost parameter will vary depending on each case, and can be expressed as a function of the current choice d_{it} , the past choice $d_{i,t-1}$, the alternative used to construct the inequality d'_{it} , and whether the individual is retiring or is joining the system. See the Appendix for derivations of the equivalent inequalities for these cases. Taking all cases into consideration, we can write the general inequality:

$$\begin{aligned} & \alpha w_{it} (p_{jt} - p_{j't}) + \delta \cdot \Delta (d_{it}, d_{i,t-1}, d'_{it}, \beta, T_i) \\ & + \beta^{T_i-t} \cdot B_{it} \cdot E \left[(r_{jt} - r_{j't}) \prod_{\tau=t+1}^{T_i} (1 + r_{j^*\tau}) \mid \Omega_{it} \right] \\ & \geq \epsilon_{ij't} - \epsilon_{ijt} \end{aligned} \quad (3.11)$$

where:

$$\Delta (d_{it}, d_{i,t-1}, d'_{it}, \beta) = \begin{cases} 1 + \beta & \text{if } d_{it} = d_{i,t-1} \\ -1 + \beta & \text{if } d_{it} \neq d_{i,t-1} \text{ and } d'_{it} = d_{i,t-1} \\ \beta & \text{if } (d_{it} \neq d_{i,t-1} \text{ and } d'_{it} \neq d_{i,t-1}) \text{ or } d_{i,t-1} = \emptyset \\ 1 & \text{if } t+1 = T_i \text{ and } d_{it} = d_{i,t-1} \\ -1 & \text{if } t+1 = T_i \text{ and } d_{it} \neq d_{i,t-1} \text{ and } d'_{it} = d_{i,t-1} \\ 0 & \text{if } t+1 = T_i \text{ and } d_{it} \neq d_{i,t-1} \text{ and } d'_{it} \neq d_{i,t-1} \end{cases} \quad (3.12)$$

and $d_{i,t-1} = \emptyset$ denotes a newcomer.

Note that under this model, the econometrician doesn't observe individuals' expected return differences across firms or their beliefs about how balances will compound over time under the choice path $\{j_\tau^*\}_{\tau=t+1}^{T_i}$. Before taking 3.11 to the data, one needs to take a stand on dealing with ex-

pected future returns $E \left[(r_{jt} - r_{j't}) \prod_{\tau=t+1}^{T_i} (1 + r_{j^*\tau}) | \Omega_{it} \right]$. This work replaces this expression with $(r_{jt} - r_{j't}) (1 + \omega)^{T_i - t - 1}$, where ω is a parameter to be estimated representing the mean per-period expected return. Replacing, the revealed preference inequality in equation 3.11 becomes:

$$\begin{aligned} & \alpha w_{it} (p_{jt} - p_{j't}) + \delta \cdot \Delta (d_{it}, d_{i,t-1}, d'_{it}, \beta) \\ & + \beta^{T_i - t} \cdot B_{it} \cdot (r_{jt} - r_{j't}) \cdot (1 + \omega)^{T_i - t - 1} \\ & \geq \epsilon_{ij't} - \epsilon_{ijt} + \eta_{ij't} - \eta_{ijt} \end{aligned} \quad (3.13)$$

where

$$\eta_{ij't} = \beta^{T_i - t} \cdot B_{it} \cdot \left(E \left[r_{jt} \cdot \prod_{\tau=t+1}^{T_i} (1 + r_{j^*\tau}) | \Omega_{it} \right] - r_{jt} \cdot (1 + \omega)^{T_i - t - 1} \right) \quad (3.14)$$

This approach implies parameterizing the beliefs regarding future returns under path $\{j^*_\tau\}_{\tau=t+1}^{T_i}$ as $(1 + \omega)^{T_i - t - 1}$, that is, that expected returns in the future are constant over time and across individuals. This allows us to recover the average expected return across the population for the time period under study. Doing so introduces an additional unobservable, $\eta_{ij't} - \eta_{ijt}$, that is composed of both the specification error from parameterizing beliefs regarding future returns and the expectational error from replacing the expectation of returns with its realization. To identify the model, an exclusion restriction between this unobservable and instruments must hold. The following section describes this issue in more detail. One could have a more complex parameterization of these beliefs, with variation across both time and observable characteristics of individuals. However, this would greatly increase the computational expense of estimation. As a first approximation, understanding average beliefs across the population in the time period in question seems interesting in and of itself, and incorporating greater complexity in modelling is left for future research.

Other simple parameterizations of $E \left[(r_{jt} - r_{j't}) \prod_{\tau=t+1}^{T_i} (1 + r_{j^*\tau}) | \Omega_{it} \right]$ are certainly possible. For example, one could impose that the relevant discount rate β is equal to $\frac{1}{1+r_{j^*\tau}}$, so that returns compounding drops out. Such a model would predict that conditional on account balance the switching probability is constant across individuals with different ages, whereas the posited model predicts that conditional on account balance younger individuals will be more likely to switch, a feature of the data. Another possibility would be to replace $E[(r_{jt} - r_{j't}) | \Omega_{it}]$ with a parametric expectation formation model that takes into account previous returns realizations. One could think of a two-step procedure, where first an expectation formation model is estimated, and then the fits of that model are substi-

tuted into $E[(r_{jt} - r_{j't}) | \Omega_{it}]$. The main argument against implementing this parameterization is the computational complexity involved with nesting the estimation of such a model into the confidence intervals of the latent variable integration procedure that will be used for estimation.

The end result of this derivation is a revealed preference inequality that is a function of current period differences in pricing and switching costs across firms, as well as differences in the expected present discounted value of retirement savings at the time of retirement. The following section discusses different alternatives for taking this model to the data, as well as the assumptions needed to make these estimation strategies valid.

4 Estimation Strategies

This section discusses alternative strategies for taking equation 3.13 to the data. The key difference between these alternative strategies will be how to deal with the unobservable components on the right-hand side of said equation. To begin, if one is willing to assume that across the population $E[\epsilon_{ij't} - \epsilon_{ijt} + \eta_{ij't} - \eta_{ijt}] = 0$, then straightforward application of traditional moment inequality estimators will suffice to recover the parameters of interest. However, this assumption is problematic for the unobserved preference difference $\epsilon_{ij't} - \epsilon_{ijt}$, as it implies that there is no selection in any relevant characteristic of the product that is omitted from the utility specification by the econometrician. If any characteristics are omitted, such as advertising or product differentiation, one would expect selection to occur along this unobservable dimension of firm quality, such that $E[\epsilon_{ij't} - \epsilon_{ijt}] < 0$. As a result, we need a strategy to deal with this selection.

An alternative estimation strategy could be to apply an instrumental variable and continue using traditional moment inequality methods. For notational simplicity, let $\zeta_{ij't} - \zeta_{ijt} \equiv \epsilon_{ij't} - \epsilon_{ijt} + \eta_{ij't} - \eta_{ijt}$. In the moment inequality framework, an instrumental variable z_{ijt} has to satisfy three conditions:

1. No Sign Changes:

$$z_{ijt} > 0 \forall i, j, t.$$

2. First Stage:

$$E \left[z_{ijt} \left(\alpha w_{it} (p_{jt} - p_{j't}) + \delta \cdot \Delta(d_{it}, d_{i,t-1}, d'_{it}, \beta) + \beta^{T_i-t} \cdot B_{it} \cdot (r_{jt} - r_{j't}) \cdot (1 + \omega)^{T-t-1} \right) \right] \neq 0$$

3. Exclusion Restriction:

$$E[z_{ijt} (\zeta_{ij't} - \zeta_{ijt})] \geq 0$$

The first condition requires that the instrument always has the same sign, so that multiplying both

sides of the moment inequality by the instrument does not flip the inequality for some observations but not for others. The second condition is a standard requirement, that the instrument has a first stage. Finally, the exclusion restriction in this case is an inequality, as if the interaction between the instrument and the unobservables has the correct sign, one can set:

$$\begin{aligned} E \left[z_{ijt} \left(\alpha w_{it} (p_{jt} - p_{j't}) + \delta \cdot \Delta (d_{it}, d_{i,t-1}, d'_{it}, \beta) + \beta^{T_i-t} \cdot B_{it} \cdot (r_{jt} - r_{j't}) \cdot (1 + \omega)^{T-t-1} \right) \right] \\ \geq E [z_{ijt} (\zeta_{ij't} - \zeta_{ijt})] \geq 0 \end{aligned} \quad (4.1)$$

and simply work with the moment:

$$E \left[z_{ijt} \left(\alpha w_{it} (p_{jt} - p_{j't}) + \delta \cdot \Delta (d_{it}, d_{i,t-1}, d'_{it}, \beta) + \beta^{T_i-t} \cdot B_{it} \cdot (r_{jt} - r_{j't}) \cdot (1 + \omega)^{T-t-1} \right) \right] \geq 0 \quad (4.2)$$

Unfortunately, instrumenting in this fashion does not take care of the selection problem induced by unobserved preference heterogeneity. To see this, assume that we have an instrument such that $cov(z_{ijt}, \epsilon_{ijt} | d_{it}^* = j) = 0$. Since,

$$E [z_{ijt} \epsilon_{ijt} | d_{it}^* = j] = cov(z_{ijt}, \epsilon_{ijt} | d_{it}^* = j) + E [z_{ijt} | d_{it}^* = j] \cdot E [\epsilon_{ijt} | d_{it}^* = j] \quad (4.3)$$

the fact that $E [z_{ijt}] > 0$ (due to the no sign changes condition) and $E [\epsilon_{ijt}] > 0$ (due to selection) implies that $E [z_{ijt} \epsilon_{ijt} | d_{it}^* = j] > 0$, and thus that the exclusion restriction is violated. As a result, direct application of instrumental variables is not an option in settings where selection on unobservables is relevant.

Ho and Pakes (2014) and Pakes et al. (2014) propose a “matched pairs strategy” to get around this issue. The idea is to divide the sample into observable characteristic bins, and to sum equation 3.13 for individuals who are in the same bin but who make different choices. Under this strategy, equation 3.13 becomes:

$$\begin{aligned} & \alpha (w_{it} - w_{i't}) (p_{jt} - p_{j't}) + \\ & + \delta \cdot (\Delta (d_{it}, d_{i,t-1}, d'_{it}, \beta) - \Delta (d_{i't}, d_{i',t-1}, d'_{i't}, \beta)) \\ & + \beta^{T_i-t} \left[B_{it} - \beta^{T_{i'}-T_i} \cdot (1 + \omega)^{T_{i'}-T_i} \cdot B_{i't} \right] (r_{jt} - r_{j't}) (1 + \omega)^{T_i-t-1} \\ & \geq \zeta_{ij't} - \zeta_{i'j't} - \zeta_{ijt} + \zeta_{i'jt} \end{aligned} \quad (4.4)$$

If $E[\zeta_{ij't} - \zeta_{i'j't} - \zeta_{ijt} + \zeta_{i'jt}] \geq 0$, or $E[z_{ii'jj't}(\zeta_{ij't} - \zeta_{i'j't} - \zeta_{ijt} + \zeta_{i'jt})] \geq 0^{11}$, then this strategy will appropriately control for selection bias. In general, this will not be the case, as one would expect $E[\epsilon_{ijt}] > E[\epsilon_{i'jt}]$ and $E[\epsilon_{i'j't}] \geq E[\epsilon_{ij't}]$, as individual i chooses firm j and individual i' chooses j' . Formally, this strategy only controls for selection bias in unobserved preference heterogeneity if, within bins of observables, unobserved preference heterogeneity is identical. In practice, however, if preference heterogeneity does not vary significantly within bins of observables, this can be a reasonable assumption.

Other solutions to this issue are discussed in Dickstein and Morales (2013). For binary choice settings, the authors argue that either parameterizing the distribution of the unobservable or normalizing the unobservable can solve the problem. The latter solution is not available in multiple choice settings, while the former implies solving for $E[\epsilon_{ijt}|d_{it}^* = j]$, which requires solving a dynamic choice problem. Since this is precisely the difficulty that moment inequalities was trying to avoid, in this setting these proposed solutions do not give us significant traction.

This work aims to build on this literature by applying latent variable integration methods (Galichon and Henry (2011); Schennach (2014)) to solve the problem posed by unobserved preference heterogeneity. The idea behind these methods is to pick a point in parameter space and to find the distribution of the unobservables conditional on the observables that minimizes a test statistic for the null hypothesis that said point is in the identified set. If this “most adverse distribution” allows us to reject the null, then the point is rejected. To introduce this estimator formally, some notation is needed. Following the notation in Schennach (2014), assume we have a moment $g(Z, U, \theta)$ that is a function of observable characteristics Z , unobservable characteristics U , and parameters θ . Denote the support of the unobservables by \mathcal{U} , and let $\mathcal{P}_{\mathcal{U}|\mathcal{Z}}$ denote the set of all regular conditional probability measures supported on \mathcal{U} (or any of its measurable subsets) given events that are measurable subsets of \mathcal{Z} . Let the marginal distribution of Z be supported on some set \mathcal{Z} , let the distribution of U conditional on $Z = z$ be supported on or inside the set \mathcal{U} for an $z \in \mathcal{Z}$. Let $\pi \in \mathcal{P}_{\mathcal{Z}}$ denote the probability measure of the observable variables, with π_0 denoting the true probability measure of the observables.

One can then write the identified set as:

$$\Theta_0 = \left\{ \theta \in \Theta : \inf_{\mu \in \mathcal{P}_{\mathcal{U}|\mathcal{Z}}} \|E_{\mu \times \pi_0}[g(U, Z, \theta)]\| = 0 \right\}$$

¹¹Denoting the instrument by $z_{ii'jj't}$ to show that it can be a function of both individuals' characteristics and of both firms' characteristics.

Note that determining whether a given θ is in the identified set requires searching over the space of conditional distributions of the unobservable $\mathcal{P}_{\mathcal{U}|Z}$ for a distribution that minimizes the value of the moment. Although this is clearly an infeasible problem, Schennach (2014) shows that it is isomorphic to a parametric problem that can actually be solved. Formally, Theorem 1 in Schennach (2014) states that for any $\theta \in \Theta$ and $\pi \in \mathcal{P}_Z$,

$$\inf_{\mu \in \mathcal{P}_{\mathcal{U}|Z}} \|E_{\mu \times \pi_0} [g(U, Z, \theta)]\| = 0$$

if and only if

$$\inf_{\gamma \in \mathbb{R}^{d_g}} \|E_{\pi} [\tilde{g}(Z, \theta, \gamma)]\| = 0$$

where

$$\tilde{g}(Z, \theta, \gamma) \equiv \frac{\int g(u, z, \theta) \exp(\gamma' g(u, z, \theta)) d\rho(u|z; \theta)}{\int \exp(\gamma' g(u, z, \theta)) d\rho(u|z; \theta)} \quad (4.5)$$

That is, the problem of searching over the space of conditional probability distributions of the unobservable given the observables can be replaced by a parametric problem of finding the parameters γ that minimize a weighted average of the moments under a distribution of the unobservables that belongs to a specific exponential family. Crucially, this exponential family has the “most adverse” property: it can span the range of values of the expectation of the moments generated if one were searching over the whole space of conditional probability distributions. The dimensionality of γ , g , is simply the number of moments. Note that the integrals required to calculate this weighted average are taken with respect to a distribution of the unobservable $\rho(u|z; \theta)$. The choice of $\rho(u|z; \theta)$ has no impact on the statistical properties of the estimator if it meets the following two conditions:

1. $\text{supp } \rho(\cdot|z; \theta) = \mathcal{U} \forall z \in \mathcal{Z}$.
2. $E_{\pi} [\ln E_{\rho(\cdot|z; \theta)} \exp(\gamma' g(U, X, \theta) | Z)]$ exists and is twice differentiable in γ for all $\gamma \in \mathbb{R}^{d_g}$.

The first condition states that the support of the distribution of the unobservable from which we will sample is equal to the support of the true unobservable, while the second imposes that a moment generating function-like quantity exists. Schennach (2014) shows how one can always construct a $\rho(u|z; \theta)$ that satisfies these conditions, and as a result this feature of the estimator is not restrictive.

Applying this estimator to the aforementioned dynamic discrete choice setting requires specifying

the moments and the unobservables we will be working with. Chesher et al. (2013) show that a multinomial discrete choice model is identified using a revealed preference moment and an independence restriction between an instrument and the unobservable. Recall that we had derived the inequality:

$$\begin{aligned} & \alpha w_{it} (p_{jt} - p_{j't}) + \delta \cdot \Delta (d_{it}, d_{i,t-1}, d'_{it}, \beta) \\ & + \beta^{T_i-t} \cdot B_{it} \cdot (r_{jt} - r_{j't}) \cdot (1 + \omega)^{T_i-t-1} \\ & \geq \zeta_{ij't} - \zeta_{ijt} \end{aligned} \quad (4.6)$$

Since this inequality must hold for all $j' \in \mathcal{J}$, we have the following revealed preference moment:

$$E \left[\sum_{j \in \mathcal{J}} 1 [d_{it}^* = j] \cdot \left[\sum_{j' \in \mathcal{J} - \{j\}} 1 [\alpha \cdot w_{it} (p_{jt} - p_{j't}) + \delta \cdot \Delta (d_{it}, d_{i,t-1}, d'_{it}, \beta) + \beta^{T_i-t} \cdot B_{it} \cdot (r_{jt} - r_{j't}) \cdot (1 + \omega)^{T_i-t-1} < \zeta_{ij't} - \zeta_{ijt}] \right] \right] = 0 \quad (4.7)$$

Formally implementing an independence restriction in this setting would require an infinite set of inequalities as additional moments, as the support of the unobservable is continuous. While there are feasible ways to implement such a restriction (Chernozhukov et al. (2013); Andrews and Shi (2014)), this is a difficult problem. As an alternative, Schennach (2014) suggests introducing a series of interactions of higher order moments of the instrument and the unobservable. This will lead to a larger confidence set than if the full independence restriction were implemented, but in practice this difference is likely to be negligible. Therefore, the model will also include the following restrictions¹²:

$$\begin{aligned} E [z_{ij't} \zeta_{ijt}] &= 0 \\ E [z_{ij't} \zeta_{ijt}^2] &= 0 \end{aligned} \quad \forall j, j' \in \mathcal{J} \quad (4.8)$$

That is, whenever instruments vary at the firm level, the model will impose restrictions across firms as well as within firms. Note that these restrictions are implemented for all firms in the data, not just the chosen firm. The reason for doing so is that some of the observables may be correlated, through selection, with ζ_{ijt} conditional on choosing j . As a result, any instrument with a first stage will also be correlated with ζ_{ijt} with this conditioning. However, over the population there is no selection on ζ_{ijt} , and as a result regular instrument validity arguments apply.

Recall that ζ_{ijt} is composed of two terms: unobserved preference heterogeneity ϵ_{ijt} , and the combination of specification and updating error η_{ijt}^2 . The instruments that will identify the model are lagged

¹²Note that for the second set of moments to make sense, the instrument must be de-meanned, as otherwise these moments cannot hold.

returns and changes in wages. The exclusion restriction is that each of these instruments is independent of the sum of these unobservables. Lagged returns $r_{j',t-1}$ and unobserved preference heterogeneity ϵ_{ijt} are independent if past returns only affect quality beliefs through future returns expectations. The main argument against this assumption is that past returns affect advertising or salesforce efforts. However, as shown in the previous section, at this frequency advertising and salesforce efforts are not significantly responding to returns differences. Lagged returns and η_{ijt} will be independent under a rational expectations model, as all past returns are in the information set. Furthermore, this instrument is relevant if previous returns change past choices, as it will shift the identity of the firm an individual is locked in to, or if previous returns are correlated with current returns. As for changes in wages, the exclusion restriction with respect to unobserved preference heterogeneity implies that changes in wages cannot change preferences, or lead to differential exposure to advertising and salesforces in the immediate month the wage change takes place. Furthermore, it also requires that individuals who are more likely to suffer wage changes have no differential unobserved preference for firms. The exclusion restriction regarding the wage change and η_{ijt} will also be met under a rational expectations model, for the same reasons as lagged returns. This instrument is relevant, as individuals whose wage changes face different prices.

As argued by Heckman (1981), a key challenge in identifying state dependence (switching costs) is being able to separate the effect of persistent preference heterogeneity. In this model, that challenge is tackled by the exclusion restrictions. Serial correlation of unobserved preference heterogeneity term ϵ_{ijt} doesn't invalidate the instruments in this case, as the moments are not conditional on current or previous choices. To gain intuition on identification of the switching cost parameter, focus on a simpler model, with two firms (j and j') and no returns differences across firms. Define $Y_{it} = \begin{cases} 1 & \text{if switch} \\ 0 & \text{otherwise} \end{cases}$. Let $X_{ijj't} \equiv w_{it}(p_{jt} - p_{j't})$, $\epsilon_{ijj't} \equiv \epsilon_{ijt} - \epsilon_{ij't}$, and assume $\epsilon \sim F$. Then:

$$\begin{aligned}
Y_{it} = 0 | d_{i,t-1} = j &\Rightarrow -X_{ijj't} + \delta(1 + \beta) \geq -\epsilon_{ijj't} \\
Y_{it} = 0 | d_{i,t-1} = j' &\Rightarrow X_{ijj't} + \delta(1 + \beta) \geq \epsilon_{ijj't} \\
Y_{it} = 1 | d_{i,t-1} = j &\Rightarrow -X_{ijj't} + \delta(1 - \beta) \leq -\epsilon_{ijj't} \\
Y_{it} = 1 | d_{i,t-1} = j' &\Rightarrow X_{ijj't} + \delta(1 - \beta) \leq \epsilon_{ijj't}
\end{aligned} \tag{4.9}$$

Dropping unnecessary subscripts, these conditions generate the following bounds for the distribution of the unobservable conditional on the state (the previous choice) and the value of the instrument

(See Appendix F for derivations).

$$\begin{aligned} \Pr[X - \delta(1 + \beta) < x, Y = 1|j, z] &\leq F(x|j, z) \leq 1 - \Pr[X - \delta(1 - \beta) > x, Y = 0|j, z] \\ \Pr[X + \delta(1 - \beta) < x, Y = 0|j', z] &\leq F(x|j', z) \leq 1 - \Pr[X + \delta(1 + \beta) > x, Y = 1|j', z] \end{aligned} \quad (4.10)$$

Combining inequalities across states and imposing independence between ϵ and z , we have that:

$$\begin{aligned} \sup_z \{ \pi_j \Pr[X - \delta(1 + \beta) < x, Y = 1|j, z] + \pi_{j'} \Pr[X + \delta(1 - \beta) < x, Y = 0|j', z] \} \\ \leq F(x) \leq \\ 1 - \sup_z \{ \pi_j \Pr[X - \delta(1 - \beta) > x, Y = 0|j, z] + \pi_{j'} \Pr[X + \delta(1 + \beta) > x, Y = 1|j', z] \} \end{aligned} \quad (4.11)$$

That is, we will reject δ if for any value of x , we have that:

$$\begin{aligned} 1 < \sup_z \{ \pi_j \Pr[X - \delta(1 + \beta) < x, Y = 1|j, z] + \pi_{j'} \Pr[X + \delta(1 - \beta) < x, Y = 0|j', z] \} \\ + \sup_z \{ \pi_j \Pr[X - \delta(1 - \beta) > x, Y = 0|j, z] + \pi_{j'} \Pr[X + \delta(1 + \beta) > x, Y = 1|j', z] \} \end{aligned} \quad (4.12)$$

Loosely, as $\delta \rightarrow \infty$ then δ will be rejected if $\pi_j \sup_z \{\Pr[Y = 1|j, z]\} + \pi_{j'} \sup_z \{\pi_{j'} \Pr[Y = 1|j', z]\} = 1$, where π_j is firm j 's share. That is, a sufficient condition to identify an upper bound on the switching cost parameter using only the revealed preference restrictions from our model and independence between the instruments and the unobservable is that for every state there exists a value of the instruments such that everyone switches. Analogously, as $\delta \rightarrow -\infty$, δ will be rejected if $\pi_{j'} \sup_z \{\Pr[Y = 0|j', z]\} + \pi_j \sup_z \{\Pr[Y = 0|j, z]\} = 1$. That is, if for every state there exists a value of the instruments such that no one switches. A formal version of this argument is presented in Appendix F.

Note that the following argument relied on a fixed value of the discount rate β . Since precisely identifying discount rates is notoriously difficult, this paper will impose that $\beta = 0.95$ yearly rather than attempt to identify it. One can then interpret $0.95 \times (1 + \omega)^{12}$ as the yearly premium to a dollar of returns in the pension system, relative to a dollar of commissions.

To summarize, this section has shown how one can follow the arguments in Bajari et al. (2007) and Pakes et al. (2014) to turn a dynamic choice problem into a static moment inequality. This simplification is useful, but straightforward application of moment inequality methods is problematic due to selection bias in the unobservable. Using latent variable integration methods, one can take advantage of this simplification while imposing an exclusion restriction that controls for unobserved

preference heterogeneity. Having laid out a road map for estimation, the following section discusses further implementation details of the estimation procedure.

5 Implementation

This section discusses how the ELVIS estimator is implemented in this setting. The revealed preference inequality presented in the previous section is:

$$E \left[\sum_{j \in \mathcal{J}} 1 [d_{it}^* = j] \cdot \left[\sum_{j' \in \mathcal{J} - \{j\}} 1 [\alpha \cdot w_{it} (p_{jt} - p_{j't}) + \delta \cdot \Delta (d_{it}, d_{i,t-1}, d'_{it}, \beta) + \beta^{T_i-t} \cdot B_{it} \cdot (r_{jt} - r_{j't}) \cdot (1 + \omega)^{T_i-t-1} < \zeta_{ij't} - \zeta_{ijt}] \right] \right] = 0 \quad (5.1)$$

In practice, there are a few differences between this equation and what is taken to the data. First, models that are built on parametric utility functions require location and scale normalizations, as utility is ordinal. In this case, the fact that the revealed preference moment is built off of differences in utility across choices negates the need for a location normalization. The scale normalization will be $\alpha = -1$, so that we can interpret switching cost estimates in dollars of commissions. Second, choice sets vary across time due to entry and exit of firms. Finally, since individuals are free to distribute their money between funds within a company, in practice the relevant rate of return is the weighted average across funds, using each individual's balance allocation as weights. As a result, the moment that is taken to the data is:

$$E \left[\sum_{j \in \mathcal{J}} 1 [d_{it}^* = j] \cdot \left[\sum_{j' \in \mathcal{J} - \{j\}} 1 [-w_{it} (p_{jt} - p_{j't}) + \delta \cdot \Delta (d_{it}, d_{i,t-1}, d'_{it}, \beta) + \beta^{T_i-t} \cdot B_{it} \cdot (r_{ijt} - r_{ij't}) \cdot (1 + \omega)^{T_i-t-1} < \zeta_{ij't} - \zeta_{ijt}] \right] \right] = 0 \quad (5.2)$$

where w_{it} is the relevant wage, capped at the maximum contribution level for period t ; p_{jt} is firm j 's commission rate in period t , β is the discount rate, assumed to be 5% yearly; $r_{ijt} = \frac{\sum_{i \in \{A,B,C,D,E\}} B_{it}^f r_{jt}^f}{B_{it}}$ is the relevant rate of return in period t ; and each period is one month.

As mentioned in the previous section, the additional moments that identify the model are the exclusion restrictions:

$$\begin{aligned} E [z_{ij't} \zeta_{ijt}] &= 0 \\ E [z_{ij't} \zeta_{ijt}^2] &= 0 \end{aligned} \quad \forall j, j' \in \mathcal{J}_t \quad (5.3)$$

where instruments are lagged returns $r_{j,t-1}$ and changes in wages $w_{it} - w_{i,t-1}$.

The aforementioned moments are the basis for calculating $\tilde{g}(Z, \theta, \gamma)$ in equation 4.5. Having obtained $\tilde{g}(Z, \theta, \gamma)$, one can test whether θ is in the identified set by solving:

$$L_n(\theta, \gamma) = \min_{\gamma} \tilde{g}(Z, \theta, \gamma)' V(\theta, \gamma)^{-1} \tilde{g}(Z, \theta, \gamma) \quad (5.4)$$

This is standard continuous updating GMM (CUE) (Hansen et al. (1996)). Owen (2001) shows that $n \cdot L_n(\theta, \gamma) \xrightarrow{d} \chi_q^2$, where q is the rank of the variance-covariance matrix $V(\theta_0, \gamma_0)$. As a result, the chi-squared critical values with degrees of freedom equal to the number of moments are conservative for the null hypothesis that θ is in the identified set¹³. Unfortunately, this function is non-convex and thus difficult to minimize. However, Newey and Smith (2004) show that the CUE is a member of the class of generalized empirical likelihood (GEL) estimators. In particular, the CUE is numerically identical to the GEL estimator with quadratic criterion function. This allows us to re-write the minimization problem using the primal of the GEL minimization problem:

$$\begin{aligned} \min_{\{\pi_i\}_{i=1}^N, \gamma} \sum_{i=1}^N (n\pi_i - 1)^2 \text{ s.t} \\ \sum_{i=1}^N \pi_i \tilde{g}_i^m(Z, \theta, \gamma) = 0 \forall m = 1, \dots, M \\ 0 \leq \pi_i, \sum_{i=1}^N \pi_i = 1 \end{aligned}$$

Note that this is an MPEC problem (Conlon (2013)), that $\pi_i \tilde{g}_i^m(Z, \theta, \gamma)$ is convex in γ (Schennach (2014)) and linear in π for each M , and that the objective function is quadratic in π , so the problem is very similar to Dubé et al. (2012)'s formulation of the BLP demand model. As in that setting, the hessian of the Lagrangian is sparse, and modern solvers such as Knitro can quickly solve the problem once the gradient and hessian and their sparsity patterns are correctly programmed.

Recall that to implement the ELVIS estimator one needs to select a distribution of the unobservables conditional on the observables $\rho(u|z; \theta)$ from which to draw from. Schennach (2014) shows that this decision has no impact on the statistical properties of the estimator under some regularity conditions, which include that its' support matches the true support of the unobservables. In practice, this is unlikely to be the case. However, if the chosen distribution of the unobservable has larger support than the true distribution, the estimator will be conservative. For the purposes of this paper, the

¹³Because rank $V(\theta_0, \gamma_0)$ is weakly less than the number of moments.

distribution of the unobservables conditional on the observables is a mean zero normal distribution with a standard deviation of 500 dollars. In practice, this implies that the support of the distribution of the unobservables conditional on the observables used to solve the integrals via simulation is bounded between -4000 and 4000 dollars for each unobservable in the current formulation, which uses 2000 simulation draws of the vector of unobservables $[u_{ijt}]_{j \in \mathcal{J}_t}$. Finally, for computational reasons the results in the following section are obtained using a random subsample of 100,000 individuals. Future iterations of this paper will attempt to work with the full sample.

6 Results

Figure 7 presents the estimated confidence set, obtained by solving the empirical likelihood primal program for a grid of points, where red points indicate rejections and blue points acceptances. The switching cost parameter estimates range from \$1,200 dollars to at least \$6,000 dollars, although an upper bound is not identified. These numbers are consistent with the notion that forward-looking individuals who are choosing to remain with expensive firms are giving up a significant amount of money to do so. To understand why an upper bound on switching costs is not found, reconsider the identification arguments discussed in Section 4. There, it was argued that an upper bound on the switching cost parameter will be identified if for each state, which in this case is the identity of the firm an individual is locked in to, we can find a value of the instrument such that everyone switches. At the same time, a lower bound will be identified if for each state we can find a value of the instrument such that no one switches. Since the sample consists of very few switchers, it is reasonable that without adding additional restrictions on the distribution of the unobservable one cannot identify an upper bound on the switching cost parameter. Despite this flaw, the lower bound on switching costs is informative of consumer behavior and suggests that individuals will be very inelastic. At the same time, note that finding switching costs does not imply, by itself, that prices are higher than they would be in a world with no such costs. As argued by Dubé et al. (2010) and Cabral (2012), among others, switching costs may actually lower prices if they are sufficiently low. The following section aims to determine the impact of switching costs on pricing in this market by simulating a no switching cost counterfactual. It shows that even at the lower bound of switching cost estimates the magnitude of said costs is enough to raise prices.

The parameter estimates also show that individuals behave as if they expect balances to compound at a monthly rate between 0.9% and 1.3%, which is equivalent to a yearly return between 11.4% and

16.8%. As a benchmark, Table 8 shows historical returns for each fund in the system. The column labelled “All-Time” shows average yearly returns for fund C from the creation of the system until July 2012. Note that fund C was the only available fund until 2000, when fund E was introduced. The column labelled “From September 2002” shows average yearly returns for the remaining funds since their inception until July 2012. This table shows that yearly returns for the system have been significantly lower than the parameter estimates presented above. For example, yearly returns for Fund C have averaged 8.73% since the inception of the system and 5.03% since September 2002. As a result, these findings suggest an overvaluation of the impact that returns differences across firms will have on account balances at retirement. One interpretation for this result is that by normalizing the commissions coefficient to 1, all values are expressed in dollars of commissions. If commissions are less salient than returns, as suggested anecdotically by the firms’ advertising campaigns, then the returns coefficient will be magnified by the fact that an extra dollar of commissions is worth less than an extra dollar of the individual’s account balance. This result has a similar flavor to Abaluck and Gruber (2011)’s finding that participants in Medicare Part D value premiums an order of magnitude more than they value out of pocket expenditures, or Ellison and Ellison (2009) finding that shoppers of computer memory modules are more sensitive to price differences than to tax differences across firms. This argument also has implications for the interpretation of the switching cost parameter, as the lower bound is \$1,200 dollars of commissions. That is, this parameter should not be interpreted as predicting that the mean participant in this system would not switch for \$1,000 dollars cash, but that they would not switch for savings of \$1,200 dollars in commissions.

Relative to standard estimation procedures, this work suggests a methodology that does not assume that consumers are myopic and that makes no structural assumptions on the distribution of the error term. To quantify the magnitude of these assumptions, Table 9 presents results of a standard multinomial logit model with firm fixed effects. To be specific, the utility of choosing firm j is modelled as:

$$u_{ijt} = -\delta \cdot [d_{i,t-1} \neq j] - w_{it} \cdot p_{jt} + \kappa \cdot r_{ijt} \cdot B_{it} + \vartheta_j + \epsilon_{ijt} \quad (6.1)$$

where, as before, rates of return are denoted as varying at the individual level because workers can distribute funds across the 5 available funds (within company) as they see fit. This methodology produces a switching cost estimate of \$117 dollars, an order of magnitude below the previous results. Furthermore, the returns coefficient is statistically insignificant, suggesting that individuals do not

respond to differences in returns across firms. To be precise, the estimate is that a dollar of returns is equivalent to 3.18×10^{-5} dollars of commissions. Clearly, then, imposing myopic consumers and a logit error term leads to significantly different conclusions regarding how consumers behave in this market.

7 Counterfactual Analysis

The previous section reported a confidence set for the flow utility parameters, but by themselves these results do not inform us about the key economic question at hand: what is the effect of switching costs on prices in this market? To answer this question, we need to perform counterfactual analysis, determining what prices would be if there were no switching costs. Having recovered the static parameters of a dynamic demand function is not enough to finish this task, as to obtain a demand function we need to consider the role of unobserved preference heterogeneity. This section presents a strategy for recovering the joint distribution of unobserved preferences across firms, and uses it to solve for equilibrium prices in a setting with no switching costs.

Recall, from equation 3.1, that we have modelled individuals as solving a dynamic program when choosing firms, with flow utility as a function of prices and unobserved preference heterogeneity, and a terminal payoff equal to the final retirement savings balance. In a world with no switching costs, the probability firm j is chosen simplifies to:

$$\Pr(u(j, d_{i,t-1}, X_{ijt}, \epsilon_{ijt}) - u(j', d_{i,t-1}, X_{ij't}, \epsilon_{ij't}) + \beta^{T_i-t} \cdot B_{it} \cdot E \left[(r_{jt} - r_{j't}) \prod_{\tau=t+1}^{T_i} (1 + r_{j^*\tau}) \mid \Omega_{it} \right] \geq 0 \forall j' \in \mathcal{J}_t - \{j\}) \quad (7.1)$$

That is, the probability of choosing a certain firm is a function of current period flow utility differences, and of the present discounted value of the expected difference in returns across firms. To compare these two quantities, differences in returns need to be compounded by the appropriate expected return rate, and then discounted to the present. Substituting for the chosen flow utility specification, this probability becomes:

$$\Pr \left(-w_{it} (p_{jt} - p_{j't}) + \beta^{T_i-t} \cdot B_{it} \cdot (r_{jt} - r_{j't}) \cdot (1 + \omega)^{T_i-t-1} \geq \zeta_{ij't} - \zeta_{ijt} \forall j' \in \mathcal{J}_t - \{j\} \right) \quad (7.2)$$

That is, if we knew the distribution of the unobservables, we could back out choice probabilities under the no switching cost counterfactual. One alternative to estimate the distribution of the unobservables

would be to plug in the parameter estimates into the choice model. Recall that choosing firm j in two consecutive periods implies that:

$$\begin{aligned} & u(j, j, X_{ijt}, \epsilon_{ijt}) - u(j', j, X_{ij't}, \epsilon_{ij't}) \\ & \geq \beta \cdot (E[V_{i,t+1}(j', B_{it}) | \Omega_{it}] - E[V_{i,t+1}(j, B_{it}) | \Omega_{it}]) \end{aligned} \tag{7.3}$$

And that we can build analogous inequalities for switchers. Substituting using the model's parameter estimates, we get that:

$$\begin{aligned} & -w_{it}(p_j - p_{j'}) + \hat{\delta} \\ & \geq \epsilon_{ij't} - \epsilon_{ijt} + \beta \cdot (E[V_{i,t+1}(j', B_{it}) | \Omega_{it}] - E[V_{i,t+1}(j, B_{it}) | \Omega_{it}]) \end{aligned} \tag{7.4}$$

Although knowing \hat{w} gives us some information about $E[V_{i,t+1}(j, B_{it}) | \Omega_{it}]$ for every firm j , we would still need further assumptions to find these values. It would be possible to estimate a distribution for $\epsilon_{jt} + E[V_{i,t+1}(j, B_{it}) | \Omega_{it}]$ for each firm, but this distribution would have a greater variance of heterogeneity across choices than the distribution of the unobservable that is relevant for counterfactual analysis, likely leading to markups that are too high.

Another alternative would be to use the distribution of the unobservables conditional on the observables that solves the ELVIS minimization problem to obtain the relevant distributions for counterfactual analysis. Recall that the ELVIS problem requires finding that weights γ such that:

$$\inf_{\gamma \in \mathbb{R}^{d_g}} \|E_{\pi}[\tilde{g}(Z, \theta, \gamma)]\| = 0 \tag{7.5}$$

Recall that $\tilde{g}(Z, \theta, \gamma)$ is a weighted average of the moment condition under a specific exponential family that can reproduce the same range of values of the expectation of the moments as the set of every possible conditional distribution of the unobservable. Then the set of $\hat{\gamma}(\theta)$'s that attain this minimum give the specific distributions in this family that minimize the expected value of the moments. In practice, this set may be a singleton even when θ is accepted. This will be the case whenever the minimum of the objective function is below the critical value but above 0. In these cases, a specific member of the exponential family used in ELVIS will minimize the expected value of the moments, and we can draw from that member to perform counterfactual analysis.

This procedure is restrictive, in the sense that now the conditional distribution of the unobservables is assumed to be a member of the exponential family, and this need not be the case. The reason why

this exponential family is used in ELVIS is because it has the “most adverse” property: the range of values of the expectation of the moments obtained by the distributions in this family spans the range of values one would obtain if searching over the whole space of conditional distributions of the unobservables. That is, for the purposes of testing structural parameters, searching over this family is equivalent to searching over the unrestricted space of conditional distributions of the unobservables. However, this does not mean that the conditional distribution of the unobservables that solves the unrestricted problem is the same as the conditional distribution of the unobservables that solves the ELVIS problem. That will only be the case if the conditional distribution of the unobservables that solves the unrestricted problem happens to be in the exponential family. Nevertheless, since some assumption on the distribution of the unobservables is required for counterfactual analysis, assuming that the distribution is in the exponential family seems in line with current estimation methods.

To be more specific, for each individual i and grid point s we can calculate:

$$w_{ist}(\hat{\gamma}(\theta)) = \frac{\exp(\hat{\gamma}(\theta)'g(u_s, z_{it}, \theta))}{\sum_s \exp(\hat{\gamma}(\theta)'g(u_s, z_{it}, \theta))} \quad (7.6)$$

To move from the distribution of the unobservables conditional on the observables to an unconditional distribution, one can average out these weights across the population for each grid point and form:

$$w_s(\hat{\gamma}(\theta)) = \frac{1}{N \cdot T} \sum_{t=1}^T \sum_{i=1}^N \frac{\exp(\hat{\gamma}(\theta)'g(u_s, z_{it}, \theta))}{\sum_s \exp(\hat{\gamma}(\theta)'g(u_s, z_{it}, \theta))} \quad (7.7)$$

With this estimate of the distribution of the unobservables, one can calculate choice probabilities by simulation for each accepted parameter vector. For the purposes of this exercise, I use the distribution of unobservables obtained when switching costs are \$1,200 and expected monthly returns are 0.9%. Furthermore, I assume that each PFA knows the distribution of unobservables in the population, and take 1000 draws for each individual. Assuming that marginal costs are equal to the price of the cheapest firm ever observed, Planvital’s current 0.47% commission, times the average salary of their active affiliates, I use iterated best responses to find the market equilibrium for December 2011¹⁴.

Results from this exercise are in Table 10. The second column shows actual loads charged by each firm on December 2011, while the third presents counterfactual loads calculated when all firms have the same returns realizations. Note that loads are significantly lower without switching costs, consistent with the notion that large switching costs lead to price increases. In this counterfactual, Planvital,

¹⁴That is, using the observed salaries and returns realization of that month.

the most expensive firm in 2011, drops its load from 19.09% to 4.58%, while Provida, the market leader, drops from 13.34% to 8.00%. On average, prices drop by 46%. The fourth column introduces returns heterogeneity, using each firm’s returns realizations for December 2011. Prices rise relative to the case with equal returns, but they are still below the no switching cost counterfactual. Part of this price increase is due to the excessive compounding of returns, as argued in the previous section. In order to identify the effect of over-optimistic returns expectations on prices in the no-switching cost counterfactual, the final column keeps the returns realizations of the previous column, but compounds them using historical returns for Fund C. The average difference between these two columns is 0.43%, so that over-confidence in returns expectations would lead to workers paying at least 0.43% more of their wages in commissions in a world with no switching costs¹⁵.

It is important to mention that there are reasons to think that these prices are likely to be an over-estimate of equilibrium prices if there were no switching costs. First, because the marginal cost assumption is likely to be conservative, shifting all prices up. And second, because the unobserved heterogeneity distribution includes the effects of any investments made by firms to differentiate themselves from the competition. If some of these investments are only undertaken because there are switching costs, then the current calculations have too much differentiation relative to the world with no switching costs.

Regarding the welfare effects of switching costs, note that in the current model these higher prices do not lead to welfare losses, as this only considers mandatory savings, so that any price paid is a mere transfer between firms and individuals. To obtain welfare losses from these higher prices one needs either different weights for firms and consumers in the social welfare function, or an estimate of the elasticity of contributions from informal sector workers. It would be interesting to estimate voluntary savers’ price elasticity in order to determine to what extent switching costs inhibit retirement saving for this population. This is left for future research.

8 Conclusion

This paper estimates switching costs in Chile’s privatized pension market, motivated by the finding of low switching rates and high price dispersion across firms. It finds that individuals behave as if they face a switching cost of at least 1,200 dollars, and that due to this switching cost prices

¹⁵I qualify this sentence with an “at least” because one could argue that using the historical returns for Fund C from its inception is already an overestimate, as returns in recent years have been significantly lower.

are, on average, around 2.2 times higher. Furthermore, it estimates the average returns expectation for individuals in the market, and finds that this expectation (between 11.4% and 16.8% yearly) is significantly higher than what observed returns have been in the past. This over-valuation of returns differences across firms is found to lead to prices that are, on average, 43 basis points above prices with no switching costs. Overall, these results suggest that a policy intervention aimed at reducing switching costs would lower prices. Examples of such interventions include simple reforms such as reducing the current informational requirements for online switching, and more complex options such as changing the default action from staying in the current firm to switching to the cheapest alternative. Since prices are a transfer between consumers and firms, the magnitude of the effect of lowering switching costs on welfare will depend on the elasticity of voluntary workers to prices, as this is the only margin through which quantity can be withheld. Estimating this elasticity is left for future research. The sign of this effect is unambiguous, however, such that lowering switching costs would increase welfare.

As the previous literature has recognized, there are several complications that make estimating switching costs difficult. First, as originally argued by Heckman (1981), separately identifying unobserved and persistent preference heterogeneity from state dependence is tough. Second, because the presence of switching costs implies that rational consumers should be forward looking, choosing goods by taking into account not only their current price and characteristics but also the expected evolution of these variables over time. And third, because firms also should be forward looking, maximizing the present discounted value of profits by counterbalancing the investment and the harvesting motives. To deal with these challenges, researchers have either imposed that consumers are myopic and only consider current period characteristics and prices when making their choices, or have been forced to explicitly model consumers' beliefs regarding the future evolution of characteristics.

This paper builds on this literature by estimating switching costs in pension plan choice as well as their impact on pricing, while developing a methodology that takes into account the aforementioned challenges and that is broadly applicable to settings where researchers are interested in dynamic demand models. This methodology relies on revealed preference inequalities to simplify the dynamic problem, following Bajari et al. (2007) and Pakes et al. (2014), while using Entropic Latent Variables Integration by Simulation (ELVIS) (Schennach (2014)) to deal with selection in unobserved and possibly persistent preference heterogeneity. The model is identified through exclusion restrictions, which in this setting will take the form of an independence restriction between a set of instruments, changes in wages and lagged returns, and an unobservable that is composed of both unobserved preference het-

erogeneity and expectational error. Since prices are a percentage of income, changes in wages create pricing differences across firms, helping trace out the relationship between switching costs and prices. Lagged returns affect previous choices, and as a result shift the current firm an individual is locked in to. Crucially, this estimation strategy neither assumes that consumers are myopic nor requires the econometrician to model beliefs about the evolution of future characteristics of the good. Furthermore, no assumptions beyond the aforementioned exclusion restriction are needed regarding the distribution of unobserved preference heterogeneity. Relative to traditional demand estimation frameworks, such as BLP (Berry et al. (1995)) or maximum simulated likelihood, this model relaxes constraints that are imposed on the distribution of the unobservables at the cost of a more stringent exclusion restriction and set identification. Furthermore, relative to recent work in dynamic demand estimation (Handel (2013) and Gowrisankaran and Rysman (2012)), this method requires fewer constraints on the distribution of the unobservable and does not require a model for the evolution of consumer beliefs regarding future characteristics of goods, again at the expense of a more stringent exclusion restriction and set identification.

There are several interesting avenues for future research stemming from the results in this paper. The first would be to investigate the drivers behind differentiation across firms and the estimation of more flexible returns expectation models. Determining how differentiation comes about in settings where firms are regulated to be similar is interesting, particularly if it is affected by investments such as advertising. In such a setting, there may be interesting welfare implications of allowing investment in vertical differentiation.

Another interesting question is whether specific reforms that lower switching costs, such as auctioning off the right to be the default firm for all consumers and allowing attentive individuals to opt out, would lead to lower prices and to entry and exit of firms. This requires solving a fixed point problem between forward-looking firms and consumers, a very complex task from a methodological standpoint, as well as identifying or making assumptions on the probability that consumers that opt out of the default firm will become inattentive in the future.

References

- Jason Abaluck and Jonathan Gruber. Choice inconsistencies among the elderly: Evidence from plan choice in the medicare part d program. *American Economic Review*, 101(4):1180–1210, 2011.
- Daniel A. Akerberg. Advertising, learning, and consumer choice in experience good markets: an empirical examination. *International Economic Review*, 44(3):1007–1040, 08 2003. URL <http://ideas.repec.org/a/ier/iecrev/v44y2003i3p1007-1040.html>.
- B. Aguirre and P. Obregon. Aumentan los premios por cambiarse de afp: hasta \$60 mil para los cotizantes de altas rentas. *El Mercurio*, 2014. URL <http://www.economiaynegocios.cl/noticias/noticias.asp?id=116568>.
- Donald W.K. Andrews and Xiaoxia Shi. Nonparametric inference based on conditional moment inequalities. *Journal of Econometrics*, 179(1):31–45, 2014. URL <http://ideas.repec.org/a/eee/econom/v179y2014i1p31-45.html>.
- Patrick Bajari, C. Lanier Benkard, and Jonathan Levin. Estimating dynamic models of imperfect competition. *Econometrica*, 75(5):1331–1370, 09 2007. URL <http://ideas.repec.org/a/ecm/emetrp/v75y2007i5p1331-1370.html>.
- Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica*, 63(4):841–90, July 1995. URL <http://ideas.repec.org/a/ecm/emetrp/v63y1995i4p841-90.html>.
- Luis Cabral. Switching costs and equilibrium prices. 2012.
- Victor Chernozhukov, Christian Hansen, and Michael Jansson. Finite sample inference for quantile regression models. *Journal of Econometrics*, 152(2):93–103, October 2009. URL <http://ideas.repec.org/a/eee/econom/v152y2009i2p93-103.html>.
- Victor Chernozhukov, Sokbae Lee, and Adam M. Rosen. Intersection bounds: Estimation and inference. *Econometrica*, 81(2):667–737, 03 2013. URL <http://ideas.repec.org/a/ecm/emetrp/v81y2013i2p667-737.html>.
- Andrew Chesher, Adam M. Rosen, and Konrad Smolinski. An instrumental variable model of multiple discrete choice. *Quantitative Economics*, 4(2):157–196, 07 2013. URL <http://ideas.repec.org/a/ecm/quante/v4y2013i2p157-196.html>.

- Raj Chetty, Adam Looney, and Kory Kroft. Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–1177, 2009. ISSN 00028282. doi: 10.1257/aer.99.4.1145.
- Christopher T. Conlon. The empirical likelihood mpec approach to demand estimation. December 2013. URL <http://ssrn.com/abstract=2331548>.
- Peter Diamond and Salvador Valdés. Social security reforms. In B. Bosworth, R. Dornbusch, and R. Labán, editors, *Chilean Economy: Policy Lessons and Challenges*, chapter 4. Brookings Institution, Washington, DC, 1993.
- Michael Dickstein and Eduardo Morales. Accounting for expectational and structural error in binary choice problems: A moment inequality approach. Technical report, NBER Working Paper No. 19486, 2013.
- Fabian Duarte and Justine S. Hastings. Fettered consumers and sophisticated firms: Evidence from Mexico’s privatized social security market. NBER Working Papers 18582, National Bureau of Economic Research, Inc, December 2012. URL <http://ideas.repec.org/p/nbr/nberwo/18582.html>.
- J.-P. Dubé, Che-Lin Su, and Jeremy T. Fox. Improving the numerical performance of blp static and dynamic discrete choice random coefficients demand estimation. *Econometrica*, 80(5):2231–2267, September 2012.
- Jean-Pierre Dubé, Günter J. Hitsch, and Peter E. Rossi. State dependence and alternative explanations for consumer inertia. *RAND Journal of Economics*, 41(3):417–445, 2010. URL <http://ideas.repec.org/a/bla/randje/v41y2010i3p417-445.html>.
- Glenn Ellison and Sara Fisher Ellison. Tax sensitivity and home state preferences in internet purchasing. *American Economic Journal: Economic Policy*, 1(2):53–71, 2009. ISSN 1945-7731. doi: 10.1257/pol.1.2.53.
- Tülin Erdem and Michael P. Keane. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science*, 15(1):1–20, 1996. URL <http://ideas.repec.org/a/inm/ormksc/v15y1996i1p1-20.html>.
- Joseph Farrell and Paul Klemperer. *Coordination and Lock-In: Competition with Switching Costs and Network Effects*, volume 3 of *Handbook of Industrial Organization*, chapter 31, pages 1967–2072. Elsevier, 00 2007. URL <http://ideas.repec.org/h/eee/indchp/3-31.html>.

- Joseph Farrell and Carl Shapiro. Dynamic competition with switching costs. *RAND Journal of Economics*, 19(1):123–137, Spring 1988. URL <http://ideas.repec.org/a/rje/randje/v19y1988ispringp123-137.html>.
- Alfred Galichon and Marc Henry. Set identification in models with multiple equilibria. *The Review of Economic Studies*, 78(4):1264–1298, 2011. doi: 10.1093/restud/rdr008. URL <http://restud.oxfordjournals.org/content/78/4/1264.abstract>.
- Gautam Gowrisankaran and Marc Rysman. Dynamics of consumer demand for new durable goods. *Journal of Political Economy*, 120(6):1173 – 1219, 2012. URL <http://EconPapers.repec.org/RePEc:ucp:jpolec:doi:10.1086/669540>.
- Gabriela Gurovic. Análisis del desempeño de los multifondos en el sistema de pensiones chileno: Benchmarks por grupos y por clases de activos. Master’s thesis, Instituto de Economía, Pontificia Universidad Católica de Chile, 2005.
- Philip A. Haile and Elie Tamer. Inference with an incomplete model of english auctions. *Journal of Political Economy*, 111(1):1–51, February 2003. URL <http://ideas.repec.org/a/ucp/jpolec/v111y2003i1p1-51.html>.
- Benjamin R. Handel. Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, 103(7):2643–82, December 2013. URL <http://ideas.repec.org/a/aea/aecrev/v103y2013i7p2643-82.html>.
- Lars Peter Hansen, John Heaton, and Amir Yaron. Finite-sample properties of some alternative gmm estimators. *Journal of Business and Economic Statistics*, 14(3):262–80, July 1996. URL <http://ideas.repec.org/a/bes/jnlbes/v14y1996i3p262-80.html>.
- Justine S. Hastings and Lydia Tejada-Ashton. Financial literacy, information, and demand elasticity: Survey and experimental evidence from mexico. NBER Working Papers 14538, National Bureau of Economic Research, Inc, December 2008. URL <http://ideas.repec.org/p/nbr/nberwo/14538.html>.
- Justine S. Hastings, Ali Hortagsu, and Chad Syverson. Advertising and competition in privatized social security: The case of mexico. NBER Working Papers 18881, National Bureau of Economic Research, Inc, March 2013. URL <http://ideas.repec.org/p/nbr/nberwo/18881.html>.

- James J. Heckman. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process and some monte carlo evidence. In Charles Manski and Daniel McFadden, editors, *Structural Analysis of Discrete Data with Econometric Applications*, chapter 4, pages 179–195. MIT Press, Cambridge, MA, 1981.
- Igal Hendel and Aviv Nevo. Intertemporal price discrimination in storable goods markets. *American Economic Review*, 103(7):2722–51, December 2013. URL <http://ideas.repec.org/a/aea/aecrev/v103y2013i7p2722-51.html>.
- Kate Ho and Ariel Pakes. Hospital choices, hospital prices and financial incentives to physicians. *American Economic Review (forthcoming)*, 2014.
- Elena Krasnokutskaya and Petra Todd. Investor behavior and fund performance under a privatized retirement accounts system: Evidence from chile. Working Papers wp209, University of Michigan, Michigan Retirement Research Center, September 2009. URL <http://ideas.repec.org/p/mrr/papers/wp209.html>.
- Fernando Luco. *Switching Costs and Competition in Retirement Investment*. PhD thesis, Northwestern University, 2014.
- Eduardo Morales, Gloria Sheu, and Andrés Zahler. Gravity and extended gravity: Using moment inequalities to estimate a model of export entry. NBER Working Papers 19916, National Bureau of Economic Research, Inc, February 2014. URL <http://ideas.repec.org/p/nbr/nberwo/19916.html>.
- Whitney K. Newey and Richard J. Smith. Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica*, 72(1):219–255, 01 2004. URL <http://ideas.repec.org/a/ecm/emetrp/v72y2004i1p219-255.html>.
- Kathleen Nosal. Estimating switching costs for medicare advantage plans. 2012.
- OECD. Pensions at a glance 2005: Retirement-income systems in OECD and G20 countries. *OECD Publishing*, 2005.
- OECD. Pensions at a glance 2011: Retirement-income systems in OECD and G20 countries. *OECD Publishing*, 2011.
- OECD. Survey of investment regulation of pension funds. *OECD Publishing*, 2012.

- José A. Olivares. *On the Chilean Pension Funds Market*. PhD thesis, The University of Texas at Austin, 2004.
- Art B. Owen. *Empirical likelihood*. CRC Press LLC, 2001.
- Ariel Pakes, Jack Porter, Kate Ho, and Joy Ishii. Moment inequalities and their application. *Econometrica (forthcoming)*, 2014.
- Maria Polyakova. *Regulation of insurance with adverse selection and switching costs: Evidence from Medicare Part D*. PhD thesis, Massachusetts Institute of Technology, 2014.
- Claudio Raddatz and Sergio Schmukler. Deconstructing herding: Evidence from pension fund investment behavior. *Journal of Financial Services Research*, 43(1):99–126, February 2013. URL <http://ideas.repec.org/a/kap/jfsres/v43y2013i1p99-126.html>.
- Susanne M. Schennach. Entropic latent variable integration via simulation. *Econometrica*, 82(1):345–385, 01 2014. URL <http://ideas.repec.org/a/wly/emetrp/v82y2014i1p345-385.html>.
- David Tuesta. Factors behind the administrative fees of private pension systems: an international analysis. *Journal of Pension Economics and Finance*, pages 1–24, 2013.
- Eduardo Walker. Desempeño financiero de las carteras de renta fija de los fondos de pensiones: Ha tenido desventajas ser grande? *Latin American Journal of Economics - formerly Cuadernos de Economía*, 30(89):1–34, 1993a.
- Eduardo Walker. Desempeño financiero de las carteras accionarias de los fondos de pensiones: Ha tenido desventajas ser grande? *Latin American Journal of Economics - formerly Cuadernos de Economía*, 30(89):35–76, 1993b.
- F Zúñiga. Desempeño de los fondos de pensiones: Impacto de las restricciones legales. Master’s thesis, Instituto de Economía, Pontificia Universidad Católica de Chile, 1992.
- Salvador Zurita and Carlos Jara. Desempeño financiero de fondos de pensiones. *Estudios Públicos*, Fall, 1999.

A International Commission Comparison

To map the commissions in the Chilean system to expense ratios, consider an individual who works for 40 years at a starting salary of \$1000 dollars, which grows 2% annually in real terms. Assume that the real return rate is 5%, and define “Accumulated Commissions at Retirement” to be the sum of commissions paid throughout the individual’s lifetime, compounded by the real rate of return. Column 2 of Table 11 shows the “Retirement Load”, defined as:

$$PDV\ Load = \frac{Accumulated\ Commissions\ at\ Retirement}{Accumulated\ Commissions\ at\ Retirement + Account\ Balance\ at\ Retirement} \quad (A.1)$$

if this individual works in Chile. Column 3 shows the expense ratio that would generate the same retirement load for an identical individual who is saving in the US, or the “Equivalent Expense Ratio”. This table also assumes that the only commission faced by the US worker is an expense ratio. If this individual is in the Chilean pension system and chooses the cheapest PFA between 2002 and 2011, she pays 1.14% of her salary as a commission every month. The retirement load for this individual is 10.23%, which is also the PDV load of an investment vehicle that charges an expense ratio of 46 basis points, or 0.46%. If this individual were paying the mean commission rate in Chile during this period, the equivalent expense ratio would be 66 basis points, while if this individual were paying the maximum commission rate observed in Chile, the equivalent expense ratio would be 88 basis points. For comparison, as of November of 2014 the “Vanguard Target Retirement 2050 Fund” has an expense ratio of 18 basis points, the “Fidelity Freedom 2050 Fund” has an expense ratio of 78 basis points, and the MIT retirement plan offers investment options with expense ratios starting at 11 basis points. Therefore, the Chilean commission rates seem high relative to what one could purchase in the US, particularly considering that the inexpensive options that are available in the US were not available in Chile during this period¹⁶. However, one can certainly find more expensive alternatives in the US system.

B Differentiation on Returns

Along with the aforementioned investment limits, there is a return band regulation that specifies that by PFA, fund and month, the annualized monthly average return for the last 36 months cannot be

¹⁶Reforms introduced after 2011 have led to further drops in commission rates in Chile. The minimum equivalent expense ratio in 2015 is 20 basis points.

lower than the minimum of (1) the industry weighted average annualized monthly average return for the last 36 months, weighted by fund assets, minus 4 percentage points for funds A and B and minus 2 percentage points for funds C, D, and E; and (2) the industry weighted average annualized monthly average return for the last 36 months, weighted by fund assets, minus one half the absolute value of said return. If any PFA falls below this band, it must cover the difference between its return and the band floor. At the same time, until 2008 if any PFA's return was higher than the aforementioned industry weighted average return plus the minimum of (1) and (2), the difference was not accrued by customers. Instead, it was kept in a yield fluctuation account, which is used to cover the event that the PFA does not realize the minimum return in the future (Olivares (2004), Krasnokutskaya and Todd (2009)).

While setting investment limits makes it difficult for a PFA to significantly differentiate itself from its competitors through its investment strategy, the return band regulation makes it unprofitable: if a firm outperforms the market, benefits to its customers are capped, while if it under-performs, it must cover the losses out its the shareholder's pockets. Raddatz and Schmukler (2013) analyze the incentives created by this return band regulation, and document that pension fund administrators exhibit herding behavior in their investment decisions, particularly in assets for which there is less market information and in periods where risk increases. Overall, there is an extensive literature that argues that although PFAs have different return realizations each period, one should not expect one PFA to consistently out-perform the competition (Walker (1993a,b); Zúñiga (1992); Zurita and Jara (1999); Diamond and Valdés (1993); Gurovic (2005)). Figure 8 shows monthly returns, by fund, between November 2002 and July 2012, while Table 12 calculates pairwise correlations across companies for the same period. Note that returns are highly correlated across companies, particularly for the riskier funds.

As a simple test of the hypothesis that price dispersion is due to vertical differentiation in returns, Table 14 studies whether commission rates are correlated with returns, by regressing PFA i 's monthly return for fund f on its commission rate that month (p_{it}) and its fund f monthly return for the previous month, controlling for date and fund effects, for the period between November 2002 and July 2012:

$$r_{ift} = \alpha_0 + \alpha_1 p_{it} + \alpha_2 r_{if,t-1} + \delta_t + \gamma_f + \epsilon_{ift} \quad (\text{B.1})$$

The first column of this table presents results combining all funds, and shows that the null hypothesis that there is no correlation between commission rates and returns cannot be rejected. Columns 2

through 6 present results for each fund, A through E¹⁷. Recall that A is the riskiest fund and E is the safest. There is no correlation between commission rates and returns for risky funds, and a negative correlation between returns and commission rates for the safest funds. An increase in commission rates of 100 percentage points is associated with a decrease in returns of 0.11 percentage points, so this is not an economically significant effect.

Finally, the way retirement is structured offers a final piece of evidence regarding differentiation on returns. Since 2004, any individual who wishes to retire in Chile must enter an exchange called SCOMP¹⁸. This system automatically sends all relevant information about the individual to PFAs and to life insurance companies. Life insurance companies offer annuities, and bid freely on each individual. It is common for retirees to have hundreds of offers, generated by each life insurance company offering annuity contracts with different terms over features such as guarantee periods and deferral periods. PFAs offer a service called “programmed withdrawal”, which is a front loaded alternative to an annuity that draws down savings using a pre-determined and regulated declining payout schedule. The key aspect of this system is that in order to retire an individual must observe all offers and actively choose an option, so there is no logistical incentive to remain at the same PFA. If an individual chooses programmed withdrawal, the PFA continues to invest their money, and the individual accrues any gains or losses. PFAs earn money on these consumers by charging a fee that is a percentage of the withdrawal amount. Figure 9 plots these fees for the period between 2002 and 2012. Notice that fees are starkly similar, with most companies charging 1.25% of the amount withdrawn. If firms had differential investment capabilities, and were able to price on this in the accumulation phase of a worker’s life, they should also be able to price on this during the retirement period, as higher returns still increase retirement payouts in this stage. However, that does not seem to be the case.

C Evidence from 2010’s Auction Reform

One recent regulatory reform provides further evidence that is consistent with the predictions of switching cost models. In 2010, a reform to auction off the right to serve new customers for their first two years in the labor force was implemented. That is, new workers cannot choose PFAs starting in 2010. Instead, they are bound for two years from the date in which they start working to the PFA that offers the lowest commission rate in the auction, which takes place every two years. Both existing PFAs

¹⁷Obviously, these regressions do not have fund effects.

¹⁸Sistema de Consultas y Ofertas de Montos de Pension.

and new entrants are allowed to bid, and the winner must set a commission rate equal to or lower than its bid for the duration of the period. The auction rules also force PFAs to offer the same commission rate to all their customers if they win. Table 13 shows bids, expressed as a percentage of income, the loads that those bids imply (“Bid Load”), and the loads these companies were charging at the time (“Current Load”). In 2010, a new entrant, Modelo, won the auction with a load of 10.23%. This was the first time entry had been observed in this market since the early 90’s. Note that Planvital, the most expensive company in the system, lost with a bid load of 10.63%, almost half the 19.09% load they were charging at the time. In 2012 Modelo won again, with a load of 7.15%, and this time Planvital was the only other existing company that bid, with a bid load of 7.83%. At this time Planvital was still charging 19.09%. An entrant that never materialized, Regional, also bid and lost. Finally, in 2014 Planvital won the auction, with a bid load of 4.49%, almost a fifth of the 19.09% they were charging at the time.

Several facts from these auctions are consistent with the notion that demand inertia is a relevant force in this market. First, the fact that there was entry into the market for the first time since the early 90’s could be attributed to this being a market with large fixed costs and demand inertia, where it is difficult to profitably enter. This also explains why Regional was willing to enter if it won the auction but does not enter after losing it. Second, Planvital’s and Modelo’s behavior fits in nicely with the predictions of the overlapping generations switching cost model in Farrell and Shapiro (1988). Using a two-firm overlapping generations model, this work finds that one firm will specialize on new customers, setting a low price, and the other will specialize in old customers, setting a high price. As time passes, firms reverse roles, with the previously cheap firm raising its price to extract rents from its locked in customers, and the previously expensive firm lowering prices to rebuild its consumer base, depleted from consumer exit from the market. Modelo’s entry, as well as Planvital’s price drop in 2014, coincide with the latter prediction. It remains to be seen whether Modelo will raise its price in 2016, when it is first allowed to do so.

Would it be possible to explain this behavior if there were no switching costs? In such a setting, the only difference between entry after winning the auction and entry if the auction did not exist is the fact that in the former case the entire mass of new consumers is locked-in to the entrant for their first two years in the system, while in the latter only some consumers would have chosen the entrant. Therefore, the only reason why we see entry now taking place, and the only reason why Regional chooses not to enter after losing the auction, is because of the marginal profits derived from

having a fraction of new customers locked-in for only two years. This is because in a world with no switching costs only the individuals who would have chosen the entrant anyway would stay with the entrant after the lock-in period. Considering that the inflow of new customers is small relative to the stock of existing customers, this is implausible. I calculate that Modelo's PDV of revenues derived from four years of locked-in consumers obtained after winning the first auction¹⁹ to be around \$81 million dollars, which corresponds to 9.3% of the PDV of system revenues for the just the first year of Modelo's existence²⁰. Furthermore, a model with no switching costs would rationalize Planvital's bidding behavior by arguing that the profits derived from having the mass of new customers locked in for their first two years is greater than the inframarginal rents lost from lowering their load for four years. This is also implausible, particularly for the Planvital's behavior in the second and third auctions. I estimate Planvital's increase in revenues if it had won the auction in 2010 to be \$16 million dollars. Considering that the second and third auctions had Planvital bidding much more aggressively than in the first, it is unlikely that these bids would have been profitable if Planvital didn't expect that the presence of switching costs would keep a large fraction of workers in Planvital after their 2 year mandatory period expired.

¹⁹Recall that individuals are mandated to stay in Modelo for their first two years in the system, so someone who starts working in the last month of Modelo's period (August 2012) is locked-in to Modelo until August 2014.

²⁰To do this calculation, I assume a 5% discount rate and that everyone who chose Modelo between 2010 and 2012 is a new worker. To the extent that some individuals switched to Modelo, and considering that some locked-in individuals would have chosen Modelo if it had entered without the auction, this calculation is an overestimate.

D Tables and Figures

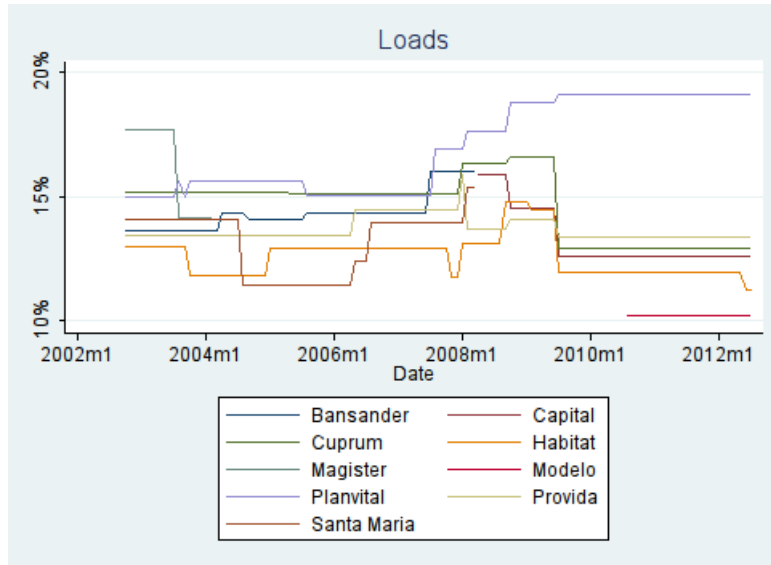


Figure 1: Loads by PFA, 2002 to 2012

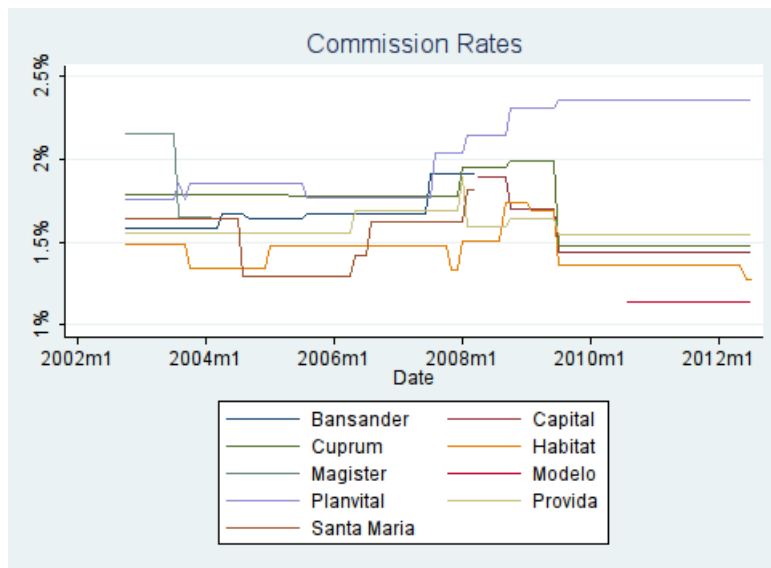


Figure 2: Commission Rates by PFA, 2002 to 2012

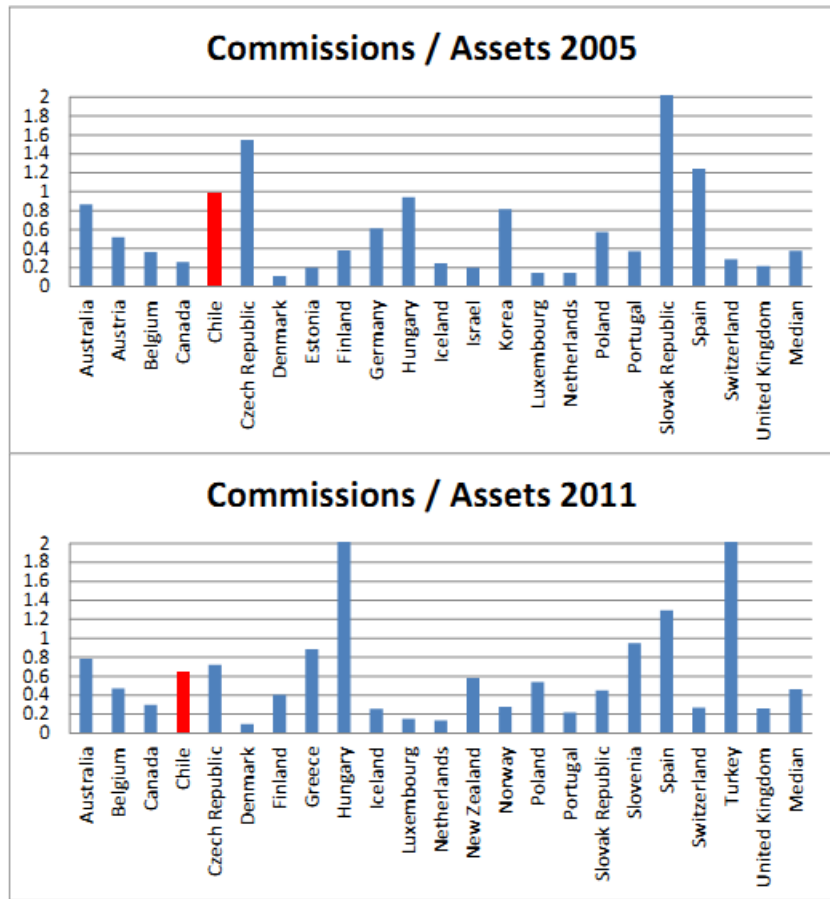


Figure 3: Comparison Across Pension Systems

| Probabilities of Choosing the Absorbing Firm in January 2007 | | |
|--|---------------------|---------------------|
| | (1) | (2) |
| Absorbed Firms' Customers | 0.267*** (0.034) | 0.475*** (0.033) |
| Absorbing Firms' Customers | | 0.651*** (0.005) |
| Constant | 0.274*** (0.002) | 0.094*** (0.002) |
| Observations | 35,020 | 35,020 |

Notes: This table presents the results of the linear regression of whether an individual chooses an absorbing firm in January 2007 on whether an individual chooses an absorbed firm at the time of merger and whether an individual chooses an absorbing firm at the time of merger. Only mergers where one merger partner disappears ("Absorbed Firm") and the other merging partner continues ("Absorbing Firm") until January 2007 are included. Mergers that involve the creation of a new brand are excluded. "Absorbing Firm's Customers" are individuals who chose the firm that disappears in the month before the merger. "Absorbed Firm's Customers" are individuals who chose the firm that continues in the month before the merger. There are 7 mergers that qualify under these criteria: Planvital and Invierta (1993), Provida and El Libertador (1995), Santa Maria and Banguardia (1995), Planvital and Concordia (1996), Provida and Union (1998), Provida and Proteccion (1999), and Planvital and Magister (2004). Individuals are matched to firms using data on fixed commissions paid. Due to data constraints I cannot identify Magister's consumers in 2004, so this merger is not considered in the analysis. All regressions include merger fixed effects.

Figure 5: Probabilities of Choosing the Absorbing Firm in January 2007

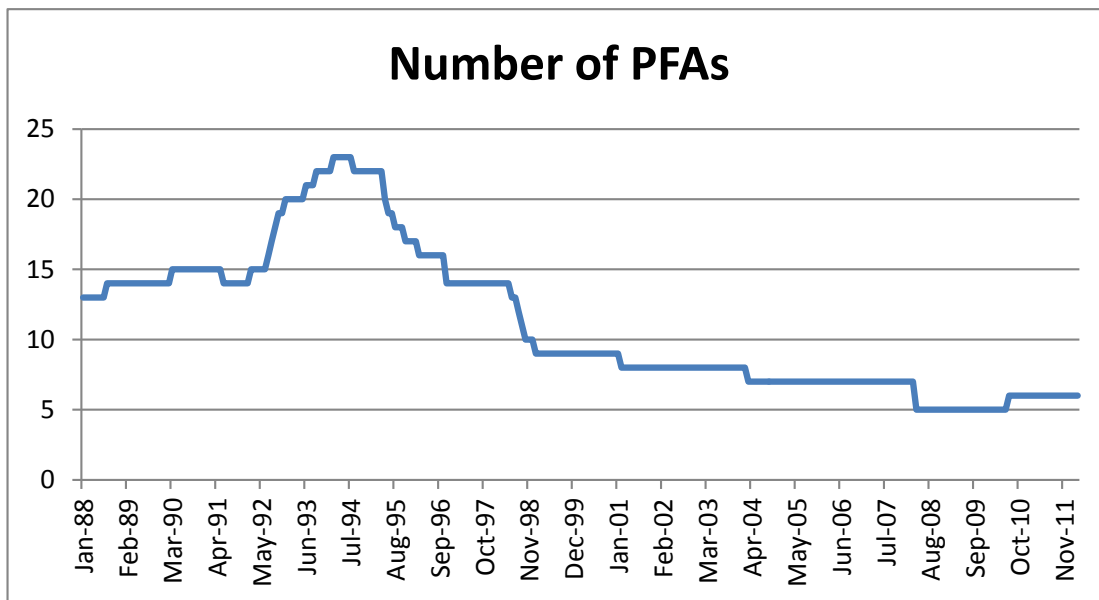


Figure 4: Number of PFAs

| Fund | Equity | | | Bonds | | | Retail Investment Funds | | |
|------|--|--|-------------------------|--|--|---|--------------------------------|---|--|
| | Variable Income Securities (min. - max.) | Domestic Public Limited Companies (sub-limit.) | Government Bonds (max.) | Convertible Bonds, local and foreign (sub-limit) | Not Investment Grade Bonds (sub-limit) | Closed and Open Ended Mutual Funds (max.) | Mutual Fund Shares (sub-limit) | Committed Payments for Closed-Ended Funds | |
| A | 40% - 80% | 60% | 40% | 30% | 5% | 40% | 5% of the value | 2% | |
| B | 25% - 60% | 50% | 40% | 30% | 4% | 30% | of the fund, per | 2% | |
| C | 15% - 40% | 30% | 50% | 10% | 3% | 20% | issuer. | 2% | |
| D | 5% - 20% | 15% | 70% | 10% | 2% | 10% | | 2% | |
| E | 0% - 5% | 5% | 80% | 3% | 0% | 5% | | 2% | |

Notes: Limits for closed and open ended mutual funds include shares and committed payments. Only funds approved by the risk rating commission are eligible. Investments in private investment funds are not allowed, and neither are direct investments in real estate.

Table 1: Investment Caps, by Fund

| Summary Statistics for Sample | |
|--|--------|
| % Female | 44.5% |
| Mean Age | 41.5 |
| Mean Monthly Wages, US\$ | 491.6 |
| Mean Non-zero Monthly Wages, US\$ | 908.2 |
| % Zero Wage | 45.9% |
| % Always Zero Wage | 21.5% |
| Mean Account Balance, US\$ | 14,211 |
| Mean Account Balance if Employed, US\$ | 20,238 |

Table 2: Summary Statistics

| Summary Statistics on Switches | |
|--------------------------------|--------|
| Monthly Switching Rate | 0.31% |
| Yearly Switching Rate | 3.18% |
| Never Switch | 89.25% |
| Switch Once | 7.29% |
| Switch Twice | 1.77% |
| Switch Three Times or More | 1.69% |
| Mean Wage at Switch | 1,237 |
| Mean Age at Switch | 38.2 |
| Mean Balance at Switch | 26,006 |

Notes: This table presents summary statistics on switching behavior for the sample period (January 2007-December 2011).

Table 4: Switching Statistics

| Switching and Wage Changes | | | |
|----------------------------|-------------|--------------------------------|------|
| | (1) | Relative Switching Probability | |
| Wage Change (in dollars) | 8.91E-06*** | \$500 Increase | 2.40 |
| | (0.000) | \$100 Increase | 1.28 |
| Constant | 0.003*** | \$10 Increase | 1.03 |
| | (0.000) | \$100 Decrease | 0.72 |
| Individual Fixed Effects | | Yes | |
| Observations | | 1189327 | |

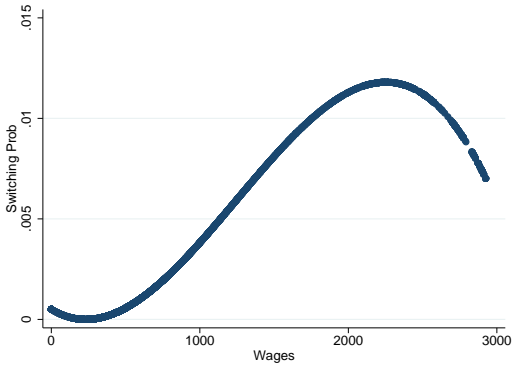
Notes: This table reports results of the regression of a dummy for whether an individual switches PFAs in period t on the wage change (in dollars) between period t-1 and period t. Regressions include individual fixed effects and heteroskedasticity-robust standard errors. The column labelled "Relative Switching Probability" posits wage changes and displays each group's switching probability relative to the no wage change baseline.

Table 5: Switching and Wage Changes

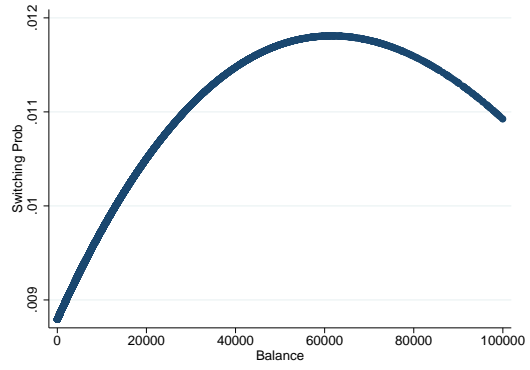
| Summary Statistics For Account Balances, in Dollars | | | | | | | | |
|---|----|-----|-------|-------|--------|--------|--------|---------|
| Percentile | 1% | 10% | 25% | 50% | 75% | 90% | 95% | 99% |
| Balance (US\$) | 40 | 200 | 2,060 | 6,460 | 16,540 | 34,980 | 53,800 | 115,300 |
| Balance, Always 0 Wage | 0 | 96 | 367 | 1,506 | 4,917 | 12,330 | 19,892 | 51,104 |

Notes: This table presents summary statistics for account balances, for all individuals in the sample and for individuals who always have a zero wage in the sample. All values are in US dollars.

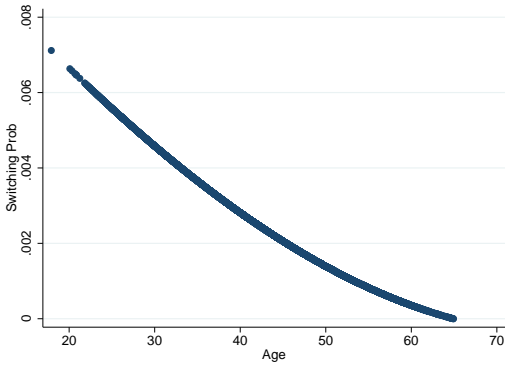
Table 3: Summary Statistics for Account Balances



(a) Predicted Switching Probability and Wage



(b) Predicted Switching Probability and Account Balance



(c) Predicted Switching Probability and Age

Figure 6: Correlations between Switching and Observables

| Table: Correlation between Salesforce / Advertising and Returns | | | | | |
|---|-----------------------|------------------------|------------------------|-------------------------|-------------------------|
| Panel A: Salesforce (# of workers) | | | | | |
| Fund | A | B | C | D | E |
| Monthly Returns | 142.1 (189.57) | 176.7 (274.12) | 314.4 (438.42) | 606.9 (793.21) | -433.5 (674.95) |
| Semester Returns | -179.8*** (55.28) | -262.8*** (79.42) | -413.9*** (135.21) | -567.7** (263.13) | 692.2** (325.11) |
| N | 222 | 222 | 222 | 222 | 222 |
| Panel B: Advertising Expenditure (US\$) | | | | | |
| Fund | A | B | C | D | E |
| Monthly Returns | -3,885.7 (9768.74) | -3,045.8 (13807.41) | -7,017.7 (21440.07) | -25,072.7 (37023.32) | 12,792.8 (44465.35) |
| Semester Returns | -146.2 (3353.04) | 1,479.5 (4742.76) | 3,726.5 (7756.29) | 9,806.5 (14367.18) | 34,717.0* (19837.85) |
| N | 615 | 615 | 615 | 615 | 615 |

Notes: This table presents regressions of number of salesforce workers and advertising expenditures on monthly and semester returns, by fund. All specifications include PFA fixed effects and heteroskedasticity-robust standard errors.

* Significant at 10% ** Significant at 5% *** Significant at 1%

Table 6: Salesforces / Advertising and Returns

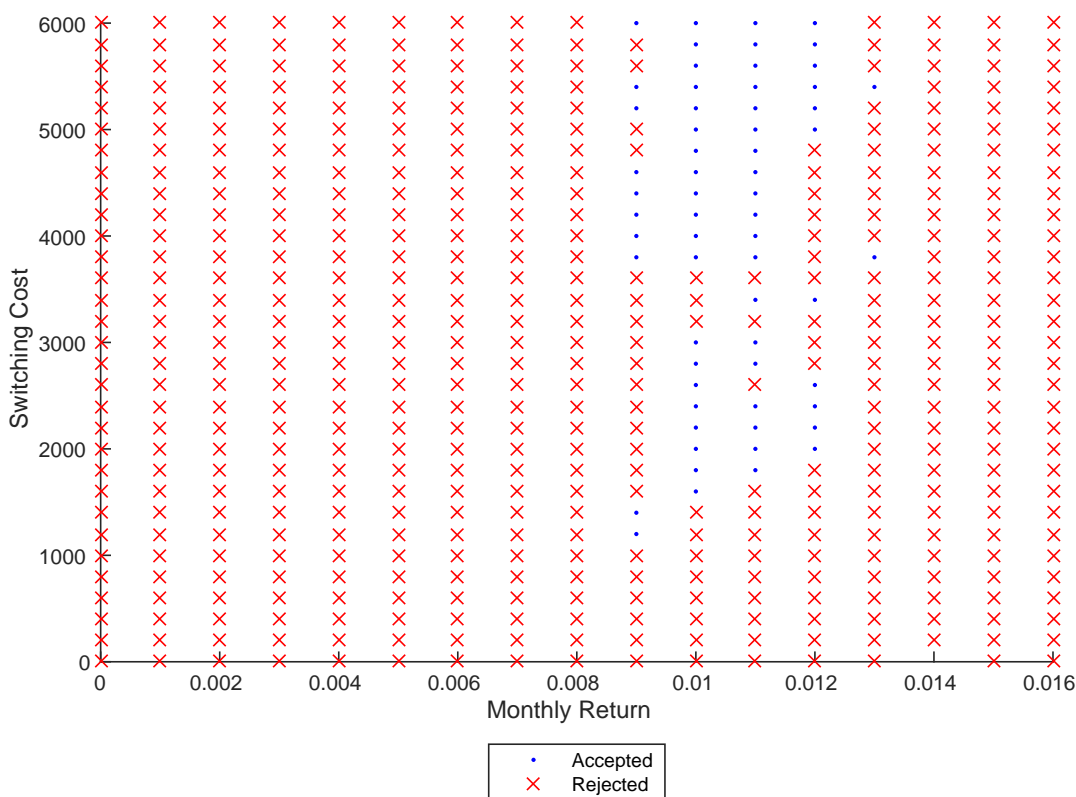


Figure 7: Parameter Estimates

PDV of Commission Savings Between Cheapest Path and Observed Path

| | | Panel A: All Accounts | | | | | | | | | |
|---|--|--------------------------------------|-----|-----|-----|-----|-------|-------|-------|-------|--|
| Percentile | | 1% | 5% | 10% | 25% | 50% | 75% | 90% | 95% | 99% | |
| PDV Commission Savings | | 0 | 0 | 0 | 1 | 176 | 690 | 1,475 | 2,110 | 2,990 | |
| PDV Commission Savings + Balance Difference | | -132 | -11 | 1 | 68 | 424 | 1,119 | 2,157 | 3,048 | 4,821 | |
| | | Panel B: Excluding Inactive Accounts | | | | | | | | | |
| Percentile | | 1% | 5% | 10% | 25% | 50% | 75% | 90% | 95% | 99% | |
| PDV Commission Savings | | 0 | 1 | 4 | 31 | 373 | 868 | 1,686 | 2,288 | 3,050 | |
| PDV Commission Savings + Balance Difference | | -128 | -4 | 21 | 198 | 630 | 1,347 | 2,402 | 3,320 | 5,037 | |

Notes: This table presents statistics that summarize how much money individuals stand to save, in PDV terms, from switching to a cheaper firm. The row titled "PDV Commission Savings" takes each individuals' observed choice path between January 2007 and December 2011, and compares the PDV of commissions paid at that path versus at the cheapest path. For the time period after 2011, the comparison assumes individuals stay at their December 2011 firm until retirement, that wages are fixed at their December 2011 level, and that commissions follow the observed path until 2015, at which point they remain constant. To calculate PDVs, a discount rate of 5% is assumed. The row titled "PDV Commission Savings + Balance Difference" adds to the previous amount the difference in account balances in December 2011 (or at retirement for earlier retirees) from picking the cheapest path versus the observed choice path. Panel A includes all accounts in the database, and Panel B excludes accounts without any contributions in the sample period.

Table 7: PDV of Commission Savings Between Cheapest and Observed Paths

| Historical Yearly Return Rates | | |
|--------------------------------|----------|---------------------|
| Fund | All-Time | From September 2002 |
| A | - | 6.43% |
| B | - | 5.57% |
| C | 8.73% | 5.03% |
| D | - | 4.56% |
| E | - | 3.80% |

Notes: This table shows yearly returns, averaged across companies, for the time period between July 1981 and July 2012 ("All Time") as well as for the period between September 2002 to July 2012. Note that fund C was the only fund available until May 2000, when fund E was introduced. The remaining funds began in September of 2002.

Table 8: Historical Returns

| Myopic Multinomial Logit Parameter Estimates | |
|--|-------------------------|
| | (1) |
| Switching Cost | 116.98*** (41.77) |
| Returns Coefficient | 3.18E-05 (2.08E-005) |
| Firm Fixed Effects | Yes |

Notes: This table presents parameter estimates for a multinomial logit PFA choice model. The relevant characteristics are switching cost, relevant price (the product of individual wages and fees), and relevant returns (the product of individual balances in each fund and returns for that fund). This specification includes firm fixed effects.

* Significant at 10% ** Significant at 5% *** Significant at 1%

Table 9: Myopic Multinomial Logit Estimates

| Prices under a No Switching Cost Counterfactual | | | | |
|---|-------------|-------------------------|------------------------------|-----------------------------|
| Company | Actual Load | Counterfactual Loads | | |
| | | No Return Heterogeneity | Estimated Return Compounding | Observed Return Compounding |
| Capital | 12.60% | 7.83% | 8.09% | 7.92% |
| Cuprum | 12.90% | 7.49% | 7.92% | 7.58% |
| Habitat | 11.97% | 5.12% | 5.21% | 5.21% |
| Modelo | 10.23% | 6.37% | 6.80% | 6.37% |
| Planvital | 19.09% | 4.58% | 4.94% | 4.67% |
| Provida | 13.34% | 6.54% | 8.00% | 6.63% |

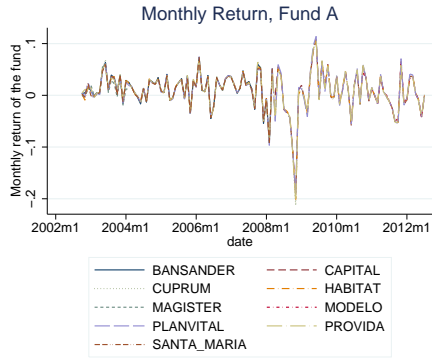
Notes: This table presents loads charged by firms in December of 2011, as well as simulated loads for the same date under a no switching cost counterfactual. The column labelled "Estimated Return Compounding" assumes individuals compound returns differences across firms according to parameter estimates, while the column marked "Observed Return Compounding" assumes individuals compound said differences using Fund C's historical monthly return from September 2002 to December 2011. Finally, the column labelled "No Return Heterogeneity" assumes that there are no returns differences across firms.

Table 10: Counterfactual Pricing

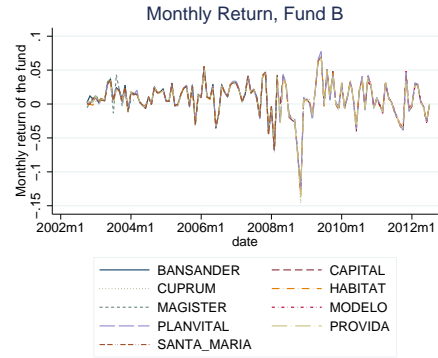
| Commission Comparison between Chile and the US | | |
|--|-----------------|--------------------------|
| PFA Load | Retirement Load | Equivalent Expense Ratio |
| 1.14% (Min) | 10.23% | 0.46% |
| 1.65% (Avg) | 14.16% | 0.66% |
| 2.36% (Max) | 19.09% | 0.92% |

Notes: This table maps commission rates under the Chilean system to expense ratios in the US, by comparing commissions paid for an individual who works for 40 years at a starting salary of US\$ 1000. It assumes said salary grows 2% annually in real terms and that returns in Chile and in the US are 5% annually in real terms. PFA Commission Rate is the percentage of wages that PFAs charge. Retirement Load is defined to be the ratio between the accumulated commissions paid at retirement, compounded using the real rate of return, and the sum of this variable plus the account balance at retirement. The Equivalent Expense Ratio is defined as the Expense Ratio that makes the US Retirement Load equal to the Chilean Retirement Load.

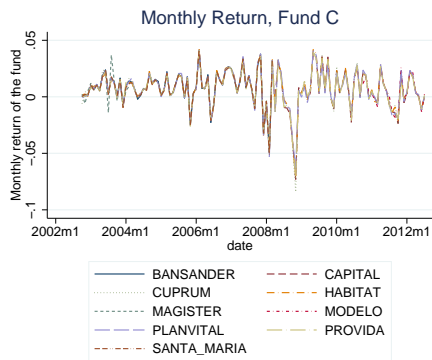
Table 11: Mapping Commissions to Expense Ratios



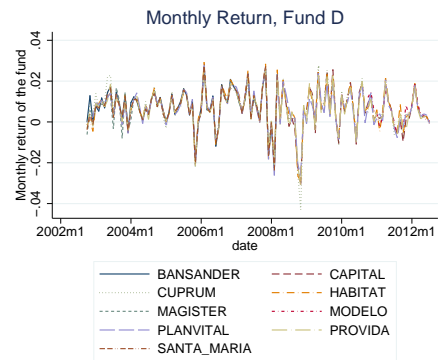
(a) Fund A



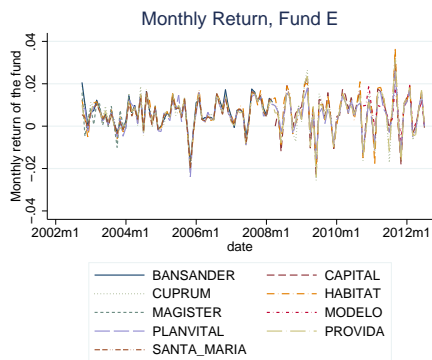
(b) Fund B



(c) Fund C



(d) Fund D



(e) Fund E

Figure 8: PFA Returns, by Fund

| Pairwise Correlations in Monthly Returns Across PFAs | | | | | | | | |
|--|-----------|---------|--------|---------|----------|--------|-----------|---------|
| Fund A | | | | | | | | |
| Firm | Bansander | Capital | Cuprum | Habitat | Magister | Modelo | Planvital | Provida |
| Bansander | 1 | | | | | | | |
| Capital | - | 1 | | | | | | |
| Cuprum | 0.9961 | 0.9911 | 1 | | | | | |
| Habitat | 0.9954 | 0.9919 | 0.9969 | 1 | | | | |
| Magister | 0.9396 | - | 0.9485 | 0.9365 | 1 | | | |
| Modelo | - | 0.9912 | 0.991 | 0.9919 | - | 1 | | |
| Planvital | 0.9946 | 0.9869 | 0.9942 | 0.9954 | 0.939 | 0.939 | 1 | |
| Provida | 0.9969 | 0.9901 | 0.9972 | 0.9964 | 0.9385 | 0.9385 | 0.9938 | 1 |
| Santa Maria | 0.9949 | - | 0.9919 | 0.9959 | 0.9094 | 0.9094 | 0.996 | 0.9892 |
| Fund B | | | | | | | | |
| Bansander | 1 | | | | | | | |
| Capital | - | 1 | | | | | | |
| Cuprum | 0.9943 | 0.9894 | 1 | | | | | |
| Habitat | 0.9921 | 0.9923 | 0.9957 | 1 | | | | |
| Magister | 0.8504 | - | 0.8734 | 0.8339 | 1 | | | |
| Modelo | - | 0.99 | 0.9911 | 0.9906 | - | 1 | | |
| Planvital | 0.9943 | 0.9887 | 0.9919 | 0.9948 | 0.8658 | 0.8658 | 1 | |
| Provida | 0.9962 | 0.9916 | 0.9964 | 0.9973 | 0.8443 | 0.8443 | 0.9954 | 1 |
| Santa Maria | 0.9957 | - | 0.9944 | 0.9938 | 0.8521 | 0.8521 | 0.9955 | 0.9951 |
| Fund C | | | | | | | | |
| Bansander | 1 | | | | | | | |
| Capital | - | 1 | | | | | | |
| Cuprum | 0.9904 | 0.987 | 1 | | | | | |
| Habitat | 0.9964 | 0.9903 | 0.9873 | 1 | | | | |
| Magister | 0.8267 | - | 0.7835 | 0.819 | 1 | | | |
| Modelo | - | 0.9611 | 0.9464 | 0.9506 | - | 1 | | |
| Planvital | 0.9968 | 0.9845 | 0.9804 | 0.9933 | 0.8405 | 0.8405 | 1 | |
| Provida | 0.9976 | 0.9915 | 0.9902 | 0.9944 | 0.8205 | 0.8205 | 0.9911 | 1 |
| Santa Maria | 0.9975 | - | 0.9893 | 0.9962 | 0.8464 | 0.8464 | 0.997 | 0.9967 |
| Fund D | | | | | | | | |
| Bansander | 1 | | | | | | | |
| Capital | - | 1 | | | | | | |
| Cuprum | 0.9604 | 0.9733 | 1 | | | | | |
| Habitat | 0.9756 | 0.9673 | 0.9539 | 1 | | | | |
| Magister | 0.8435 | - | 0.8554 | 0.8646 | 1 | | | |
| Modelo | - | 0.9208 | 0.9028 | 0.8635 | - | 1 | | |
| Planvital | 0.975 | 0.9527 | 0.9457 | 0.9711 | 0.9056 | 0.9056 | 1 | |
| Provida | 0.9816 | 0.9822 | 0.9678 | 0.9798 | 0.8561 | 0.8561 | 0.9662 | 1 |
| Santa Maria | 0.9863 | - | 0.9629 | 0.9891 | 0.8902 | 0.8902 | 0.9892 | 0.9916 |
| Fund E | | | | | | | | |
| Bansander | 1 | | | | | | | |
| Capital | - | 1 | | | | | | |
| Cuprum | 0.9598 | 0.9324 | 1 | | | | | |
| Habitat | 0.944 | 0.9466 | 0.9459 | 1 | | | | |
| Magister | 0.7177 | - | 0.7738 | 0.7571 | 1 | | | |
| Modelo | - | 0.7719 | 0.7625 | 0.7452 | - | 1 | | |
| Planvital | 0.9289 | 0.9481 | 0.936 | 0.9517 | 0.6654 | 0.6654 | 1 | |
| Provida | 0.9504 | 0.967 | 0.9652 | 0.9604 | 0.7268 | 0.7268 | 0.9578 | 1 |
| Santa Maria | 0.9381 | - | 0.9303 | 0.961 | 0.7122 | 0.7122 | 0.9626 | 0.9786 |

Notes: This table presents pair-wise correlations of monthly returns across PFAs, by fund, for the period between November 2002 and July 2012. Each pairwise correlation that includes a firm that exits the market (Bansander, Magister, Santa Maria) or a firm that enters the market (Capital, Modelo) considers only the overlapping timeframe of the pair.

Table 12: Correlation in Monthly Returns Across PFAs, by Fund

| Auction Bids for the Right to Serve New Consumers, 2010-2014 | | | | | | | | | |
|--|-------|----------|--------------|-------|----------|--------------|-------|----------|--------------|
| PFA | 2010 | | | 2012 | | | 2014 | | |
| | Bid | Bid Load | Current Load | Bid | Bid Load | Current Load | Bid | Bid Load | Current Load |
| Modelo | 1.14% | 10.23% | - | 0.77% | 7.15% | 10.23% | 0.72% | 6.72% | 7.15% |
| Planvital | 1.19% | 10.63% | 19.09% | 0.85% | 7.83% | 19.09% | 0.47% | 4.49% | 19.09% |
| Habitat | 1.21% | 10.79% | 11.97% | - | - | - | - | - | - |
| Cuprum | 1.32% | 11.66% | 12.89% | - | - | - | - | - | - |
| Regional | - | - | - | 1.04% | 9.42% | - | - | - | - |

Notes: This table reports results of the 2010, 2012 and 2014 auctions for the right to serve new consumers for their first two years in the system. Bid reports each company's bid as a percentage of income, Bid Load reports the loads these bids imply, and Current Load reports the load existing firms were actually charging at that time.

Table 13: Auction Bids, 2010-2014

| Correlation Between Returns, Commission Rates and Lagged Returns | | | | | | |
|--|---------------------|------------------|---------------------|-------------------|---------------------|----------------------|
| | Fund A | Fund B | Fund C | Fund D | Fund E | |
| All Funds | (2) | (3) | (4) | (5) | (6) | |
| Commission Rate | -0.040 (0.100) | 0.008 (0.045) | -0.015 (0.036) | -0.034 (0.034) | -0.065** (0.031) | -0.113*** (0.039) |
| Lagged Monthly Returns | 0.187*** (0.036) | 0.091 (0.057) | 3.39E-04 (0.101) | -0.094 (0.112) | 0.076 (0.057) | 0.018 (0.068) |
| Date Fixed Effect | x | x | x | x | x | x |
| Fund Fixed Effect | x | | | | | |
| Observations | 3430 | 686 | 686 | 686 | 686 | 686 |

Notes: This table reports estimates of the OLS regression of monthly returns on a one month lag of returns and on the commission rate charged by each PFA. Column 1 reports results including returns from all 5 funds, while columns 2-6 show fund specific results. All specifications include a date fixed effect. Robust standard errors are in parentheses.

* Significant at 10% ** Significant at 5% *** Significant at 1%

Table 14: Correlation Between Returns and Commission Rates

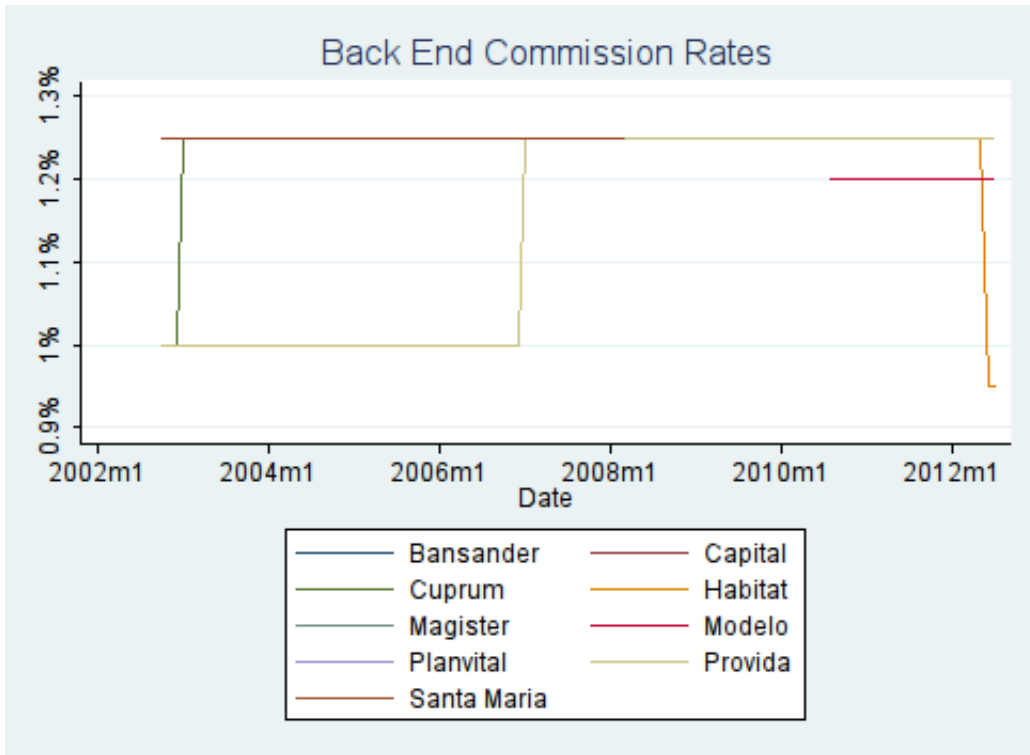


Figure 9: Back-end Commissions

E Derivations

If we observe $d_{i,t}^* = j$, $d_{i,t-1}^* = j'$, there are two relevant cases: comparing $d_{i,t}^* = j$ with $d'_{i,t} = j'$, or staying in the same firm, and comparing $d_{i,t}^* = j$ with $d'_{i,t} = j''$, or switching to a different firm. In the first case, we have that:

$$\begin{aligned} & u(j, j', X_{ijt}, \epsilon_{ijt}) - u(j', j', X_{ij't}, \epsilon_{ij't}) \\ & \geq \beta \cdot (E[V_{i,t+1}(j', B_{it}) | \Omega_{it}] - E[V_{i,t+1}(j, B_{it}) | \Omega_{it}]) \end{aligned} \quad (\text{E.1})$$

As before, we can bound the difference in continuation values using $\{j_\tau^*\}_{\tau=t+1}^{T_i}$:

$$\begin{aligned} & E[V_{i,t+1}(j', B_{it}) | \Omega_{it}] - E[V_{i,t+1}(j, B_{it}) | \Omega_{it}] \geq \\ & E[u(j_{t+1}^*, j', X_{ij,t+1}, \epsilon_{ij,t+1}) - u(j_{t+1}^*, j, X_{ij,t+1}, \epsilon_{ij,t+1}) | \Omega_{it}] \\ & + \beta^{T_i-t-1} \cdot E \left[B_{iT} \left(\{j_\tau^*\}_{\tau=t+1}^{T_i}, B_{i,t+1}(j') \right) - B_{iT} \left(\{j_\tau^*\}_{\tau=t+1}^{T_i}, B_{i,t+1}(j) \right) | \Omega_{it} \right] \\ & \geq -\gamma + \beta^{T_i-t-1} \cdot B_{it} \cdot E \left[(r_{j't} - r_{jt}) \prod_{\tau=t+1}^T (1 + r_{j^*\tau}) | \Omega_{it} \right] \end{aligned} \quad (\text{E.2})$$

Which means we can re-write E.1 as:

$$\begin{aligned} & \alpha w_{it} (p_{jt} - p_{j't}) + \delta(-1 + \beta) \\ & + \beta^{T_i-t} \cdot B_{it} \cdot E \left[(r_{j't} - r_{jt}) \prod_{\tau=t+1}^T (1 + r_{j^*\tau}) | \Omega_{it} \right] \\ & \geq \epsilon_{ij't} - \epsilon_{ijt} \end{aligned} \quad (\text{E.3})$$

In the second case, we get that:

$$\begin{aligned} & u(j, j', X_{ijt}, \epsilon_{ijt}) - u(j'', j', X_{ij't}, \epsilon_{ij't}) \\ & \geq \beta \cdot (E[V_{i,t+1}(j'', B_{it}) | \Omega_{it}] - E[V_{i,t+1}(j, B_{it}) | \Omega_{it}]) \end{aligned} \quad (\text{E.4})$$

Again, using the same $\{j_\tau^*\}_{\tau=t+1}^{T_i}$ as before, we get that:

$$\begin{aligned}
& E [V_{i,t+1} (j'', B_{it}) | \Omega_{it}] - E [V_{i,t+1} (j, B_{it}) | \Omega_{it}] \geq \\
& E [u (j_{t+1}^*, j'', X_{ij,t+1}, \epsilon_{ij,t+1}) - u (j_{t+1}^*, j, X_{ij,t+1}, \epsilon_{ij,t+1}) | \Omega_{it}] \\
& + \beta^{T_i-t-1} \cdot E \left[B_{iT} \left(\{j_\tau^*\}_{\tau=t+1}^{T_i}, B_{i,t+1} (j'') \right) - B_{iT} \left(\{j_\tau^*\}_{\tau=t+1}^{T_i}, B_{i,t+1} (j) \right) | \Omega_{it} \right] \\
& \geq -\delta + \beta^{T_i-t-1} \cdot B_{it} \cdot E \left[(r_{j''t} - r_{jt}) \prod_{\tau=t+1}^T (1 + r_{j^*\tau}) | \Omega_{it} \right]
\end{aligned} \tag{E.5}$$

Which means we can re-write E.1 as:

$$\begin{aligned}
& \alpha w_{it} (p_{jt} - p_{j't}) + \delta \beta \\
& + \beta^{T_i-t} \cdot B_{it} \cdot E \left[(r_{j't} - r_{jt}) \prod_{\tau=t+1}^T (1 + r_{j^*\tau}) | \Omega_{it} \right] \\
& \geq \epsilon_{ij't} - \epsilon_{ijt}
\end{aligned} \tag{E.6}$$

The problem is slightly different for individuals who are just entering the system (“newcomers”) and for individuals who are retiring (“retirees”). The former are solving the following maximization:

$$\begin{aligned}
E [V_{it} (\emptyset, B_{it}) | \Omega_{it}] & \equiv \max_{\{j_\tau \in \mathcal{J}_\tau\}_{\tau=t}^{T_i}} \left\{ u (j_t, \emptyset, X_{ij_t}, \epsilon_{ij_t}) + \sum_{\tau=t+1}^{T_i} \beta^{\tau-t} \cdot E [u (j_\tau, j_{\tau-1}, X_{ij_\tau}, \epsilon_{ij_\tau}) | \Omega_{it}] \right. \\
& \quad \left. + \beta^{T_i-t} \cdot E \left[B_{iT_i} \left(\{j_\tau\}_{\tau=t}^{T_i}, B_{it} \right) | \Omega_{it} \right] \right\} \\
& = \max_{j_t} u (j_t, \emptyset, X_{ij_t}, \epsilon_{ij_t}) + \beta \cdot E [V_{i,t+1} (j_t, B_{it}) | \Omega_{it}]
\end{aligned} \tag{E.7}$$

where $d_{i,t-1} = \emptyset$ denotes the fact that no previous choice has been made. If a newcomer picks firm j , we have that:

$$\begin{aligned}
& u (j, \emptyset, X_{ij_t}, \epsilon_{ij_t}) - u (j', \emptyset, X_{ij't}, \epsilon_{ij't}) \\
& \geq \beta \cdot (E [V_{i,t+1} (j', B_{it}) | \Omega_{it}] - E [V_{i,t+1} (j, B_{it}) | \Omega_{it}])
\end{aligned} \tag{E.8}$$

Note that the same bounding argument applies for the difference in continuation values, and we have that:

$$\begin{aligned}
& \alpha w_{it} (p_{jt} - p_{j't}) + \delta \beta \\
& + \beta^{T_i-t} \cdot B_{it} \cdot E \left[(r_{j't} - r_{jt}) \prod_{\tau=t+1}^T (1 + r_{j^*\tau}) \mid \Omega_{it} \right] \\
& \geq \epsilon_{ij't} - \epsilon_{ijt}
\end{aligned} \tag{E.9}$$

Finally, retirees face no dynamic implications of their choice beyond the final compounding of their account balances.

$$E [V_{i,T_i-1} (j, B_{i,T_i-1}) \mid \Omega_{i,T_i-1}] \equiv \max_{j_{T_i-1}} u (j_{T_i-1}, j, X_{ijT}, \epsilon_{ijt}) + \beta \cdot E [B_{iT_i} (j_{T_i-1}, B_{i,T_i-1}) \mid \Omega_{i,T_i-1}] \tag{E.10}$$

As a result, observing $d_{i,T_i-2}^* = j$ and $d_{i,T_i-1}^* = j$ implies that:

$$\begin{aligned}
& u (j, j, X_{ij,T_i-1}, \epsilon_{ij,T_i-1}) - u (j', j, X_{ij',T_i-1}, \epsilon_{ij',T_i-1}) \\
& \geq \beta (E [B_{iT_i} (j', B_{i,T_i-1}) \mid \Omega_{i,T_i-1}] - E [B_{iT_i} (j, B_{i,T_i-1}) \mid \Omega_{i,T_i-1}])
\end{aligned} \tag{E.11}$$

$$\begin{aligned}
& \alpha w_{i,T_i-1} (p_{j,T_i-1} - p_{j',T_i-1}) + \delta \\
& + \beta \cdot B_{i,T_i-1} \cdot E [r_j - r_{j'} \mid \Omega_{i,T_i-1}] \\
& \geq \epsilon_{ij't} - \epsilon_{ijt}
\end{aligned} \tag{E.12}$$

While observing $d_{i,T_i-2}^* = j'$ and $d_{i,T_i-1}^* = j$ implies that:

$$\begin{aligned}
& u (j, j', X_{ij,T_i-1}, \epsilon_{ij,T_i-1}) - u (j', j', X_{ij',T_i-1}, \epsilon_{ij',T_i-1}) \\
& \geq \beta (E [B_{iT_i} (j', B_{i,T_i-1}) \mid \Omega_{i,T_i-1}] - E [B_{iT_i} (j, B_{i,T_i-1}) \mid \Omega_{i,T_i-1}])
\end{aligned} \tag{E.13}$$

$$\begin{aligned}
& \alpha w_{i,T_i-1} (p_{j,T_i-1} - p_{j',T_i-1}) - \delta \\
& + \beta \cdot B_{i,T_i-1} \cdot E [r_j - r_{j'} \mid \Omega_{i,T_i-1}] \\
& \geq \epsilon_{ij't} - \epsilon_{ijt}
\end{aligned} \tag{E.14}$$

as well as:

$$\begin{aligned}
& u(j, j', X_{ij, T_i-1}, \epsilon_{ij, T_i-1}) - u(j'', j', X_{ij', T_i-1}, \epsilon_{ij', T_i-1}) \\
& \geq \beta (E[B_{iT_i}(j', B_{i, T_i-1}) | \Omega_{i, T_i-1}] - E[B_{iT_i}(j, B_{i, T_i-1}) | \Omega_{i, T_i-1}])
\end{aligned} \tag{E.15}$$

$$\begin{aligned}
& \alpha w_{i, T_i-1} (p_{j, T_i-1} - p_{j', T_i-1}) \\
& + \beta \cdot B_{i, T_i-1} \cdot E[r_j - r_{j'} | \Omega_{i, T_i-1}] \\
& \geq \epsilon_{ij't} - \epsilon_{ijt}
\end{aligned} \tag{E.16}$$

This defines all the cases in equation 3.12.

F Identification of the Switching Cost Parameter

This Appendix provides a formal argument for the identification of upper and lower bounds of the switching cost parameter in a simplified version of the model used in this paper. Assume that there are two firms, j and j' , in the market. Define $Y_{it} = \begin{cases} 1 & \text{if switch} \\ 0 & \text{otherwise} \end{cases}$. Let $X_{ijj't} \equiv w_{it} (p_{jt} - p_{j't})$, $\epsilon_{ijj't} \equiv \epsilon_{ijt} - \epsilon_{ij't}$, and ignore returns differences across firms. Then:

$$\begin{aligned}
Y_{it} = 0 | d_{i, t-1} = j & \Rightarrow -X_{ijj't} + \delta(1 + \beta) \geq -\epsilon_{ijj't} \\
Y_{it} = 0 | d_{i, t-1} = j' & \Rightarrow X_{ijj't} + \delta(1 + \beta) \geq \epsilon_{ijj't} \\
Y_{it} = 1 | d_{i, t-1} = j & \Rightarrow -X_{ijj't} + \delta(1 - \beta) \leq -\epsilon_{ijj't} \\
Y_{it} = 1 | d_{i, t-1} = j' & \Rightarrow X_{ijj't} + \delta(1 - \beta) \leq \epsilon_{ijj't}
\end{aligned}$$

Assume $\epsilon \sim F$. Then, dropping subscripts:

$$\begin{aligned}
\Pr[X < x, Y = 1 | j, z] &= \Pr[\epsilon < X - \delta(1 + \beta) < x - \delta(1 + \beta) | j, z] \\
&\leq \Pr[\epsilon < x - \delta(1 + \beta) | j, z] = F(x - \delta(1 + \beta) | j, z) \\
\Pr[X > x, Y = 0 | j, z] &= \Pr[\epsilon > X - \delta(1 - \beta) | j, z] \\
&\leq \Pr[\epsilon > x - \delta(1 - \beta) | j, z] = 1 - F(x - \delta(1 - \beta) | j, z) \\
\Pr[X > x, Y = 1 | j', z] &= \Pr[x + \delta(1 + \beta) < X + \delta(1 + \beta) < \epsilon | j', z] \\
&\leq \Pr[x + \delta(1 + \beta) < \epsilon | j', z] = 1 - F(x + \delta(1 + \beta) | j', z) \\
\Pr[X < x, Y = 0 | j', z] &= \Pr[x + \delta(1 - \beta) > X + \delta(1 - \beta) > \epsilon | j', z] \\
&\leq \Pr[x + \delta(1 - \beta) > \epsilon | j', z] = F(x + \delta(1 - \beta) | j', z)
\end{aligned}$$

Then:

$$\Pr[X - \delta(1 + \beta) < x, Y = 1|j, z] \leq F(x|j, z) \leq 1 - \Pr[X - \delta(1 - \beta) > x, Y = 0|j, z]$$

$$\Pr[X + \delta(1 - \beta) < x, Y = 0|j', z] \leq F(x|j', z) \leq 1 - \Pr[X + \delta(1 + \beta) > x, Y = 1|j', z]$$

Combining the two inequalities:

$$\begin{aligned} & \pi_j \Pr[X - \delta(1 + \beta) < x, Y = 1|j, z] + \pi_{j'} \Pr[X + \delta(1 - \beta) < x, Y = 0|j', z] \\ & \leq F(x|z) \leq \\ & 1 - \pi_j \Pr[X - \delta(1 - \beta) > x, Y = 0|j, z] - \pi_{j'} \Pr[X + \delta(1 + \beta) > x, Y = 1|j', z] \end{aligned}$$

Imposing independence ($\epsilon \perp\!\!\!\perp z$):

$$\begin{aligned} & \sup_z \{ \pi_j \Pr[X - \delta(1 + \beta) < x, Y = 1|j, z] + \pi_{j'} \Pr[X + \delta(1 - \beta) < x, Y = 0|j', z] \} \\ & \leq F(x) \leq \\ & \inf_z \{ 1 - \pi_j \Pr[X - \delta(1 - \beta) > x, Y = 0|j, z] - \pi_{j'} \Pr[X + \delta(1 + \beta) > x, Y = 1|j', z] \} \end{aligned}$$

To simplify notation, I will work with the following definitions:

$$\sup_z \Pr[Y = 1|j, z] = \Pr[Y = 1|j, z_j], \quad \sup_z \Pr[Y = 1|j', z] = \Pr[Y = 1|j', z_{j'}]$$

$$\bar{X}^{y,j} \equiv \max_X \{X : Y = y, j\}, \quad \underline{X}^{Y,j} \equiv \min_X \{X : Y = y, j\}$$

$$\bar{\delta} \equiv \inf_x \left\{ \max \left[\frac{\bar{X}^{1,j} - x}{1 + \beta}, \frac{\bar{X}^{0,j} - x}{1 - \beta}, \frac{x - \underline{X}^{0,j'}}{1 - \beta}, \frac{x - \underline{X}^{1,j'}}{1 + \beta} \right] \right\}$$

Claim 1. If $\pi_j \Pr[Y = 1|j, z_j] + \pi_{j'} \Pr[Y = 1|j', z_{j'}] = 1$ and either:

1. $\exists \delta^* > \max \left[\bar{\delta}, \frac{\underline{X}^{0,j'} - \underline{X}^{1,j'}}{2\beta} \right]$: for any $x' \in \left(\underline{X}^{0,j'} + \delta^*(1 - \beta), \underline{X}^{1,j'} + \delta^*(1 + \beta) \right)$,
 $\Pr[X + \delta^*(1 - \beta) < x', Y = 0|j', z_{j'}] > 0$.
2. $\exists \delta^* > \max \left[\bar{\delta}, \frac{\bar{X}^{1,j} - \bar{X}^{0,j}}{2\beta} \right]$: for any $x'' \in \left(\bar{X}^{1,j} - \delta^*(1 + \beta), \bar{X}^{0,j} - \delta^*(1 - \beta) \right)$,
 $\Pr[X - \delta^*(1 - \beta) > x'', Y = 0|j, z_j] > 0$.

then $\forall \delta > \delta^*$, δ can be rejected.

Proof. If: □

$$x + \delta(1 + \beta) > \bar{X}^{1,j} \ \& \ x - \delta(1 - \beta) < \underline{X}^{0,j'} \Rightarrow$$

$$\sup_z \{ \pi_j \Pr[X - \delta(1 + \beta) < x, Y = 1|j, z] + \pi_{j'} \Pr[X + \delta(1 - \beta) < x, Y = 0|j', z] \} = \pi_j \sup_z \Pr[Y = 1|j, z]$$

and

$$x + \delta(1 - \beta) > \bar{X}^{0,j} \& x - \delta(1 + \beta) < \underline{X}^{1,j'} \Rightarrow$$

$$\inf_z \{1 - \pi_j \Pr[X - \delta(1 - \beta) > x, Y = 0|j, z] - \pi_{j'} \Pr[X + \delta(1 + \beta) > x, Y = 1|j', z]\} = 1 - \pi_{j'} \sup_z \Pr[Y = 1|j', z]$$

Let $\bar{\delta} = \inf_x \left\{ \max \left[\frac{\bar{X}^{1,j} - x}{1 + \beta}, \frac{\bar{X}^{0,j} - x}{1 - \beta}, \frac{x - \underline{X}^{0,j'}}{1 - \beta}, \frac{x - \underline{X}^{1,j'}}{1 + \beta} \right] \right\}$. Then if $\delta > \bar{\delta}$, all four restrictions are met.

Given $\delta > \bar{\delta}$ and $\forall x \in \left(\max \{ \bar{X}^{1,j} - \delta(1 + \beta), \bar{X}^{0,j} - \delta(1 - \beta) \}, \min \{ \underline{X}^{0,j'} + \delta(1 - \beta), \underline{X}^{1,j'} + \delta(1 + \beta) \} \right)$

(non-empty because of the definition of $\bar{\delta}$), we have that:

$$\sup_z \{ \pi_j \Pr[X - \delta(1 + \beta) < x, Y = 1|j, z] + \pi_{j'} \Pr[X + \delta(1 - \beta) < x, Y = 0|j', z] \} = \pi_j \sup_z \Pr[Y = 1|j, z]$$

and:

$$\inf_z \{ 1 - \pi_j \Pr[X - \delta(1 - \beta) > x, Y = 0|j, z] - \pi_{j'} \Pr[X + \delta(1 + \beta) > x, Y = 1|j', z] \} = 1 - \pi_{j'} \sup_z \Pr[Y = 1|j', z]$$

Then for every $\delta > \bar{\delta}$, $\forall x \in \left(\max \{ \bar{X}^{1,j} - \delta(1 + \beta), \bar{X}^{0,j} - \delta(1 - \beta) \}, \min \{ \underline{X}^{0,j'} + \delta(1 - \beta), \underline{X}^{1,j'} + \delta(1 + \beta) \} \right)$

we have that $F(x) = \pi_j$ if $\pi_{j'} \sup_z \Pr[Y = 1|j', z] + \pi_j \sup_z \Pr[Y = 1|j, z] = 1$. Fix a $\delta_0 > \bar{\delta}$, and

consider x' such that:

$$\min \left\{ \underline{X}^{0,j'} + \delta_0(1 - \beta), \underline{X}^{1,j'} + \delta_0(1 + \beta) \right\} < x' < \max \left\{ \underline{X}^{0,j'} + \delta_0(1 - \beta), \underline{X}^{1,j'} + \delta_0(1 + \beta) \right\}$$

Then we have that:

$$\begin{aligned} & \sup_z \{ \pi_j \Pr[X - \delta_0(1 + \beta) < x', Y = 1|j, z] + \pi_{j'} \Pr[X + \delta_0(1 - \beta) < x', Y = 0|j', z] \} \\ & = \sup_z \{ \pi_j \Pr[Y = 1|j, z] + \pi_{j'} \Pr[X + \delta_0(1 - \beta) < x', Y = 0|j', z] \} \end{aligned}$$

and

$$\begin{aligned} & \inf_z \{ 1 - \pi_j \Pr[X - \delta_0(1 - \beta) > x', Y = 0|j, z] - \pi_{j'} \Pr[X + \delta_0(1 + \beta) > x', Y = 1|j', z] \} \\ & = 1 - \pi_{j'} \sup_z \Pr[X + \delta_0(1 + \beta) > x', Y = 1|j', z] \end{aligned}$$

If $\delta_0 > \frac{\underline{X}^{0,j'} - \underline{X}^{1,j'}}{2\beta}$, $\underline{X}^{0,j'} + \delta_0(1 - \beta) < x' < \underline{X}^{1,j'} + \delta_0(1 + \beta)$, and:

$$\sup_z \Pr[X + \delta_0(1 + \beta) > x', Y = 1|j', z] = \sup_z \Pr[Y = 1|j', z]$$

so that:

$$\begin{aligned} & \sup_z \{ \pi_j \Pr [Y = 1|j, z] + \pi_{j'} \Pr [X + \delta_0 (1 - \beta) < x', Y = 0|j', z] \} + \pi_{j'} \sup_z \Pr [X + \delta_0 (1 + \beta) > x', Y = 1|j', z] \\ & \geq \pi_j \Pr [Y = 1|j, z_j] + \pi_{j'} \Pr [X + \delta_0 (1 - \beta) < x', Y = 0|j', z_j] + \pi_{j'} \Pr [Y = 1|j', z_{j'}] \\ & = 1 + \pi_{j'} \Pr [X + \delta_0 (1 - \beta) < x', Y = 0|j', z_j] \end{aligned}$$

so if $\Pr [X + \delta_0 (1 - \beta) < x', Y = 0|j', z_j] > 0$, we can reject δ_0 at x' . We can also use this information to reject all $\delta > \delta_0$. To see this, define $x''(\delta) \equiv x' + (\delta - \delta_0)(1 - \beta)$, such that:

$$\Pr [X + \delta_0 (1 - \beta) < x', Y = 0|j', z_j] = \Pr [X + \delta (1 - \beta) < x''(\delta), Y = 0|j', z_j]$$

Then if $\min \{ \underline{X}^{0,j'} + \delta (1 - \beta), \underline{X}^{1,j'} + \delta (1 + \beta) \} < x''(\delta) < \max \{ \underline{X}^{0,j'} + \delta (1 - \beta), \underline{X}^{1,j'} + \delta (1 + \beta) \}$, we can reject δ at x'' . Since $\underline{X}^{0,j'} + \delta_0 (1 - \beta) < x' < \underline{X}^{1,j'} + \delta_0 (1 + \beta)$, we also have that:

$$\underline{X}^{0,j'} + \delta (1 - \beta) < x'' < \underline{X}^{1,j'} + 2(\delta_0 - \delta)\beta + \delta (1 + \beta) < \underline{X}^{1,j'} + \delta (1 + \beta)$$

Therefore, if for some $\delta^* > \max \left[\bar{\delta}, \frac{\underline{X}^{0,j'} - \underline{X}^{1,j'}}{2\beta} \right]$ there exists $x' \in \left(\underline{X}^{0,j'} + \delta^* (1 - \beta), \underline{X}^{1,j'} + \delta^* (1 + \beta) \right)$ such that $\Pr [X + \delta^* (1 - \beta) < x', Y = 0|j', z_j] > 0$, then we can reject $\delta \forall \delta > \delta^*$.

We can also obtain a similar condition looking at x 's below:

$$x \in \left(\max \{ \bar{X}^{1,j} - \delta (1 + \beta), \bar{X}^{0,j} - \delta (1 - \beta) \}, \min \{ \underline{X}^{0,j'} + \delta (1 - \beta), \underline{X}^{1,j'} + \delta (1 + \beta) \} \right)$$

Fix δ_0 as before, and consider x^* such that:

$$\min \{ \bar{X}^{1,j} - \delta_0 (1 + \beta), \bar{X}^{0,j} - \delta_0 (1 - \beta) \} < x^* < \max \{ \bar{X}^{1,j} - \delta_0 (1 + \beta), \bar{X}^{0,j} - \delta_0 (1 - \beta) \}$$

Then we have that:

$$\begin{aligned} & \sup_z \{ \pi_j \Pr [X - \delta_0 (1 + \beta) < x^*, Y = 1|j, z] + \pi_{j'} \Pr [X + \delta_0 (1 - \beta) < x^*, Y = 0|j', z] \} \\ & = \pi_j \sup_z \Pr [X - \delta_0 (1 + \beta) < x^*, Y = 1|j, z] \end{aligned}$$

$$\begin{aligned} & \inf_z \{ 1 - \pi_j \Pr [X - \delta_0 (1 - \beta) > x^*, Y = 0|j, z] - \pi_{j'} \Pr [X + \delta_0 (1 + \beta) > x^*, Y = 1|j', z] \} \\ & = 1 - \sup_z \{ \pi_j \Pr [X - \delta_0 (1 - \beta) > x^*, Y = 0|j, z] + \pi_{j'} \Pr [Y = 1|j', z] \} \end{aligned}$$

If $\frac{\bar{X}^{1,j} - \bar{X}^{0,j}}{2\beta} < \delta_0$, $\bar{X}^{1,j} - \delta_0(1 + \beta) < x^* < \bar{X}^{0,j} - \delta_0(1 - \beta)$, and:

$$\pi_j \sup_z \Pr[X - \delta_0(1 + \beta) < x^*, Y = 1|j, z] = \pi_j \sup_z \Pr[Y = 1|j, z]$$

And:

$$\begin{aligned} & \pi_j \sup_z \Pr[Y = 1|j, z] + \sup_z \{ \pi_j \Pr[X - \delta_0(1 - \beta) > x^*, Y = 0|j, z] + \pi_{j'} \Pr[Y = 1|j', z] \} \\ & \geq \pi_j \Pr[Y = 1|j, z_j] + \pi_j \Pr[X - \delta_0(1 - \beta) > x^*, Y = 0|j, z_j] + \pi_{j'} \Pr[Y = 1|j', z_{j'}] \\ & = \pi_j \Pr[X - \delta_0(1 - \beta) > x^*, Y = 0|j, z_{j'}] + 1 \end{aligned}$$

So if $\Pr[X - \delta_0(1 - \beta) > x^*, Y = 0|j, z_{j'}] > 0$, we can reject δ_0 at x^* . We can also use this information to reject all $\delta > \delta_0$. To see this, define $x^{**}(\delta) \equiv x^* - (\delta - \delta_0)(1 - \beta)$, such that

$$\Pr[X - \delta_0(1 - \beta) > x^*, Y = 0|j, z_{j'}] = \Pr[X - \delta(1 - \beta) > x^{**}(\delta), Y = 0|j, z_{j'}]$$

Then if $\min \{ \bar{X}^{1,j} - \delta(1 + \beta), \bar{X}^{0,j} - \delta(1 - \beta) \} < x^{**}(\delta) < \max \{ \bar{X}^{1,j} - \delta(1 + \beta), \bar{X}^{0,j} - \delta(1 - \beta) \}$, we can reject δ at $x^{**}(\delta)$. Since $\bar{X}^{1,j} - \delta_0(1 + \beta) < x^* < \bar{X}^{0,j} - \delta_0(1 - \beta)$, we also have that:

$$\bar{X}^{1,j} - \delta(1 + \beta) < \bar{X}^{1,j} - \delta(1 + \beta) + 2\beta(\delta - \delta_0) < x^{**}(\delta) < \bar{X}^{0,j} - \delta(1 - \beta)$$

Therefore, if for some $\delta^* > \max \left[\bar{\delta}, \frac{\bar{X}^{1,j} - \bar{X}^{0,j}}{2\beta} \right]$ there exists $\bar{X}^{1,j} - \delta_0(1 + \beta) < x^* < \bar{X}^{0,j} - \delta_0(1 - \beta)$ such that $\Pr[X - \delta^*(1 - \beta) > x^*, Y = 0|j, z_{j'}] > 0$, then we can reject $\delta \forall \delta > \delta^*$.

The following figure presents a graphical representation of these arguments:

The red area indicates the set of (δ, x) for which $F(x) = \pi_j$ if $\pi_{j'} \sup_z \Pr[Y = 1|j', z] + \pi_j \sup_z \Pr[Y = 1|j, z] = 1$. The blue area indicates the set of (δ, x) that are above $\bar{\delta}$ and that meet the restrictions that define x' (right side) and x^* (left side). In this example, $\bar{\delta} > \max \left[\frac{\bar{X}^{1,j} - \bar{X}^{0,j}}{2\beta}, \frac{\underline{X}^{0,j'} - \underline{X}^{1,j'}}{2\beta} \right]$, and these constraints on δ^* are ignored. Furthermore, at (δ^*, x^*) we have that $\Pr[X - \delta^*(1 - \beta) > x^*, Y = 0|j, z_{j'}] > 0$, and as a result we can reject δ^* and all $\delta > \delta^*$ by following the ray $x^{**}(\delta)$. The graph also depicts an x' such that $\Pr[X + \delta'(1 - \beta) < x', Y = 0|j', z_j] > 0$, and the accompanying ray $x''(\delta)$.

The proof for establishing a lower bound of the switching cost parameter is analogous.

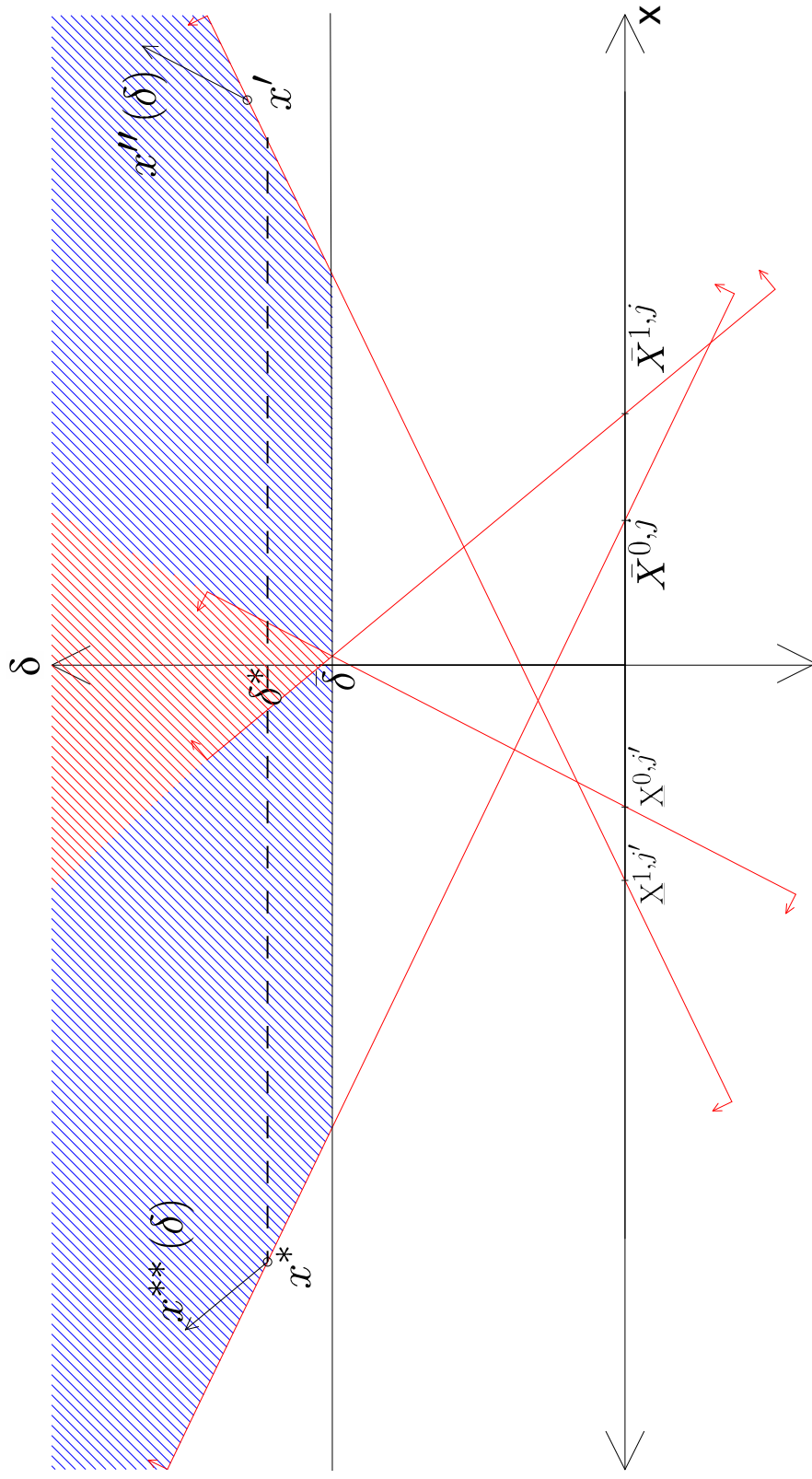


Figure 10: Identification of an upper bound on the switching cost parameter