Essays on Strategic Behavior in Government-Designed Markets

by

Gastón Illanes

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2016

Author .... Signature redacted
Department of Economics
February 12, 2016

Certified by .... Signature redacted
Nancy L. Rose
Charles P. Kindleberger Professor of Applied Economics
Thesis Supervisor

Certified by .... Signature redacted
Paulo Somaini
Assistant Professor of Economics
Stanford University Graduate School of Business
Thesis Supervisor

Accepted by .... Signature redacted
Ricardo Caballero
Ford International Professor of Economics
Chairman, Department Committee on Graduate Theses

Essays on Strategic Behavior in Government-Designed Markets

by

Gastón Illanes

Submitted to the Department of Economics
on February 12, 2016, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis studies consumer behavior and strategic interactions between firms in markets that were actively designed by governments. In such settings, government intervention is frequent, firms are often constrained in their actions, and consumer behavior may depart from what is predicted by the standard set of assumptions. The chapters of this thesis study how the current set of regulations is affecting market outcomes in different settings and what can be done to improve them.

Chapter 1 studies the Chilean pension system, where workers' mandatory contributions are administered by private companies. This market exhibits fee dispersion and low switching rates, which could be explained by firm differentiation or by switching costs. Using a novel combination of revealed preference inequalities and latent variable integration techniques, I find evidence of large switching costs, and that if these costs did not exist prices would fall to around one-half of currently observed levels.

Chapter 2 is a pre-cursor to Chapter 1, studying what would be learned from estimating demand in this market using a more standard set of techniques. I find that ignoring switching costs, individual-level heterogeneity, and endogeneity will lead to implausible demand estimates. These results are the key motivation for the use of the more sophisticated methods used in Chapter 1.

Finally, Chapter 3, written with Sarah Moshary, studies the privatization of liquor sales in Washington state. It focuses on a natural experiment induced by privatization, which creates exogenous variation in the number of elegible licensees in local liquor markets, generated by a licensure threshold requirement on store size: only stores larger than 10,000 square feet are allowed to sell liquor. We find that this regulation does not alter the total number of liquor outlets within each market. Instead, it shifts the composition of stores. Also, we find that in markets with an additional potential entrant the product mix is shifted towards cheaper products. This confirms concerns that competition in liquor markets leads to greater availability of cheap alcohol, and suggests that regulation has an effect in limiting the availability of those types of products.

Thesis Supervisor: Nancy L. Rose
Title: Charles P. Kindleberger Professor of Applied Economics

Contents

List of Figures

List of Tables

Acknowledgements

Successfully completing a Ph.D. program requires many sacrifices. No one sacrificed more for me to be able to achieve this goal than my wife, Carolina Alfero. Although my name is on the title, the cost was shared by the both of us, and I will never forget that. For this, for your support, and for being partners in this great adventure, I am forever grateful.

My advisors have been extremely generous with their time and effort, and their support has greatly increased the quality of my work. Paulo Somaini has been instrumental in challenging me, in working through problems with me, and in general in supporting me throughout this whole process. Without his guidance, my work would have been significantly poorer, and he will be the model for my future advising efforts. I hope I will be able to be as insightful and generous an advisor as he is. Nancy Rose has been a wonderful mentor, and I never left a meeting with her without learning something new about my own work. Her support through the most difficult times of this experience kept me going, and even when she was not at MIT, I always felt that she was close by. Finally, I met Victor Chernozhukov late in this process, but hearing that he liked my work was the greatest confidence boost I could have ever had. His willingness to discuss small empirical details of my job market paper astounded me, as well as his practical suggestions for solving problems that troubled me for weeks. I wish that I had come up with a project that required his help sooner.

I would also like to thank Michael Whinston, Panle Jia Barwick and Isaiah Andrews, who are not formally my advisors but were instrumental in shaping my work. Mike has dived into the depths of my job market paper with gusto, and in the process has provided critical suggestions that have made it significantly better. I am in awe of his approach towards understanding empirical work. Panle believed in my ideas from visiting days on, took me under her tutelage, and was my biggest influence during my first two years in this program. I cannot thank her enough for her impact on my training. Finally, Isaiah has helped guide me through a thicket of econometric challenges, and his ability to propose solutions to the problems I've faced and to warn me of issues I hadn't seen has been invaluable.

Many members of the MIT faculty have also helped shape my dissertation, with their helpful comments in seminars, lunches and one-on-one meetings. I would like to thank Nikhil Agarwal, Peter Diamond, Glenn Ellison, Sara Fisher Ellison, Alfred Galichon, Jon Gruber, Elena Manresa, Anna Mikusheva, Parag Pathak, James Poterba and Frank Schilbach for their contributions to my work. Outside of MIT, I would like to thank and honor the memory of Che-Lin Su, a wonderful teacher and researcher who will be deeply missed.

I cannot say enough good things about the students I've interacted with during my time at MIT. I'd like to thank my first year crew, Sarah Moshary, Ashish Shenoy and Sally Hudson. I am truly blessed to have you as my friends. I'd also like to thank Manisha Padi, who has had to suffer through trying to co-author a project with me during my time on the market. I am grateful to all the students in the IO group, particularly Vivek Bhattacharya, Dong Jae Eun, Brad Larsen, Maria Polyakova, Brad Shapiro and Daniel Waldinger. Finally, I'd like to thank my classmates Manasi Deshpande, Kathleen Easterbrook, Daan Struyven and Melanie Wasserman for their helpful suggestions regarding my job talk.

Many individuals helped me acquire the data I needed to succesfully complete these projects. I'd like to thank Augusto Iglesias, who opened doors when I had given up on ever opening them. I'd also like to thank Solange Bernstein, Olga Fuentes, Eugenio Salvo and Ximena Quintanilla at the Superintendencia de Pensiones, as well as Carlos Alvarado and Jorge Mastrangelo at the Superintendencia de Valores y Seguros.

I would like to thank my professors at Pontificia Universidad Católica de Chile. They say good teachers change students' lives, and that was certainly the case for me during my years there. In particular, I would like to thank Francisco Gallego, Juan Pablo Montero, and Claudio Sapelli, who mentored me and inspired me to become an academic.

I would like to thank my son, Santiago Illanes, who brought joy into my life when it was darkest; my grandfather, Sergio Orellana, for nurturing my academic curiosity; my grandmother, Josephine Johnson, for her unconditional love; my second mother, Maria Figueroa, for teaching me that the world was full of people living tougher lives than I had ever imagined, which ended up fueling my passion for economics; and finally my mother, Ana Maria Orellana, for being a mom and a dad and for more that I could ever put into words.

Chapter 1

Switching Costs in Pension Plan Choice

1.1 Introduction

[1]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,

[1]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 Economia 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.

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 system[2], 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]).

[2] $Load = \frac{Commission\ Paid}{Commission\ Paid + Mandatory\ Contribution}$

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 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 2002[3], so we can conclude that participants in the system are choosing firms as if returns differences will have 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

[3]For 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.

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])[4]. 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 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.

1.2 Data and Descriptive Evidence

1.2.1 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

[4]The 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.

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 all[5]. 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 allowed[6], 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:

$$Load = \frac{Monthly\,Commission\,Paid}{Monthly\,Commission\,Paid + Monthly\,Mandatory\,Contribution} \tag{1.1}$$

Figure 1.1 plots loads from 2002 through 2012, while Figure 1.2 plots commission rates for the same period. The main stylized fact to be obtained from these plots is that commission rates are disperse, 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 1.3). 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

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

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

to view pricing differences in a different light if they prices 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 1.4 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 1.1 shows a summary of investment rules for each fund (OECD [2012]). Workers are free to choose between funds[7], 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

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

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 1.4, the PFA market had over 20 active firms during the 90's. Some of these firms ceased to exist by merging with other companies, 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 criterion[8] , 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 1.2 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 1.2 studies this issue by comparing the probability that a customer of the absorbed firm chooses

[8] Planvital 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.

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.

1.2.2 Data and Descriptive Evidence

This paper works with an administrative database made available by the Chilean Pension Superintendency[9], 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 1.3 shows some key indicators for the sample. An important feature of the

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

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 1.3. 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 1.4 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 1.5 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 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 1.5 present predicted switching probabilities obtained from the following linear probability model:

$$1\left[Switch_{it}\right] = b_{age}\left(age_{it}\right) + b_{wage}\left(wage_{it}\right) + b_{bal}\left(balance_{it}\right) + \epsilon_{it} \tag{1.2}$$

where $b_{var}\left(\cdot\right)$ 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 1.6 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 1.7 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 1.8 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 period[10]. The row titled "PDV Commission Savings" reports the results obtained using a

[10]With 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

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 calculation[11] 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 the subject of the following section.

1.3 Model

This section introduces a model of pension fund administation 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

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.

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

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 $u\left(d_{it}, d_{i,t-1}, X_{ijt}, \epsilon_{ijt}\right)$, 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 $B_{i,t+1} = B_{it}\left(1 + r_{d_{it}t}\right) + c_{i,t+1}$. They update they beliefs and form $\Omega_{i,t+1}$.

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

$$\begin{aligned}E\left[V_{it}\left(d_{i,t-1},B_{it}\right)|\Omega_{it}\right]\equiv\quad & \max_{\{j_\tau\in\mathcal{J}_\tau\}_{\tau=t}^{T_i}}\left\{u\left(j_t,d_{i,t-1},X_{ijt},\epsilon_{ijt}\right)\right.\\ & +\sum_{\tau=t+1}^{T_i}\beta^{\tau-t}\cdot E\left[u\left(j_\tau,j_{\tau-1},X_{ij\tau},\epsilon_{ij\tau}\right)|\Omega_{it}\right]\\ & \left.+\beta^{T_i-t}\cdot E\left[B_{iT_i}\left(\{j_\tau\}_{\tau=t}^{T},B_{it}\right)|\Omega_{it}\right]\right\}\\ & =\max_{j_t}\ u\left(j_t,d_{t-1},X_{ijt},\epsilon_{ijt}\right)+\beta\cdot E\left[V_{i,t+1}\left(j_t,B_{i,t+1}\right)|\Omega_{it}\right]\end{aligned} \tag{1.3}$$

where $E\left[V_{it}\left(d_{i,t-1},B_{it}\right)|\Omega_{it}\right]$ 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 $E\left[B_{iT_i}\left(\{j_\tau\}_{\tau=t}^{T},B_{it}\right)|\Omega_{it}\right]$ 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 (Ackerberg [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 $\left(d^*_{i,t} = d^*_{i,t-1} = j\right)$. Then:

$$\begin{aligned} E\left[V_{it}\left(j, B_{it}\right)|\Omega_{it}\right] &= u\left(j, j, X_{ijt}, \epsilon_{ijt}\right) + \beta \cdot E\left[V_{i,t+1}\left(j, B_{i,t+1}\right)|\Omega_{it}\right] \\ &\geq u\left(j', j, X_{ij't}, \epsilon_{ij't}\right) + \beta \cdot E\left[V_{i,t+1}\left(j', B_{i,t+1}\right)|\Omega_{it}\right] \end{aligned} \tag{1.4}$$

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

$$\begin{aligned} & u\left(j, j, X_{ijt}, \epsilon_{ijt}\right) - u\left(j', j, X_{ij't}, \epsilon_{ij't}\right) \\ \geq\ & \beta \cdot \left(E\left[V_{i,t+1}\left(j', B_{i,t+1}\right)|\Omega_{it}\right] - E\left[V_{i,t+1}\left(j, B_{i,t+1}\right)|\Omega_{it}\right]\right) \end{aligned} \tag{1.5}$$

One can use revealed preference arguments to find a lower bound of the terms in the right hand side of equation 1.5. To do so, it is useful to define $\{j^*_\tau\}_{\tau=t+1}^{T_i}$ as the sequence of choices that attains

$E\left[V_{i,t+1}\left(j, B_{i,t+1}\right)|\Omega_{it}\right]$:

$$\begin{aligned}\{j^*_\tau\}_{\tau=t+1}^{T_i} \equiv \quad & \arg\max_{\{j_\tau \in \mathcal{J}_\tau\}_{\tau=t+1}^{T_i}} \Big\{ E\left[u\left(j_{t+1}, j, X_{ij,t+1}, \epsilon_{ij,t+1}\right)|\Omega_{it}\right] \\ & + \textstyle\sum_{\tau=t+2}^{T_i} \beta^{\tau-t-1} \cdot E\left[u\left(j_\tau, j_{\tau-1}, X_{ij\tau}, \epsilon_{ij\tau}\right)|\Omega_{it}\right] \\ & + \beta^{T_i-t-1} \cdot E\left[B_{iT_i}\left(\{j_\tau\}_{\tau=t+1}^{T}, B_{it}\right)|\Omega_{it}\right]\Big\}\end{aligned} \tag{1.6}$$

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:

$$\begin{aligned}E\left[V_{i,t+1}\left(j', B_{i,t+1}\right)|\Omega_{it}\right] \geq \quad & E\left[u\left(j^*_{t+1}, j', X_{ij,t+1}, \epsilon_{ij,t+1}\right)|\Omega_{it}\right] \\ & + \textstyle\sum_{\tau=t+2}^{T_i} \beta^{\tau-t-1} \cdot E\left[u\left(j^*_\tau, j^*_{\tau-1}, X_{ij\tau}, \epsilon_{ij\tau}\right)|\Omega_{it}\right] \\ & + \beta^{T_i-t-1} \cdot E\left[B_{iT}\left(\{j^*_\tau\}_{\tau=t+1}^{T_i}, B_{i,t+1}\left(j'\right)\right)|\Omega_{it}\right]\end{aligned} \tag{1.7}$$

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 $E\left[V_{i,t+1}\left(j, B_{it}\right)|\Omega_{it}\right]$. This allows us to bound the difference in continuation values in the right hand side of equation 1.5 by:

$$\begin{aligned}E\left[V_{i,t+1}\left(j', B_{i,t+1}\right)|\Omega_{it}\right] - E\left[V_{i,t+1}\left(j, B_{i,t+1}\right)|\Omega_{it}\right] \geq \\ E\left[u\left(j^*_{t+1}, j', X_{ij,t+1}, \epsilon_{ij,t+1}\right) - u\left(j^*_{t+1}, j, X_{ij,t+1}, \epsilon_{ij,t+1}\right)|\Omega_{it}\right] \\ + \beta^{T_i-t-1} \cdot E\left[B_{iT}\left(\{j^*_\tau\}_{\tau=t+1}^{T_i}, B_{i,t+1}\left(j'\right)\right) - B_{iT}\left(\{j^*_\tau\}_{\tau=t+1}^{T_i}, B_{i,t+1}\left(j\right)\right)|\Omega_{it}\right]\end{aligned} \tag{1.8}$$

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 1.8 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 $u\left(j_t, d_{t-1}, X_{ijt}, \epsilon_{ijt}\right) = -\delta \cdot 1\left[j_t \neq d_{t-1}\right] + \alpha w_{it} p_{jt} + \epsilon_{ijt}$ 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:

$$\begin{gathered} E\left[V_{i,t+1}\left(j', B_{i,t+1}\right)|\Omega_{it}\right] - E\left[V_{i,t+1}\left(j, B_{i,t+1}\right)|\Omega_{it}\right] \geq \\ -\delta + \beta^{T_i-t-1} \cdot E\left[B_{iT}\left(\{j^*_\tau\}^{T_i}_{\tau=t+1}, B_{i,t+1}\left(j'\right)\right) - B_{iT}\left(\{j^*_\tau\}^{T_i}_{\tau=t+1}, B_{i,t+1}\left(j\right)\right)|\Omega_{it}\right] \end{gathered} \tag{1.9}$$

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

$$E\left[B_{iT}\left(\{j_\tau\}^{T}_{\tau=t+1}, B_{i,t+1}\right)|\Omega_{it}\right] \equiv E\left[\prod_{\tau=t+1}^{T}\left(1+r_{j\tau}\right)B_{i,t+1} + \sum_{k=t+2}^{T}\prod_{\tau=k}^{T}\left(1+r_{j\tau}\right)c_{ik}|\Omega_{it}\right] \tag{1.10}$$

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 $B_{i,t+1}\left(j'\right) = B_{it}\left(1+r_{j't}\right) + c_{i,t+1}$ or $B_{i,t+1}\left(j\right) = B_{it}\left(1+r_{jt}\right) + c_{i,t+1}$, we have that:

$$\begin{aligned} &E\left[V_{i,t+1}\left(j', B_{i,t+1}\right)|\Omega_{it}\right] - E\left[V_{i,t+1}\left(j, B_{i,t+1}\right)|\Omega_{it}\right] \\ &\geq -\delta + \beta^{T_i-t-1} \cdot E\left[\left(B_{i,t+1}\left(j'\right) - B_{i,t+1}\left(j\right)\right)\textstyle\prod_{\tau=t+1}^{T_i}\left(1+r_{j^*\tau}\right)|\Omega_{it}\right] \\ &= -\delta + \beta^{T_i-t-1} \cdot B_{it} \cdot E\left[\left(r_{j't} - r_{jt}\right)\textstyle\prod_{\tau=t+1}^{T_i}\left(1+r_{j^*\tau}\right)|\Omega_{it}\right] \end{aligned} \tag{1.11}$$

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 1.5, which gives us that:

$$\begin{gathered} \alpha w_{it}\left(p_{jt} - p_{j't}\right) + \delta\left(1+\beta\right) \\ +\beta^{T_i-t} \cdot B_{it} \cdot E\left[\left(r_{jt} - r_{j't}\right)\textstyle\prod_{\tau=t+1}^{T_i}\left(1+r_{j^*\tau}\right)|\Omega_{it}\right] \\ \geq \epsilon_{ij't} - \epsilon_{ijt} \end{gathered} \tag{1.12}$$

Recall that this derivation assumed that the individual made the same choice in two consecutive periods, $d^*_{i,t} \doteq d^*_{i,t-1} = j$. 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:

$$\begin{gathered}\alpha w_{it}\left(p_{jt}-p_{j't}\right)+\delta\cdot\Delta\left(d_{it},d_{i,t-1},d'_{it},\beta,T_i\right)\\ +\beta^{T_i-t}\cdot B_{it}\cdot E\left[\left(r_{jt}-r_{j't}\right)\prod_{\tau=t+1}^{T_i}\left(1+r_{j^*\tau}\right)|\Omega_{it}\right]\\ \geq \epsilon_{ij't}-\epsilon_{ijt}\end{gathered} \tag{1.13}$$

where:

$$\Delta\left(d_{it},d_{i,t-1},d'_{it},\beta\right)=\begin{cases}1+\beta \text{ if } d_{it}=d_{i,t-1}\\ -1+\beta \text{ if } d_{it}\neq d_{i,t-1} \text{ and } d'_{it}=d_{i,t-1}\\ \beta \text{ if } \left(d_{it}\neq d_{i,t-1} \text{ and } d'_{it}\neq d_{i,t-1}\right) \text{ or } d_{i,t-1}=\emptyset\\ 1 \text{ if } t+1=T_i \text{ and } d_{it}=d_{i,t-1}\\ -1 \text{ if } t+1=T_i \text{ and } d_{it}\neq d_{i,t-1} \text{ and } d'_{it}=d_{i,t-1}\\ 0 \text{ if } t+1=T_i \text{ and } d_{it}\neq d_{i,t-1} \text{ and } d'_{it}\neq d_{i,t-1}\end{cases} \tag{1.14}$$

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 1.13 to the data, one needs to take a stand on dealing with expected future returns $E\left[\left(r_{jt}-r_{j't}\right)\prod_{\tau=t+1}^{T_i}\left(1+r_{j^*\tau}\right)|\Omega_{it}\right]$. This work replaces this expression with $\left(r_{jt}-r_{j't}\right)\left(1+\omega\right)^{T_i-t-1}$, where ω is a parameter to be estimated representing the mean per-period expected return. Replacing, the revealed preference inequality in equation 1.13 becomes:

$$\begin{gathered}\alpha w_{it}\left(p_{jt}-p_{j't}\right)+\delta\cdot\Delta\left(d_{it},d_{i,t-1},d'_{it},\beta\right)\\ +\beta^{T_i-t}\cdot B_{it}\cdot\left(r_{jt}-r_{j't}\right)\cdot\left(1+\omega\right)^{T_i-t-1}\\ \geq \epsilon_{ij't}-\epsilon_{ijt}+\eta_{ij't}-\eta_{ijt}\end{gathered} \tag{1.15}$$

where

$$\eta_{ijt}=\beta^{T_i-t}\cdot B_{it}\cdot\left(E\left[r_{jt}\cdot\prod_{\tau=t+1}^{T_i}\left(1+r_{j^*\tau}\right)|\Omega_{it}\right]-r_{jt}\cdot\left(1+\omega\right)^{T_i-t-1}\right) \tag{1.16}$$

This approach implies parameterizing the beliefs regarding future returns under path $\{j^*_\tau\}_{\tau=t+1}^{T_i}$ as $\left(1+\omega\right)^{T_i-t-1}$, 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 $E\left[(r_{jt} - r_{j't})\,|\Omega_{it}\right]$ 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 substituted into $E\left[(r_{jt} - r_{j't})\,|\Omega_{it}\right]$. 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.

1.4 Estimation Strategies

This section discusses alternative strategies for taking equation 1.15 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

$E\left[\epsilon_{ij't} - \epsilon_{ijt} + \eta_{ij't} - \eta_{ijt}\right] = 0$, 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 $E\left[\epsilon_{ij't} - \epsilon_{ijt}\right] < 0$. 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:
   $E\left[z_{ijt}\left(\alpha w_{it}\left(p_{jt} - p_{j't}\right) + \delta \cdot \Delta\left(d_{it}, d_{i,t-1}, d'_{it}, \beta\right) + \beta^{T_i - t} \cdot B_{it} \cdot \left(r_{jt} - r_{j't}\right) \cdot (1+\omega)^{T-t-1}\right)\right] \neq 0$
3. Exclusion Restriction:
   $E\left[z_{ijt}\left(\zeta_{ij't} - \zeta_{ijt}\right)\right] \geq 0$

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:

$$\begin{aligned} E\left[z_{ijt}\left(\alpha w_{it}\left(p_{jt} - p_{j't}\right) + \delta \cdot \Delta\left(d_{it}, d_{i,t-1}, d'_{it}, \beta\right) + \beta^{T_i - t} \cdot B_{it} \cdot \left(r_{jt} - r_{j't}\right) \cdot (1+\omega)^{T-t-1}\right)\right] \\ \geq E\left[z_{ijt}\left(\zeta_{ij't} - \zeta_{ijt}\right)\right] \geq 0 \end{aligned} \tag{1.17}$$

and simply work with the moment:

$$E\left[z_{ijt}\left(\alpha w_{it}\left(p_{jt} - p_{j't}\right) + \delta \cdot \Delta\left(d_{it}, d_{i,t-1}, d'_{it}, \beta\right) + \beta^{T_i - t} \cdot B_{it} \cdot \left(r_{jt} - r_{j't}\right) \cdot (1+\omega)^{T-t-1}\right)\right] \geq 0 \tag{1.18}$$

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 $cov\left(z_{ijt}, \epsilon_{ijt} | d^*_{it} = j\right) = 0$. Since,

$$E\left[z_{ijt}\epsilon_{ijt}|d^*_{it} = j\right] = cov\left(z_{ijt}, \epsilon_{ijt}|d^*_{it} = j\right) + E\left[z_{ijt}|d^*_{it} = j\right] \cdot E\left[\epsilon_{ijt}|d^*_{it} = j\right] \tag{1.19}$$

the fact that $E\left[z_{ijt}\right] > 0$ (due to the no sign changes condition) and $E\left[\epsilon_{ijt}\right] > 0$ (due to selection) implies that $E\left[z_{ijt}\epsilon_{ijt}|d^*_{it} = j\right] > 0$, 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 1.15 for individuals who are in the same bin but who make different choices. Under this strategy, equation 1.15 becomes:

$$\begin{aligned} &\alpha\left(w_{it} - w_{i't}\right)\left(p_{jt} - p_{j't}\right) + \\ &+\delta \cdot \left(\Delta\left(d_{it}, d_{i,t-1}, d'_{it}, \beta\right) - \Delta\left(d_{i't}, d_{i',t-1}, d'_{i't}, \beta\right)\right) \\ &+\beta^{T_i - t}\left[B_{it} - \beta^{T_{i'} - T_i} \cdot (1+\omega)^{T_{i'} - T_i} \cdot B_{i't}\right]\left(r_{jt} - r_{j't}\right)(1+\omega)^{T_i - t - 1} \\ &\geq \zeta_{ij't} - \zeta_{i'j't} - \zeta_{ijt} + \zeta_{i'jt} \end{aligned} \tag{1.20}$$

If $E\left[\zeta_{ij't} - \zeta_{i'j't} - \zeta_{ijt} + \zeta_{i'jt}\right] \geq 0$, or $E\left[z_{ii'jj't}\left(\zeta_{ij't} - \zeta_{i'j't} - \zeta_{ijt} + \zeta_{i'jt}\right)\right] \geq 0$[12], then this strategy will appropriately control for selection bias. In general, this will not be the case, as one would expect $E\left[\epsilon_{ijt}\right] > E\left[\epsilon_{i'jt}\right]$ and $E\left[\epsilon_{i'j't}\right] \geq E\left[\epsilon_{ij't}\right]$, 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 $E\left[\epsilon_{ijt}|d^*_{it} = j\right]$, which requires solving a dynamic choice problem. Since this is precisely the difficulty that moment inequalities was trying to avoid, in this setting these

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

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:

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

Note that determining whether a given θ is in the identified set requires searching over the space of conditional distributions of the unobservable $\mathcal{P}_{\mathcal{U}|\mathcal{Z}}$ 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 $\pi \in \mathcal{P}_{\mathcal{Z}}$,

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

if and only if

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

where

$$\tilde{g}(Z,\theta,\gamma) \equiv \frac{\int g(u,z,\theta)\exp(\gamma' g(u,z,\theta))\, d\rho(u|z;\theta)}{\int \exp(\gamma' g(u,z,\theta))\, d\rho(u|z;\theta)} \tag{1.21}$$

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. $supp\, \rho(\cdot|z;\theta) = \mathcal{U}\, \forall z \in \mathcal{Z}$.
2. $E_\pi\left[\ln E_{\rho(\cdot|z;\theta)} \exp(\gamma' g(U,X,\theta)\,|Z)\right]$ 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:

$$\begin{aligned} &\alpha w_{it}(p_{jt} - p_{j't}) + \delta \cdot \Delta(d_{it}, d_{i,t-1}, d'_{it}, \beta) \\ &+\beta^{T_i-t} \cdot B_{it} \cdot (r_{jt} - r_{j't}) \cdot (1+\omega)^{T_i-t-1} \\ &\geq \zeta_{ij't} - \zeta_{ijt} \end{aligned} \tag{1.22}$$

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

$$\begin{aligned} E\Big[\textstyle\sum_{j\in\mathcal{J}} 1[d^*_{it}=j] \cdot \Big[\textstyle\sum_{j'\in\mathcal{J}-\{j\}} 1[\alpha \cdot w_{it}(p_{jt}-p_{j't}) + \\ \delta\cdot\Delta(d_{it},d_{i,t-1},d'_{it},\beta) + \beta^{T_i-t}\cdot B_{it}\cdot(r_{jt}-r_{j't})\cdot(1+\omega)^{T_i-t-1} < \zeta_{ij't}-\zeta_{ijt}\Big]\Big]\Big] = 0 \end{aligned} \tag{1.23}$$

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 restrictions[13]:

$$\begin{aligned} E\left[z_{ij't}\zeta_{ijt}\right] &= 0 \\ E\left[z_{ij't}\zeta_{ijt}^2\right] &= 0 \end{aligned} \quad \forall j, j' \in \mathcal{J} \tag{1.24}$$

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 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

[13]Note that for the second set of moments to make sense, the instrument must be de-meaned, as otherwise these moments cannot hold.

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 $X_{ijj't} \equiv w_{it}\left(p_{jt} - p_{j't}\right)$, $\epsilon_{ijj't} \equiv \epsilon_{ijt} - \epsilon_{ij't}$, and assume $\epsilon \sim F$. Then:

$$\begin{aligned}
Y_{it} = 0|d_{i,t-1} = j &\Rightarrow -X_{ijj't} + \delta\left(1+\beta\right) \geq -\epsilon_{ijj't} \\
Y_{it} = 0|d_{i,t-1} = j' &\Rightarrow X_{ijj't} + \delta\left(1+\beta\right) \geq \epsilon_{ijj't} \\
Y_{it} = 1|d_{i,t-1} = j &\Rightarrow -X_{ijj't} + \delta\left(1-\beta\right) \leq -\epsilon_{ijj't} \\
Y_{it} = 1|d_{i,t-1} = j' &\Rightarrow X_{ijj't} + \delta\left(1-\beta\right) \leq \epsilon_{ijj't}
\end{aligned} \tag{1.25}$$

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).

$$\begin{aligned}
\Pr\left[X - \delta\left(1+\beta\right) < x, Y = 1|j,z\right] \leq F\left(x|j,z\right) \leq 1 - \Pr\left[X - \delta\left(1-\beta\right) > x, Y = 0|j,z\right] \\
\Pr\left[X + \delta\left(1-\beta\right) < x, Y = 0|j',z\right] \leq F\left(x|j',z\right) \leq 1 - \Pr\left[X + \delta\left(1+\beta\right) > x, Y = 1|j',z\right]
\end{aligned} \tag{1.26}$$

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

$$\begin{aligned}
&\sup_z \left\{\pi_j \Pr\left[X - \delta\left(1+\beta\right) < x, Y = 1|j,z\right] + \pi_{j'} \Pr\left[X + \delta\left(1-\beta\right) < x, Y = 0|j',z\right]\right\} \\
&\qquad\qquad \leq F\left(x\right) \leq \\
&1 - \sup_z \left\{\pi_j \Pr\left[X - \delta\left(1-\beta\right) > x, Y = 0|j,z\right] + \pi_{j'} \Pr\left[X + \delta\left(1+\beta\right) > x, Y = 1|j',z\right]\right\}
\end{aligned} \tag{1.27}$$

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

$$1 < \sup_z \{\pi_j \Pr[X - \delta(1+\beta) < x, Y = 1|j,z] + \pi_{j'} \Pr[X + \delta(1-\beta) < x, Y = 0|j',z]\} \tag{1.28}$$
$$+ \sup_z \{\pi_j \Pr[X - \delta(1-\beta) > x, Y = 0|j,z] + \pi_{j'} \Pr[X + \delta(1+\beta) > x, Y = 1|j',z]\}$$

Loosely, as $\delta \to \infty$ 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 $\delta \to -\infty$, 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.

1.5 Implementation

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

$$\begin{gathered} E\Big[\textstyle\sum_{j\in\mathcal{J}} 1\left[d^*_{it} = j\right] \cdot \Big[\textstyle\sum_{j'\in\mathcal{J}-\{j\}} 1\left[\alpha \cdot w_{it}\left(p_{jt} - p_{j't}\right) + \right. \\ \left. \delta \cdot \Delta\left(d_{it}, d_{i,t-1}, d'_{it}, \beta\right) + \beta^{T_i - t} \cdot B_{it} \cdot \left(r_{jt} - r_{j't}\right) \cdot \left(1+\omega\right)^{T_i - t - 1} < \zeta_{ij't} - \zeta_{ijt}\right]\Big]\Big] \end{gathered} = 0 \tag{1.29}$$

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:

$$\begin{array}{c} E\left[\sum_{j\in\mathcal{J}} 1\left[d^*_{it} = j\right] \cdot \left[\sum_{j'\in\mathcal{J}-\{j\}} 1\left[-w_{it}\left(p_{jt} - p_{j't}\right) + \right.\right.\right. \\ \left.\left.\left. \delta \cdot \Delta\left(d_{it}, d_{i,t-1}, d'_{it}, \beta\right) + \beta^{T_i - t} \cdot B_{it} \cdot \left(r_{ijt} - r_{ij't}\right) \cdot (1+\omega)^{T_i - t - 1} < \zeta_{ij't} - \zeta_{ijt}\right]\right]\right] \end{array} = 0 \tag{1.30}$$

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; $r_{ijt} = \frac{\sum_{i\in\{A,B,C,D,E\}} B^f_{it} r^f_{jt}}{B_{it}}$ 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:

$$\begin{array}{c} E\left[z_{ij't}\zeta_{ijt}\right] = 0 \\ E\left[z_{ij't}\zeta^2_{ijt}\right] = 0 \end{array} \quad \forall j, j' \in \mathcal{J}_t \tag{1.31}$$

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 1.21. Having obtained $\tilde{g}(Z, \theta, \gamma)$, one can test whether θ is in the identified set by solving:

$$L_n(\theta, \gamma) = \min_{\gamma} \tilde{g}(Z, \theta, \gamma)' V(\theta, \gamma)^{-1} \tilde{g}(Z, \theta, \gamma) \tag{1.32}$$

This is standard continuous updating GMM (CUE) (Hansen et al. [1996]). Owen [2001] shows that $n \cdot L_n(\theta, \gamma) \xrightarrow{d} \chi^2_q$, 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 identifed set[14]. Unfortunately, this function is non-convex and thus

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

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:

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

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 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.

1.6 Results

Figure 1.6 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 1.9 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 1.10 presents results of a standard multinomial logit model with firm fixed effects. To be specific, the utility of choosing firm j is modelled as:

$$u_{ijt} = -\delta \cdot [d_{i,t-1} \neq j] - w_{it} \cdot p_{jt} + \kappa \cdot r_{ijt} \cdot B_{it} + \vartheta_j + \epsilon_{ijt} \tag{1.33}$$

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.

1.7 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 1.3, 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:

$$\begin{aligned} &\Pr\left(u\left(j, d_{i,t-1}, X_{ijt}, \epsilon_{ijt}\right) - u\left(j', d_{i,t-1}, X_{ij't}, \epsilon_{ij't}\right)\right. \\ &\left.+\beta^{T_i - t} \cdot B_{it} \cdot E\left[\left(r_{jt} - r_{j't}\right) \prod_{\tau=t+1}^{T_i} \left(1 + r_{j^*\tau}\right) | \Omega_{it}\right] \geq 0 \forall j' \in \mathcal{J}_t - \{j\}\right) \end{aligned} \tag{1.34}$$

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:

$$\Pr\left(-w_{it}\left(p_{jt} - p_{j't}\right) + \beta^{T_i - t} \cdot B_{it} \cdot \left(r_{jt} - r_{j't}\right) \cdot (1 + \omega)^{T_i - t - 1} \geq \zeta_{ij't} - \zeta_{ijt} \forall j' \in \mathcal{J}_t - \{j\}\right) \tag{1.35}$$

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:

$$\begin{aligned} &u\left(j, j, X_{ijt}, \epsilon_{ijt}\right) - u\left(j', j, X_{ij't}, \epsilon_{ij't}\right) \\ &\geq \beta \cdot \left(E\left[V_{i,t+1}\left(j', B_{it}\right) | \Omega_{it}\right] - E\left[V_{i,t+1}\left(j, B_{it}\right) | \Omega_{it}\right]\right) \end{aligned} \tag{1.36}$$

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

$$\begin{aligned} & -w_{it}\left(p_j - p_{j'}\right) + \hat{\delta} \\ \geq\ & \epsilon_{ij't} - \epsilon_{ijt} + \beta \cdot \left(E\left[V_{i,t+1}\left(j', B_{it}\right)|\Omega_{it}\right] - E\left[V_{i,t+1}\left(j, B_{it}\right)|\Omega_{it}\right]\right) \end{aligned} \tag{1.37}$$

Although knowing $\hat{\omega}$ gives us some information about $E\left[V_{i,t+1}\left(j, B_{it}\right)|\Omega_{it}\right]$ for every firm j, we would still need further assumptions to find these values. It would be possible to estimate a distribution for $\epsilon_{jt} + E\left[V_{i,t+1}\left(j, B_{it}\right)|\Omega_{it}\right]$ 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:

$$\inf_{\gamma \in \mathbb{R}^{d_g}} \left\|E_\pi\left[\tilde{g}\left(Z, \theta, \gamma\right)\right]\right\| = 0 \tag{1.38}$$

Recall that $\tilde{g}\left(Z, \theta, \gamma\right)$ 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 $\hat{\gamma}\left(\theta\right)$'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:

$$w_{ist}\left(\hat{\gamma}\left(\theta\right)\right)=\frac{\exp\left(\hat{\gamma}\left(\theta\right)'g\left(u_s,z_{it},\theta\right)\right)}{\sum_s\exp\left(\hat{\gamma}\left(\theta\right)'g\left(u_s,z_{it},\theta\right)\right)} \tag{1.39}$$

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:

$$w_s\left(\hat{\gamma}\left(\theta\right)\right)=\frac{1}{N\cdot T}\sum_{t=1}^{T}\sum_{i=1}^{N}\frac{\exp\left(\hat{\gamma}\left(\theta\right)'g\left(u_s,z_{it},\theta\right)\right)}{\sum_s\exp\left(\hat{\gamma}\left(\theta\right)'g\left(u_s,z_{it},\theta\right)\right)} \tag{1.40}$$

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 2011[15].

Results from this exercise are in Table 1.11. 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, 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

[15] That is, using the observed salaries and returns realization of that month.

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 costs[16].

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.

1.8 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 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

[16] I 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.

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 heterogeneity 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.

Figures

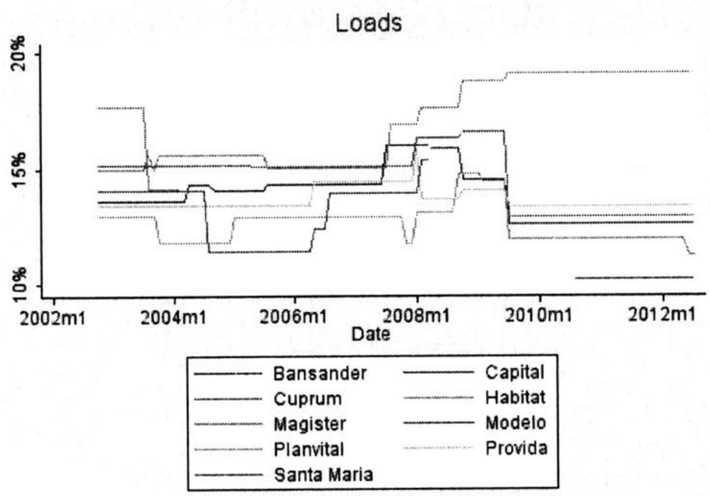


Figure 1.1: Loads by PFA, 2002 to 2012

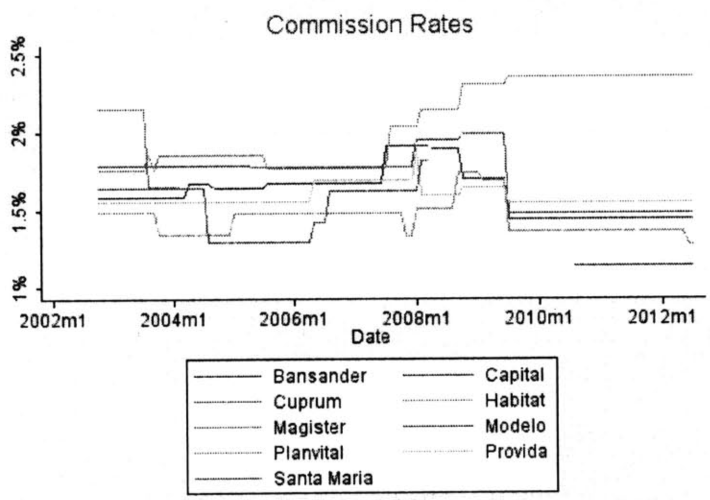


Figure 1.2: Commission Rates by PFA, 2002 to 2012

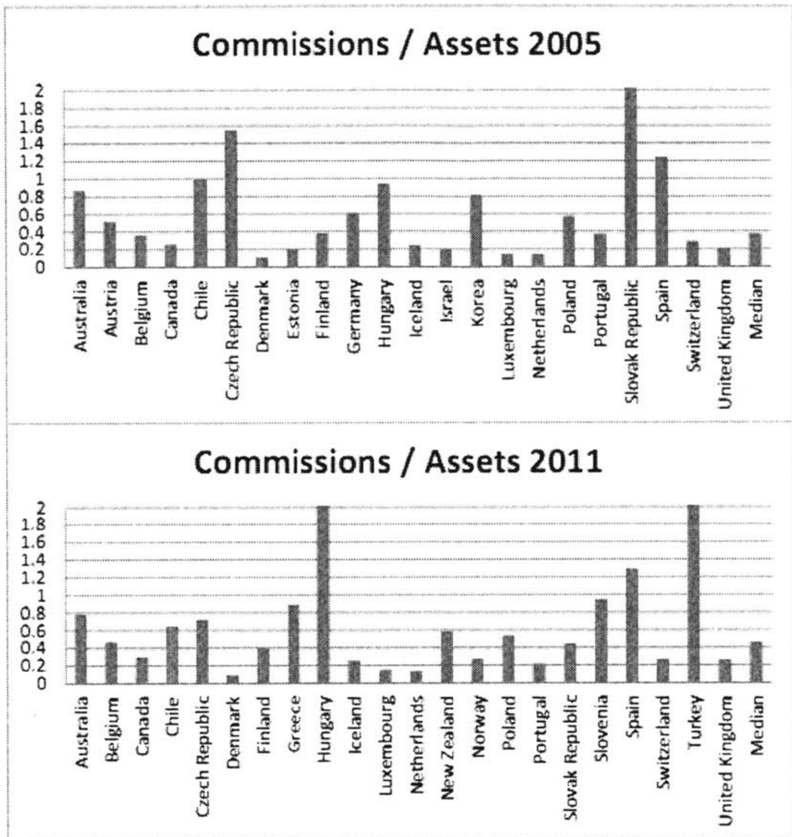


Figure 1.3: Comparison Across Pension Systems

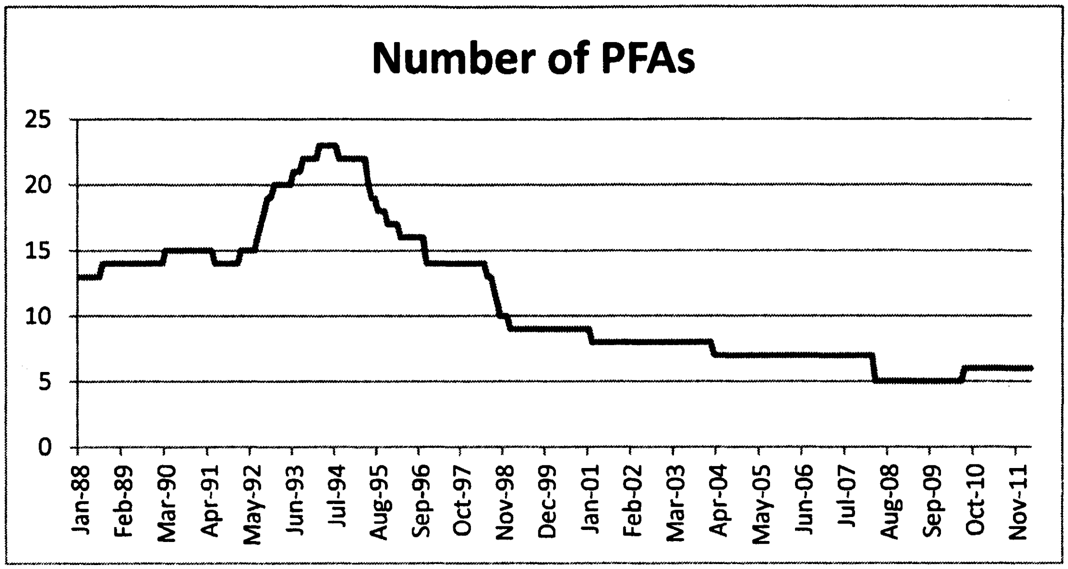


Figure 1.4: Number of PFAs

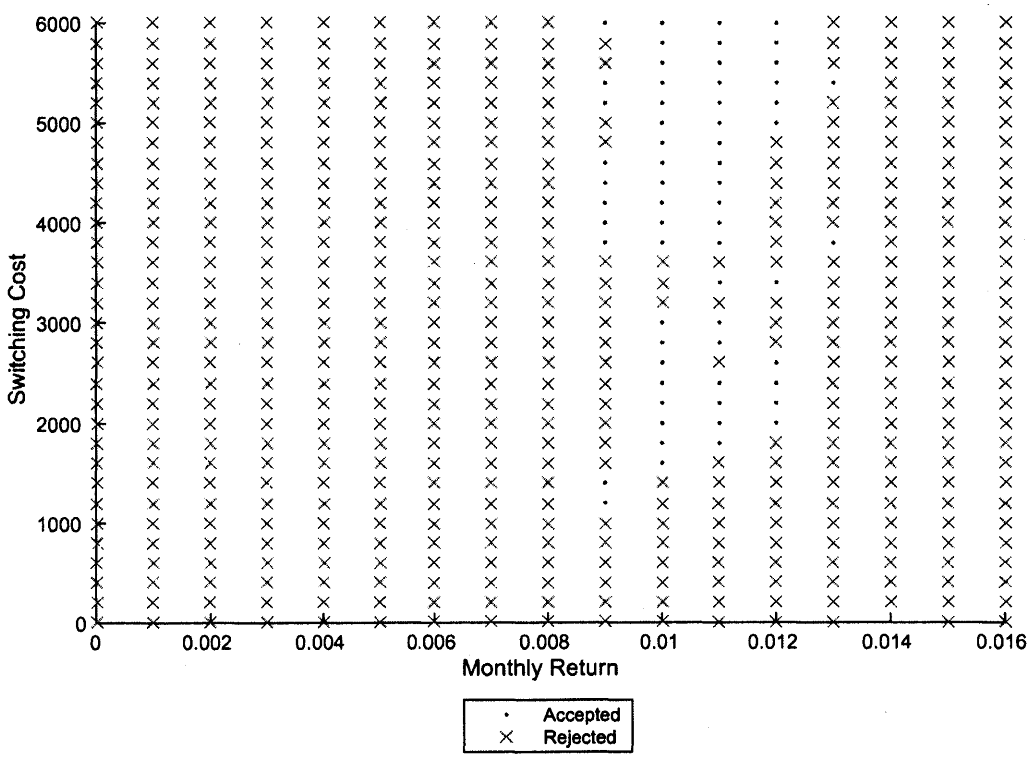
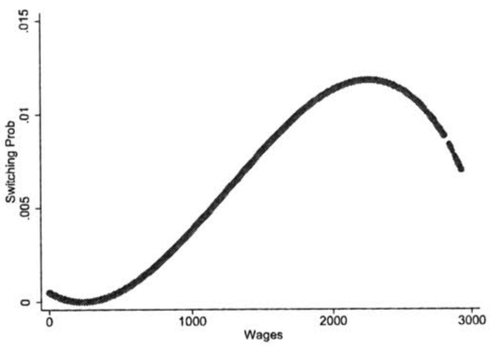
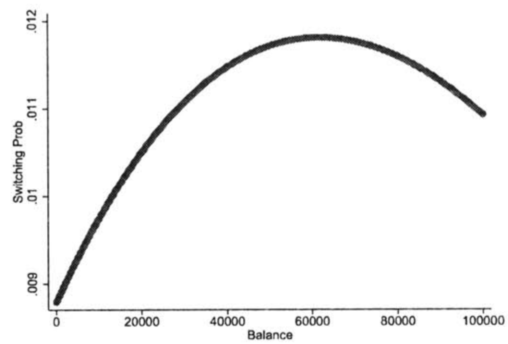


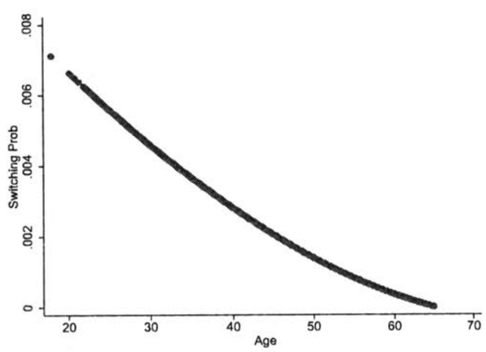
Figure 1.6: Parameter Estimates



(a) Predicted Switching Probability and Wage

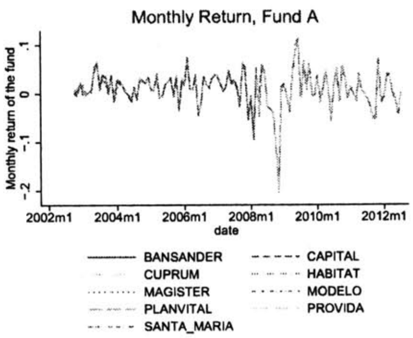


(b) Predicted Switching Probability and Account Balance

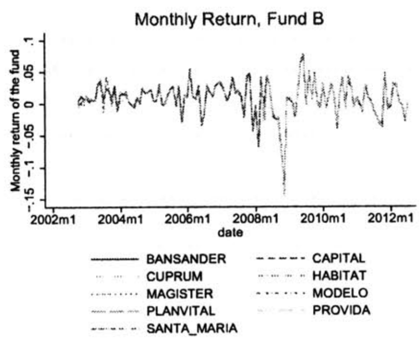


(c) Predicted Switching Probability and Age

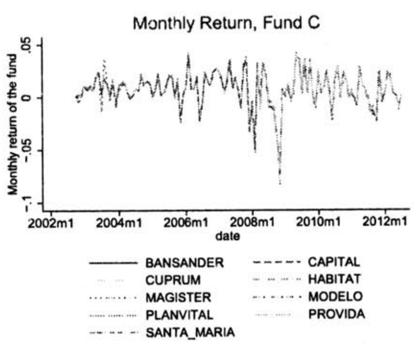
Figure 1.5: Correlations between Switching and Observables



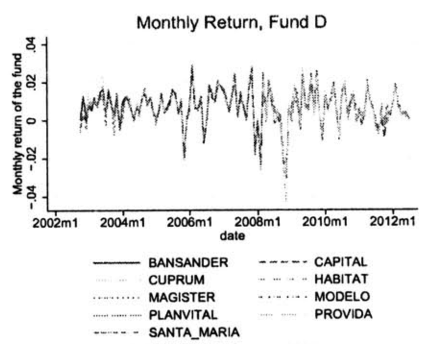
(a) Fund A



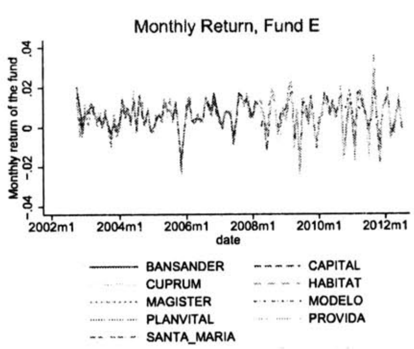
(b) Fund B



(c) Fund C



(d) Fund D



(e) Fund E

Figure 1.7: PFA Returns, by Fund

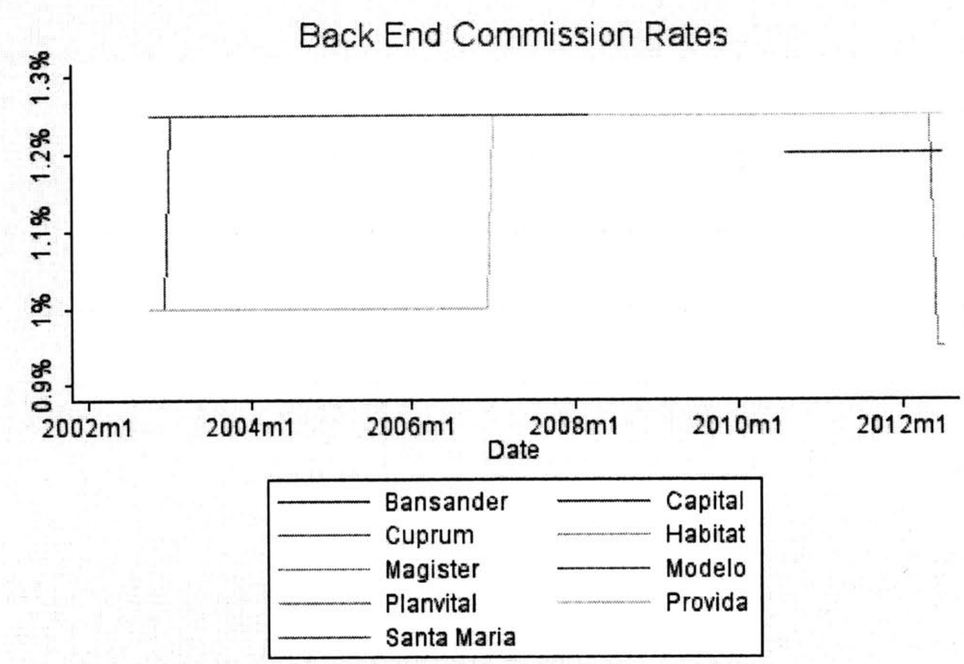


Figure 1.8: Back-end Commissions

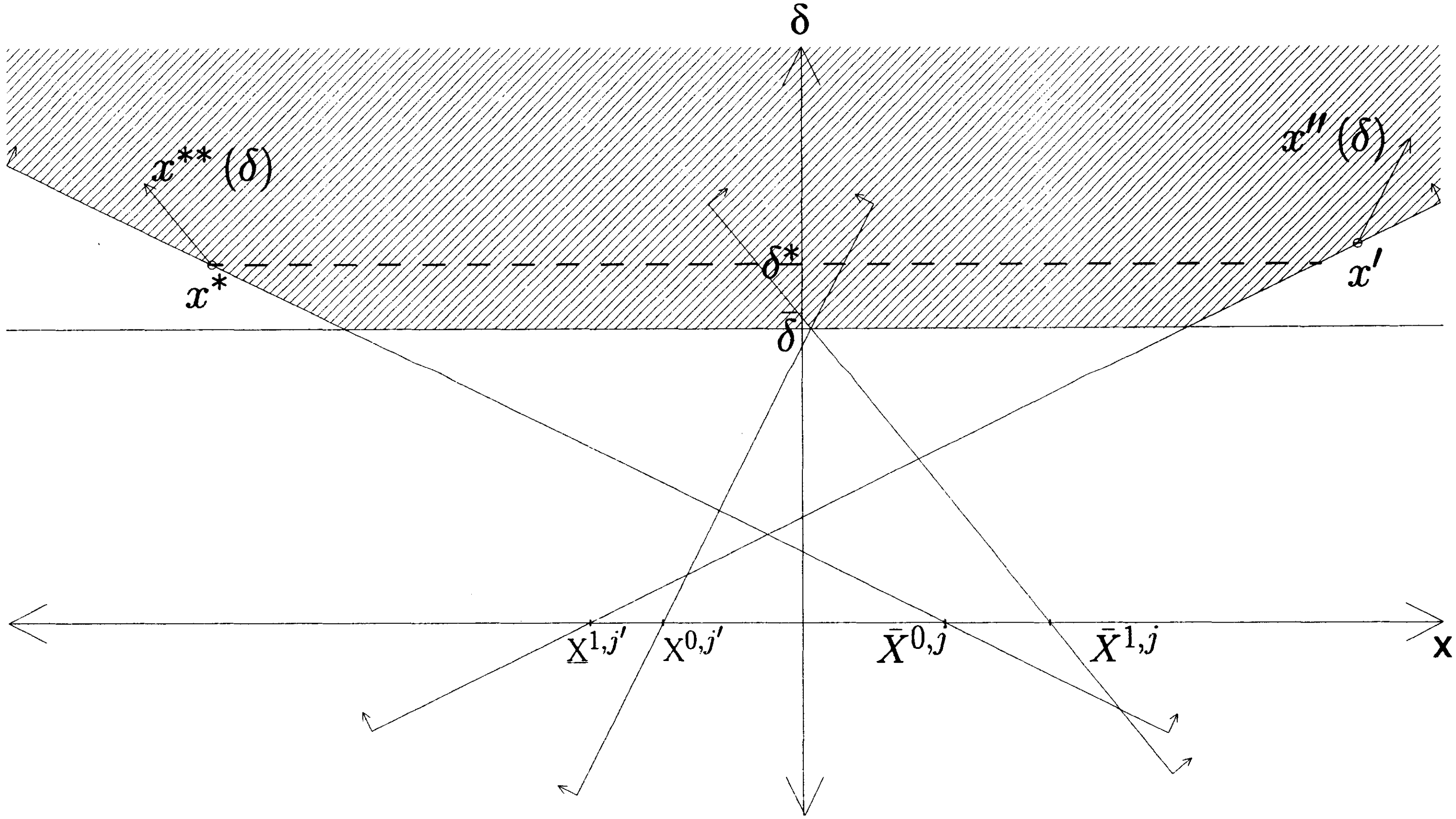


Figure 1.9: Identification of an upper bound on the switching cost parameter

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 of the fund, per issuer.	2%
B	25% - 60%	50%	40%	30%	4%	30%		2%
C	15% - 40%	30%	50%	10%	3%	20%		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 elegible. Investments in private investment funds are not allowed, and neither are direct investments in real estate.

Table 1.1: Investment Caps, by Fund

Tables

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

Table 1.2: Probabilities of Choosing the Absorbing Firm in January 2007

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 1.3: Summary Statistics

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 1.4: Summary Statistics for Account Balances

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 1.5: 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 1.6: Switching and Wage Changes

Table: Correlation between Salesforce / Advertising and Returns					
Panel A: Salesforce (# of workers)					
Fund	A	B	C	D	E
Monthly Returns	142.1	176.7	314.4	606.9	-433.5
	(189.57)	(274.12)	(438.42)	(793.21)	(674.95)
Semester Returns	-179.8***	-262.8***	-413.9***	-567.7**	692.2**
	(55.28)	(79.42)	(135.21)	(263.13)	(325.11)
N	222	222	222	222	222
Panel B: Advertising Expenditure (US$)					
Fund	A	B	C	D	E
Monthly Returns	-3,885.7	-3,045.8	-7,017.7	-25,072.7	12,792.8
	(9768.74)	(13807.41)	(21440.07)	(37023.32)	(44465.35)
Semester Returns	-146.2	1,479.5	3,726.5	9,806.5	34,717.0*
	(3353.04)	(4742.76)	(7756.29)	(14367.18)	(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 1.7: Salesforces / Advertising and Returns

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 1.9: Historical Returns

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 acounts in the database, and Panel B excludes accounts without any contributions in the sample period.

Table 1.8: PDV of Commission Savings Between Cheapest and Observed Paths

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 1.10: Myopic Multinomial Logit Estimates

Prices under a No Switching Cost Counterfactual				
		Counterfactual Loads		
Company	Actual Load	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 1.11: 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 1.12: Mapping Commissions to Expense Ratios

Pairwise Correlations in Monthly Returns Across PFAs								
Firm	Bansander	Capital	Cuprum	Habitat	Magister	Modelo	Planvital	Provida
Fund A								
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 1.13: Correlation in Monthly Returns Across PFAs, by Fund

Auction Bids for the Right to Serve New Consumers, 2010-2014									
	2010			2012			2014		
PFA	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 1.14: Auction Bids, 2010-2014

Correlation Between Returns, Commission Rates and Lagged Returns						
	All Funds	Fund A	Fund B	Fund C	Fund D	Fund E
	(1)	(2)	(3)	(4)	(5)	(6)
Commission Rate	-0.040	0.008	-0.015	-0.034	-0.065**	-0.113***
	(0.100)	(0.045)	(0.036)	(0.034)	(0.031)	(0.039)
Lagged Monthly Returns	0.187***	0.091	3.39E-04	-0.094	0.076	0.018
	(0.036)	(0.057)	(0.101)	(0.112)	(0.057)	(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 1.15: Correlation Between Returns and Commission Rates

Chapter 2

Choice in Privatized Social Security Markets

The merits and drawbacks of privatizing social security have been discussed at length in different countries, including the United States, over the last decade. Proponents of privatized systems argue that they take retirement savings away from politician's hands, where there is a temptation to spend them immediately, and that individual choice and competition would make commissions low while driving the funds towards profitable investments. Detractors, on the other hand, suggest that having many companies investing retirement savings and administering retirement benefits leads to an inefficient duplication of fixed costs, and so it is doubtful that any gain in better investment strategies will lead to larger returns when controlling for commission rates, particularly if individuals are financially illiterate or uninterested and thus competition is weak. This paper studies choice of pension fund administrator (or AFP[1]) in Chile, a country that has had a privatized and mandatory social security system since 1981, to determine consumer's substitution patterns between firms as commissions and returns vary. It will be argued that individuals should be very sensitive to commission rates, as the service of administering one's retirement savings can be thought of as a homogenous good (or very close to one), and that this sensitivity should be increasing in wages. At the same time, I will argue that individuals should not be sensitive to returns. These two hypotheses will be tested with administrative data from Chile's pension fund supervising agency, the Superintendencia de AFP (or SAFP). This exploration of consumer behavior in a mandatory privatized social security system is of interest not only because of

[1]The acronym in Spanish.

its potential public policy implications for Chile, but also because the insights gained might be useful in other settings where mandatory private systems are being enacted, such as the health care system in the United States. Indeed, if individuals are not responding to differences in commission rates by switching firms, one of the main arguments for a private system, that competition will lower costs for the consumer, is negated.

Results suggest that the effect of a 0.1 percentage point increase in commission rate above the mean commission level generates a drop in the probability that a firm is chosen that ranges from 0.3 to 12.5 percentage points, depending on the AFP, while the effect of a 100 percentage point increase in returns is an increase in the probability that a firm is chosen that varies from roughly 0.0005 to roughly 0.03, again depending on the AFP. The first result suggests wildly differing sensitivities to commission rates by AFP, suggesting that they may serve different populations with underlying differences in incentives to switch firms or in financial literacy, while the second suggests a low impact of returns. Further analysis of differences in price sensitivity by subgroups yields suggestive, but not conclusive, evidence that richer individuals are more sensitive to commission rates than the general population. These results yield interesting questions regarding consumer and firm behavior that could potentially be explored in future work.

This work proceeds as follows: Section 1 describes some relevant institutional details of the AFP system, Section 2 describes the data and the methodologies used for this analysis, Section 3 presents and discusses results, and Section 4 concludes.

2.1 Description of the AFP System

As was previously mentioned, Chile has a heavily regulated private and mandatory social security system, where workers must choose among several pension fund administrators, or AFPs. Workers contribute 10% of their monthly income to their AFP account, up to a maximum contribution that is linked to the consumer price index and that in February 2012 was at about $310 dollars or $151.100 pesos. On top of this contribution, consumers pay a fixed commission (until September 2008, when they were eliminated by law) and a variable commission rate that is a percentage of their income. That is, a consumer that chooses an AFP that charges a 3% variable commission rate pays 13% of their income to the AFP every month. Neither of these commission levels has ever been regulated. These are the only commissions that are allowed by law, so for example there cannot be commissions linked to the amount of savings in the account.

There are several regulations in place to protect workers' funds. First, AFPs 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 AFP goes bankrupt, the value of the pension funds that it manages should be unaffected. In fact, during the 1982-83 crisis, several AFPs went bankrupt, and workers' savings were not affected (Diamond and Valdés, 1993). Furthermore, AFPs 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 2.1 shows the number of firms in the market, month by month, from January 1988 to December 2011, showing that the market was relatively dynamic during the early 90's, and has stabilized since. Second, AFPs are not free to invest as they wish, but must fall within caps for exposure to variable and fixed income, foreign and domestic assets, and investment grade levels. Furthermore, since the year 2002 AFPs must offer five different funds in which savings are invested, each with a different mix of fixed and variable income securities. Funds are labeled from A to E, with A having the largest proportion of variable income securities and E the smallest. Workers are placed by default into fund C, which must invest between 40% and 15% of its funds in variable income assets. At the same time, only Fund C was available before 2002.

Overall, these regulations imply that it would be difficult for an AFP to significantly differentiate itself from its competitors through its investment strategy. In fact, Walker [1993a,b] and Zúñiga [1992] argue that no AFP was off the investment possibility frontier, while Diamond and Valdés [1993] argue that regulation impedes AFPs from choosing significantly different points on said frontier. Furthermore, Zurita and Jara [1999] show that the serial correlation in returns, while positive, is only statistically significant in 20% of the periods analyzed, and goes to zero as the time window studied expands. Having argued that returns should not be significantly different between companies, the only dimensions where firms can differentiate themselves are quality of service, number of branches, and marketing efforts.

The fact that firms compete on commission rates, rather than an actual price, introduces different incentives to individuals with different levels of income. To be more precise, the incentive to switch between companies increases when an individual's wages increase, as the real monetary gain of doing so is greater. This implies that richer individuals should be more sensitive to higher commissions, and that if there are switching costs, individuals with low wages may be locked in to firms. To change AFP, one must physically go to a new AFP´s branch on a weekday and switch[2], a process that in January 2012 took the author´s wife roughly 3 hours. Naturally, the difficulty of switching increases

[2]Online switching has recently been introduced, but the author was unable to perform a switch in January 2012, so it is not particularly easy.

if one does not live in a large city with a branch. The effect of having wages affect the incentive to switch companies on the empirical strategy will be discussed further in the next section.

2.2 Data and Empirical Methodology

This work uses two primary databases, a firm level database and an individual level database. The firm level database contains monthly information on each AFP's market share, commission rate and return on each fund (A through E) over the last 12 months, and was constructed using information available online from the SAFP. This database has information from January 1988 to December 2010. Figure 2.2 presents variable commission rates for the companies that are present in the market on December 2011, while Figure 2.3 presents fixed commission rates for these same companies.

The individual level database is a representative sample of the workers who have accounts in the AFP system, and it has monthly information on wages, fixed commissions paid, variable commissions paid, gender and age from January 1988 to December 2010. It was constructed using the Pension Histories Database, published by the SAFP. Unfortunately, it has no monthly information on which AFP each individual has chosen. Nevertheless, since AFPs usually charge different commissions, using data on the fixed commission paid and the variable commission paid one can infer to which AFP each customer belongs, a procedure suggested by Bernstein and Cabrita [2006]. To do so, I match workers to AFPs by looking separately at fixed commissions and variable commissions, and choosing the inference based on the fixed commission if they differ[3]. This generates an imputed AFP variable that I label as conservative. Then, if individuals have gaps of less than six months in their imputed AFP history, and the inferred AFP before and after the gap is the same, I fill in the gap with the same firm, and generate a second imputed AFP variable that I label as aggressive.

Figure 2.5 presents the monthly percentage of individuals in the sample that are matched using both variables. It is clear that this procedure is not perfect, and that for many years the matched percentage is low. However, from January 1999 to midway through 2008 more than 80% of the sample is matched. Figure 2.6 presents the monthly fraction of individuals in the sample that switch companies, using both imputed AFP variables. In this case, they are both identical. This finding gives me some confidence in the aggressive imputed AFP method, and therefore all future results that use imputed AFPs use this measure. Restricting attention to the period between January 1999 and June 2008, it seems that the switching pattern is reasonably stable for the first years, but then experiences massive spikes in the latter years. After careful research, I was not able to find changes in market conditions or

[3]This happens in less than 2% of the matches.

in the regulatory framework that can explain such massive shifts, so I conclude that they must be due to flaws in the matching mechanism. These flaws are plausibly due to two factors: the elimination of fixed commissions in September 2008, which eliminated one matching criterion, and the merger of two AFPs in late 2007. As a result, I restrict my analysis to the period between January 1999 and June 2007. This restriction will also be imposed on the results obtained when using data from the firm level database, as I will attempt to draw comparisons between specifications that use just firm level data and specifications that include individual characteristics. Table 2.1 shows the comparison of individual level characteristics for matched and unmatched individuals, for the full database (1988-2010) and for the estimation sample (January 1999 to June 2007). The table shows that unmatched individuals have lower wages, are older, and are slightly more likely to be males, although none of these differences are statistically significant. After the matching process and the aforementioned restriction on the time period analized, the individual level database consists of 7,791,376 observations over 90 months. During this period, there were 8 active firms, and 6 of them were present for the entire period. Figure 2.4 presents market shares for all 8 firms during the period under study.

To determine the effect of commissions and returns on the demand for each AFP's services, I model the decision of which AFP to choose using a conditional logit model as in McFadden [1974]. Assuming that individual i's latent utility from choosing AFP j in time period t can be written as:

$$u^*_{ijt} = x_{ijt}\beta + \epsilon_{ij} \tag{2.1}$$

where ϵ_{ij} are unobservables. If ϵ_{ij} are independently distributed and follow an extreme value Type 1 distribution, we have that:

$$Pr\left(y_{it} = j|x_{it}\right) = \frac{\exp\left(x_{ijt}\beta\right)}{\sum_{k=0}^{J} \exp\left(x_{ikt}\beta\right)} \tag{2.2}$$

where $y_i = \arg\max\left(u^*_{i0t}, ..., u^*_{iJt}\right)$ and $j = 0, ..., J$. When using the firm level database, this regression is run without individual level variables, by including fixed commission, variable commission, and the last 12 months' returns as regressors; when using the individual level database, the regressors include effective commissions (that is, $fixed\, commission + variable\, commission \times wages$) and the last 12 months' returns.

Since both variable and fixed commissions are endogenous variables, and in this setting there are unobserved (to the econometrician) firm characteristics, such as marketing expenses, number of branches, and quality of services, it seems clear that this endogeneity could be biasing the estimates of the effect of commissions on the probability that each firm is chosen. In order to control for this, two separate approaches are taken. For the firm level regressions, one can follow the Berry inversion (Berry [1994]) to find that $\ln(s_{jt}) = x_{jt}\beta + \delta_t + \epsilon_{ijt}$, where the δ_t captures the effect of the $-\ln\left(\sum_{k=0}^{J} \exp(x_{ikt}\beta)\right)$. Using this specification, one can then use standard instrumental variable methods to consistently estimate the utility function's parameters. Berry [1994] also suggests using rival firms' exogenous characteristics as instruments in this setting, as they are correlated with a firm's commissions through the Nash equilibrium of the game through which they compete, and as they should be uncorrelated with the firm's unobserved characteristics if they are truly exogenous. In this setting, the only available instruments are rivals' returns for the past 12 months. Clearly, each firm's returns are affected by their choices, as they decide on what to invest, but if firms choose their investment portfolios irrespective of other firms' unobserved characteristics, then these instruments will be valid. Furthermore, the evidence presented by Diamond and Valdés [1993] suggesting that firms are constrained by regulation in their choice of points on the investment possibility frontier lends some validity to these instruments. However, if firms decide to be more or less aggressive in their investment portfolio to counter rivals′ marketing expenses, for example, then these instruments will not satisfy the exclusion restriction. For the moment, I will present results using these instruments, but an exploration of other potential sources of exogenous variation is pending.

For the conditional logit model with individual characteristics, I was not able to find the appropriate Berry inversion, as shares now should depend on the sum of individual characteristics. To deal with the endogeneity of effective commissions, I use two stage residual inclusion, as presented by Terza et al. [2008]. The idea is to estimate a first stage equation of a firm′s effective commission on its' rivals returns and its own returns, and then to plug in the residual from that equation into the conditional logit. That is, in a first stage I estimate:

$$ec_{ij} = \alpha_j + r_{-j}\gamma_j + r_j\lambda_j + \nu_{ij} \tag{2.3}$$

where ec_{ij} is the effective commission paid by individual i if she chooses firm j, r_{-j} is a matrix containing each rivals return for the last 12 months, and r_j is the firms' return for the same period.

Then, in a second stage, I estimate:

$$Pr\left(y_{it} = j|x_{it}, \hat{\nu}_{ij}\right) = \frac{\exp\left(x_{ijt}\beta + \hat{\nu}_{ij}\right)}{\sum_{k=0}^{J} \exp\left(x_{ikt}\beta + \hat{\nu}_{ik}\right)} \tag{2.4}$$

where $\hat{\nu}_{ij}$ are the fitted residuals. Since $\hat{\nu}_{ij}$ is a consistent estimator of the effect of unobserved characteristics on commissions if the instruments are valid, then our estimates of β using this method will be consistent. However, since $\hat{\nu}_{ij}$ is an estimate, the standard conditional logit standard errors will be wrong, as they do not take this into account. In order to obtain the appropriate standard errors, one must bootstrap the whole procedure, and bootstrap standard errors with 50 replications are presented whenever this method is used.

2.3 Results

2.3.1 Firm Level Estimation

This section presents results obtained from regressions where only firm level characteristics are included, that is, when utility from choosing firm j is specified as $u^*_{ijt} = x_{jt}\beta + \epsilon_{ijt}$, and the elements of x are commissions and previous returns. Note that these results are a first approximation to the problem, as individuals should be concerned with the effective commission paid, which varies by wage, rather than by the commission rates charged by each firm. Nevertheless, this is a relevant starting point. For all of this section, I will call OLS results the coefficients of the OLS regression of $\ln(s_{jt}) = x_{jt}\beta + \delta_t + \epsilon_{ijt}$, which is simply the Berry inversion applied to the conditional logit model, and IV results the coefficients of the 2SLS estimation of the aforementioned equation, using the Berry instruments discussed in the previous section.

Table 2.2 presents results for these regressions, where odd columns include variable and fixed commissions as well as the last 12 months' return of Fund C as regressors, while even columns include commissions and the returns of all funds (A through E) as regressors. The right panel (Columns 5 through 8) presents results obtained when estimating using actual shares as the left hand side variable, while the left panel (Columns 1 through 4) presents results obtained when using estimated shares, calculated period by period on the matched sample of individuals, as the left hand side variable. This table presents coefficients, and not marginal effects, so the main economic intuitions regarding substitution elasticities cannot be garnered from it, but there are nonetheless several interesting takeaways of looking at the signs of these coefficients. First, Column 5 shows negative and significant OLS coefficients for commissions, as expected, as well as a positive and significant return of Fund C coefficient. When adding the return on other funds (Column 6), we see that the coefficients on commissions remain somewhat stable, while the coefficient on the return of Fund C increases dramatically and becomes very imprecise. In fact, coefficients on the return of all funds are not only imprecise but also unintuitive: they should all be positive if individuals believe that performance is correlated over time, and zero if they do not. The fact that some of them are negative casts doubt on what we are picking up when they are included. Also, signs and magnitudes do not seem to be following a pattern in terms of exposure to risk. On face value, it seems that low risk funds (D and E) have no impact on utility, while funds with more risk (A through C) do. However, estimates for low risk funds are very imprecise, and estimates for fund B are counterintuitive. A more rigorous modeling on the choice of fund and its potential correlation with choice of AFP is an area for future research.

Columns 1 and 2 replicate these regressions, but using estimated shares from the matching procedure instead of the actual market shares. These estimates are presented as a robustness check for the individual level analysis presented later, since in that case only matched data is available. Comparing columns 1 and 2 to columns 5 and 6 shows that coefficients for commissions are basically the same, and that coefficients for returns are different, although not in a statistically significant fashion. These findings give more confidence in the matching procedure and in the results of the later section.

Columns 7 and 8 replicate the regressions in columns 5 and 6, but using the instrumental variables method described previously. The overall patterns observed in the OLS coefficients remain similar, but the point estimate on variable commission rates becomes larger in absolute value, while the point estimate of fixed commission rates becomes smaller. Again, there is not much economic intuition in these numbers, so the discussion of their economic implications is postponed to the paragraphs discussing marginal effects. The last row of this table presents the joint F test of the null hypothesis that the excluded instruments are uncorrelated with fixed and variable commissions, showing that in all IV regressions the null is strongly rejected. Table 2.3 presents Hausman Tests of the null hypothesis that OLS and IV estimates are equal, showing that for both specifications and using both actual and estimated shares, we can reject the null. This is to be expected, as commissions are endogenous and correlated with omitted variables, such as service, marketing expenses and number of branches.

Table 2.4 presents marginal effects (at the mean) for the conditional logit model, with commissions and Fund C's return as regressors, and using actual shares. Values on the diagonal of this matrix present own effects, while off-diagonal elements present cross effects[4]. The first panel shows the marginal effect of increasing the return of Fund C by 100 percentage points above the mean return. When this scale is taken into consideration, one can see that these effects are truly negligible: at the most, increasing the return of Fund C by 100 percentage points increases own share by 12.3 percentage points. Thus, a more realistic number, such as a 1 percentage point increase over the mean, would yield at most a .12 percentage point increase in share[5]. Considering the regulatory barriers on portfolio choice, this evidence suggests that it is virtually impossible to gain significant market share through returns. This result is consistent with the hypothesis that past returns are not a good predictor of future returns, and as such individuals should not base their AFP choice on this variable. It is also interesting to note that some firms have significantly higher sensitivities than others: while Aportafomenta gains only 0.003 percentage points in share when increasing its Fund C return by 1 percentage point above the mean, Habitat gains 0.123. At the same time, some firms are more sensitive to competition for some rivals than from others: Santamaria loses 0.17 percentage points when Habitat raises its return by 1 percentage point above the mean, but only 0.003 points when Aportafomenta does the same. These results suggest that competition is more intense between some companies, suggesting that perhaps they differentiate into different populations. For example, the previous results could be explained if Aportafomenta were to specialize in poorer, less informed and locked in customers, while Santamaria and Habitat compete for richer, less locked in and more informed consumers.

The second panel in Table 2.4 shows the marginal effects of a 100 percentage point increase over the mean variable commission rate. Thus, charging 100 percentage points above the mean would reduce Habitat's share by 125.8 and Aportafomenta's share by 3.2, with Habitat being the most sensitive AFP and Aportafomenta the least. Using more intuitive scales, a 0.1 percentage point increase above the mean commission reduces Habitat's share by 12.5 percentage points, and Aportafomenta's by 0.3. Clearly, these two coefficients have extremely different economic implications. On the one hand, Habitat's coefficient suggests that individuals are extremely sensitive to commission rates, as charging 0.1 points above the mean commission gives Habitat a 12.5 percentage point lower share, while on the other Aportafomenta's coefficient suggests the opposite, that individuals are fairly insensitive. Results for other companies suggest that firms lose between 1.1 and 10.2 percentage points with such an increase. Overall, the drastic differences between firms suggests that either there are significant

[4] For all marginal effects matrices in this work, only the upper triangle is shown. I do this to reduce clutter, as these matrices are symmetric.

[5] Share is defined between 0 and 1.

differences in the populations that they serve, or that the model is misspecified and fails to pick up some important between firm variation in customer's sensitivity to commission rates. As was discussed previously, it does not seem reasonable to include commissions in each individual's utility function without interacting them with wages. After all, the effective commission paid is said interaction. Therefore, if consumer's wages vary between firms, one would find significantly differing marginal effects when looking at commission rates even if each firm has the same marginal effect when looking at effective commissions. The effect of using effective commissions will be studied in the next subsection.

Finally, the last panel in Table 2.4 shows the marginal effect of raising fixed commissions by one peso. Since 500 pesos is roughly one dollar, this panel shows that a one dollar increase over the mean commission rate reduces Aportafomenta's share by 0.2 percentage points and Habitat's share by 8 percentage points. Again, this is an extremely large effect for Habitat, and a very low one for Aportafomenta. Intuitively, poorer populations should be more sensitive to fixed commissions than to variable commissions, as fixed commissions will be a greater proportion of the total commission paid for these individuals. If this is the case, this runs counter to the findings of the previous paragraph, which also found Habitat to be more sensitive to variable commissions. This suggests that there could be more differences in the populations that each AFP serves that simply wages. This discussion will be touched upon further when studying the effect of effective commission paid, rather than this artificial separation between variable and fixed commission.

Table 2.5 shows the marginal effects at the mean of the Firm Level Conditional Logit Model, using instrumental variables to control for the endogeneity of fixed and variable commission rates. As in Table 2.4, this table is estimated using actual shares and controlling only for the last 12 months' return on Fund C and commissions. The first panel presents the marginal effect of increasing the return on Fund C by 100 percentage points over the mean return of Fund C. As in the estimates without controlling for endogeneity, the effects are small: a 1 percentage point increase over the mean would yield as little as a 0.003 percentage point increase in share (Aportafomenta) and at most a .15 percentage point increase in share (Habitat). These results are consistent with the previous discussion that suggested that returns do not have a relevant impact on choice probabilities. The second panel in Table 2.5 shows marginal effects for a 100 percentage point increase over the mean variable commission rate. As for returns, the main economic intuitions obtained from Table 2.4 remain: charging 0.1 points above the mean commission gives Habitat, the most price sensitive firm, a 14.3 percentage point lower share, and it gives Aportafomenta, the least price sensitive firm, a 0.25 percentage point lower share. Again, the disparities in marginal effects suggest that either the firms specialize in different populations or that the model is misspecified due to the inclusion of commission rates rather than effective commissions paid.

An important caveat of the previous discussion is that the conditional logit model imposes independence of irrelevant alternatives (IIA), that is, that the substitution between two firms depends only on the characteristics of those two firms, and not on other rival's characteristics. This assumption may be violated in practice, as is often the case when alternatives are similar (Wooldridge [2002]), and as a result the model will estimate unreasonable cross-firm marginal effects. If this is the case, then the conditional logit model will not be an appropriate choice, and other methods, such as conditional probit (Hausman and Wise [1978]) or the random coefficients multinomial logit in Berry et al. [1995] are appropriate. To test whether IIA is a restrictive assumption, Hausman and McFadden [1984] propose the following procedure: estimate the model twice, once with the full choice set and once with a restricted choice set, and perform a Hausman test of the null hypothesis that both estimates are equal. To implement this procedure, I estimate the IV Conditional Logit model with commissions and Fund C's return as regressors on the full choice set, and then again dropping three AFPs from the sample (Planvital, Provida, and Santamaria). A Hausman test of the null hypothesis that both sets of estimates are equal yields a Chi Squared statistic of 59.06 and a p-value of 0.000, rejecting the null. As a result, it seems that the assumptions of the conditional logit model are too restrictive for this setting, and that models that incorporate more flexible substitution patterns are required. The estimation of these models is left for future research.

Before moving on to the discussion of the estimates of the conditional logit model with individual specific variables, which uses the matched sample of individuals, it is important to investigate whether using the matched sample generates significantly different results than those of using the full sample. While a test of this cannot be performed for the conditional logit model with individual specific variables, a test can be performed for the conditional logit model with just firm level variables. That is, one can estimate the model with actual shares and with estimated shares to determine whether there are significant differences in coefficients and in marginal effects. The discussion of Table 2.2 already included some analysis of these differences, arguing that pairwise comparisons of coefficient estimates do not yield statistically significant differences. Table 2.6 presents marginal effects (at the mean) for the conditional logit model, using commissions and Fund C's return as regressors. Overall, the magnitudes and economic intuitions obtained when using estimated shares seem similar to those obtained when using actual shares: a 0.1 percentage point increase above the mean commission reduces Habitat's share by 12.5 percentage points, and Aportafomenta's by 0.3, just as before, while a 1 percentage point increase over the mean return of Fund C yields at most a .0012 percentage point increase in share. This is much smaller than the .12 percentage point increase found when using actual shares, but economically the interpretation is the same: increasing returns does not seem to lead to significant increases in share. Table 2.7 presents marginal effects for the same model, but instrumenting fixed and variable commissions using Berry instruments. As for the uninstrumented results, the marginal effects of increasing the return of Fund C are statistically smaller but economically similar, while the effects of increasing variable commission rates are mostly unchanged. As a result of these findings, one can be more confident in the results obtained when using the matched sample.

2.3.2 Individual Level Estimation

This section presents results obtained from regressions where firm and individual level characteristics are included, that is, when utility from choosing firm j is specified as $u^*_{ijt} = x_{ijt}\beta + \epsilon_{ijt}$. In this setting, the elements of x are the effective commission paid[6], previous returns, as well as age and gender controls. As was previously argued, effective commission paid is the appropriate price that an individual faces in this market, and as a result should yield relevant results. Table 2.8 presents the conditional logit estimates of the utility function's parameters under four different specifications. Column 1 presents results when effective commissions, Fund C's return for the last 12 months and a firm specific constant are included, column 2 adds age and gender controls to the specification in column 1, column 3 adds the return of all other funds to column 1's specification, while column 4 adds age and gender controls to the specification in column 3. There are two noteworthy findings in this table: first, the coefficient on effective commissions is positive, which is counterintuitive, and second, when controlling for the return of all funds, only Funds A and B have a significant impact on the choice, and Fund A has a counterintuitive effect. As was argued in the previous section, it is not immediately clear why a fund's return should have a negative coefficient, and thus this result merits further exploration in future work. As for the positive coefficient of effective commission, the endogeneity of commission rates could be responsible for this result. For example, if individuals face a switching cost, then some customers (those with low wages) are locked in to their firms, as a difference in commissions of a few percentage points translates to a very small monetary gain. If this is the case, then it makes sense for a firm to build a base of customers and then raise their commissions,

[6]That is, $fixed\,commission + variable\,commission \times wage$

creating the illusion of an upward sloping demand curve through coefficients that are biased upward. Alternatively, if higher commissions are correlated with unobservables that raise a consumer's utility, such as number of branches or quality of service, then the coefficient on effective commission will be biased upward, possibly resulting in a positive estimate. In either case, it is clear that an instrument is needed to control for the endogeneity of commissions.

However, before presenting results for the instrumented version of the individual level conditional logit model, a simple test of the hypothesis that the upward sloping demand result is driven by locked in customers with low wages can be constructed by reestimating the model using only individuals with high wages. Table 2.9 presents results for the same regressions as Table 2.8, but restricting the sample to those earning above 1 million pesos a month[7]. Clearly, the upward sloping demand result disappears, a finding that is consistent with the aforementioned hypothesis that this result is driven by low income earners who do not have incentives to switch AFPs. Tables 2.10 and 2.11 present marginal effects (at the mean) for both the full sample and the high earner sample versions of the individual level conditional logit model. As in Table 2.9, in Table 2.10 the effect of effective commission rates is positive, which is unreasonable. Also, the effect of Fund C's return is positive but small, as in the previous section. Due to the obvious misspecification present in this estimation, the magnitudes of the marginal effects are uninteresting and merit no further discussion. In Table 2.11, which includes only high income earners, we find marginal effects that at least are plausible, although they should also be biased upward due to the endogeneity of commissions. This table predicts that the firm with the largest marginal effect on effective commissions, Habitat, suffers a loss in share of 0.00343 percentage points when raising effective commissions by 1 peso above the mean, or of 0.2 points in share when raising effective commissions by 1 dollar above the mean, while the firm with the lowest sensitivity, Aportafomenta, loses 0.06 points when raising effective commissions by the same dollar amount. As in the previous section, the effect on Habitat is a relatively high, suggesting that individuals are willing to switch actively when commissions rise, while the effect on Aportafomenta suggests much lower sensibility. This result puts the onus on modeling heterogeneity in individual choice better, a question that for now is left for future research.

Table 2.12 presents coefficient estimates for the instrumental variables version of the conditional logit model with individual characteristics. As was mentioned in the Data and Empirical Methodology section, these estimates are obtained from a two step residual inclusion procedure, with bootstrap

[7]Roughly 2000 dollars.

standard errors. There are three important takeaways from this table: first, that even when instrumenting for the endogeneity of commissions, the effective commission coefficient is positive; second, that this finding disappears when looking at high income earners; and third, that the coefficient on the estimated error term from the first stage is significant. The first takeaway is surprising, and casts doubt either on the validity of the instruments in this setting or on the methodology used. If the instruments are invalid, as would be the case if investment portfolios are chosen taking into account other firms' unobservables, we should expect the coefficient on commissions to continue to have bias. If this is the case, then all previous regressions that use instrumental variables are also invalid, and better instruments are needed. Alternatively, it could be that the instruments are valid, and the methodology is inappropriate. To determine whether this is the case, future work is needed on implementing other ways to control for endogeneity in this setting, such as the methods used in Berry et al. [1995]. The second takeaway, that the upward sloping demand result disappears when looking at high income earners, confirms the previous paragraph's assertion that a plausible explanation for the finding that demand is upward sloping is that low income earners are locked in to firms. At the same time, it also confirms that this methodology is not correcting for endogeneity, suggesting that further work on implementing other methods is needed. Finally, for the third takeaway, a t test of the significance of $\hat{v}$ would be a regression based Hausman test of the null hypothesis that the estimates on this conditional logit model without correcting for endogeneity are equal to those obtained when correcting for endogeneity, but since we have just argued that the estimates of the instrumented model are inconsistent, this test is invalid.

Tables 2.13 and 2.14 present marginal effects for the instrumental variables version of the conditional logit model with individual characteristics, estimated with the full sample and with the high income earners sample, respectively. Table 2.13 confirms the problems discussed in the previous paragraph, as marginal effects for effective commissions have the wrong sign: own effects are positive, and cross effects are negative. As for Table 2.14, marginal effects on effective commission are of the correct sign, although we would still expect them to be biased towards zero. Nevertheless, their magnitude suggests sensibility to commission rates: the firm with the largest marginal effect, Habitat, suffers a 0.5 percentage point drop in the high income earner market when it raises its effective commission by one dollar above the mean, while the firm with the smallest marginal effect, Aportafomenta, suffers a 0.002 percentage point drop. As in the discussion of Table 2.14, it is clear from these results that more work on understanding individual level heterogeneity in the incentives to switch AFPs is needed. As for marginal effects on Fund C's return, there are no effects that are statistically significant, although this seems to be more due to low power than to zero effect of said returns on choice. This is to be expected, as the sample of high income earners is considerably smaller.

2.4 Conclusion

This work presents an initial exploration of the demand for pension fund administration in Chile's mandatory and private social security market. Using a conditional logit specification, it attempts to measure the degree to which consumers look at commissions and returns when deciding to choose a pension fund manager, or AFP, in order to determine the validity of the claim that competition will drive commissions down and result in savings for workers. The findings suggest that individuals are sensitive to commission rates and insensitive to returns, which is rational as previous returns should not be accurate predictors of future returns in this setting. At the same time, there seem to be significant disparities between firms in the effect of commissions on the probability that they are chosen, suggesting that they could be serving different populations or that elasticities are not well specified. An exploration of the latter hypothesis, via testing the independence of irrelevant alternatives assumption imposed by the conditional logit, suggests that more flexible methods should be used to determine substitution elasticities. At the same time, when controlling for a single dimension of individual level heterogeneity, heterogeneity in wages, results become inconclusive and biased, with more work needed on the estimation of models that control for endogeneity of wages. Overall, further work on modelling individual level heterogeneity and the impact of switching costs on elasticities of substitution is needed.

Nevertheless, these initial results suggest that in this setting individuals are sensitive to commission rates and insensitive to returns. Both results lend credence to the notion that individuals are behaving rationally in this market, and thus that competition should give firms incentives to lower commissions. It would be interesting to use combine these demand estimates with the estimation of the supply side game that firms are playing, in order to estimate markups and to truly test the hypothesis that competition leads to lower costs for consumers. This is left for future research.

Figures

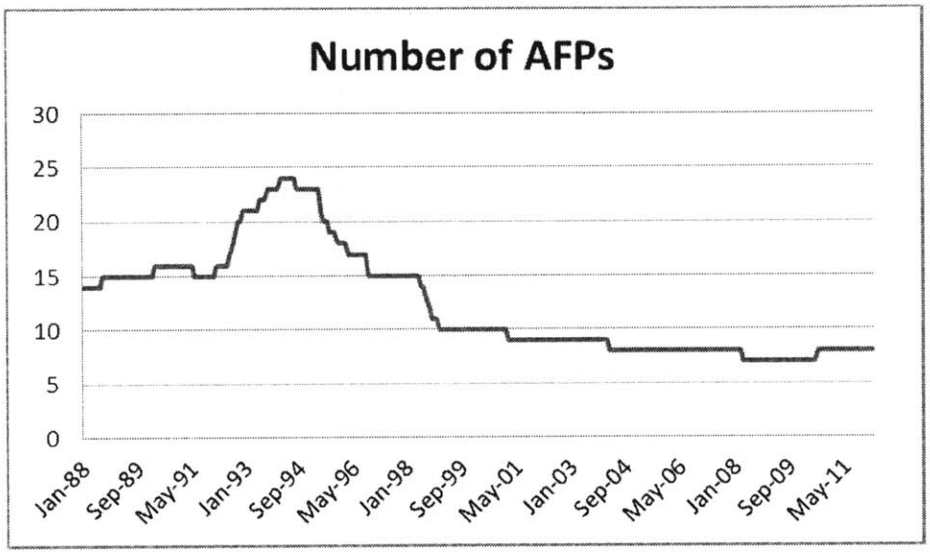


Figure 2.1: Number of Firms

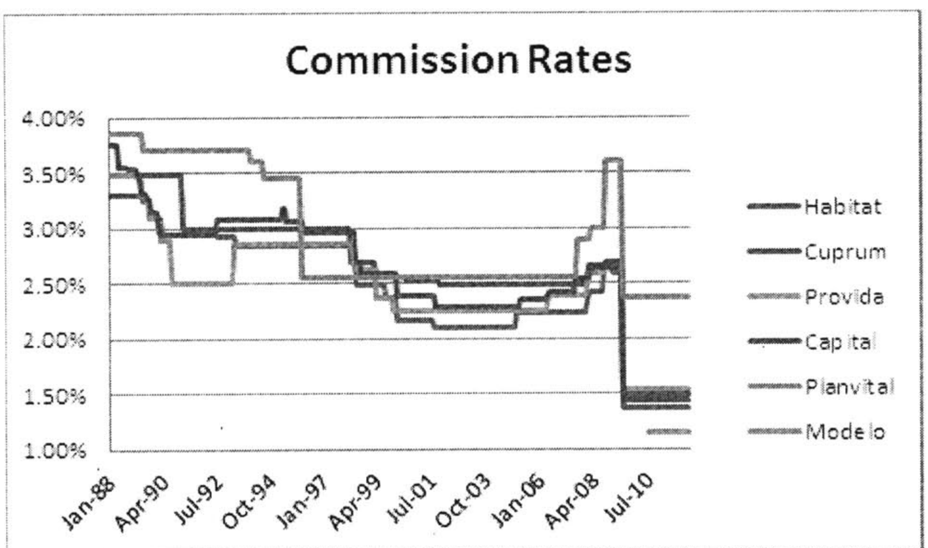


Figure 2.2: Commission Rates

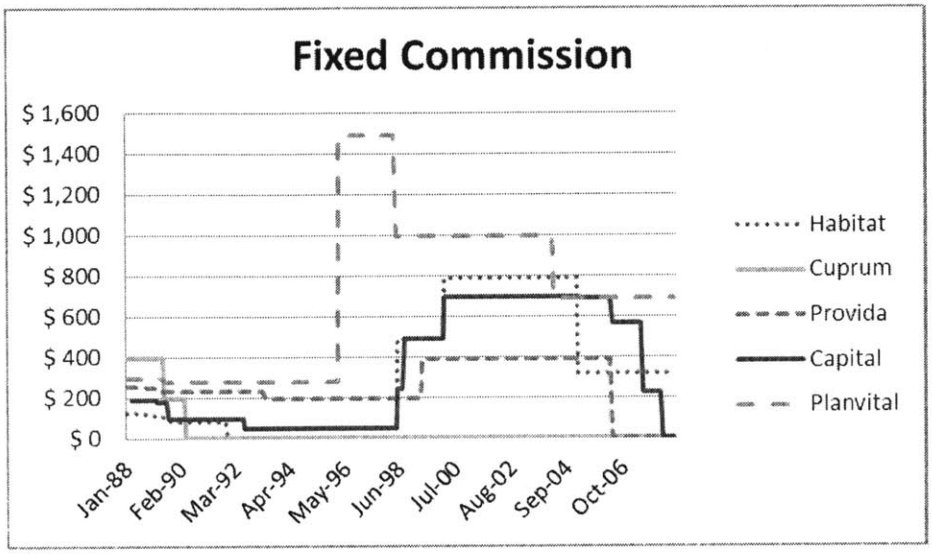


Figure 2.3: Commission Rates

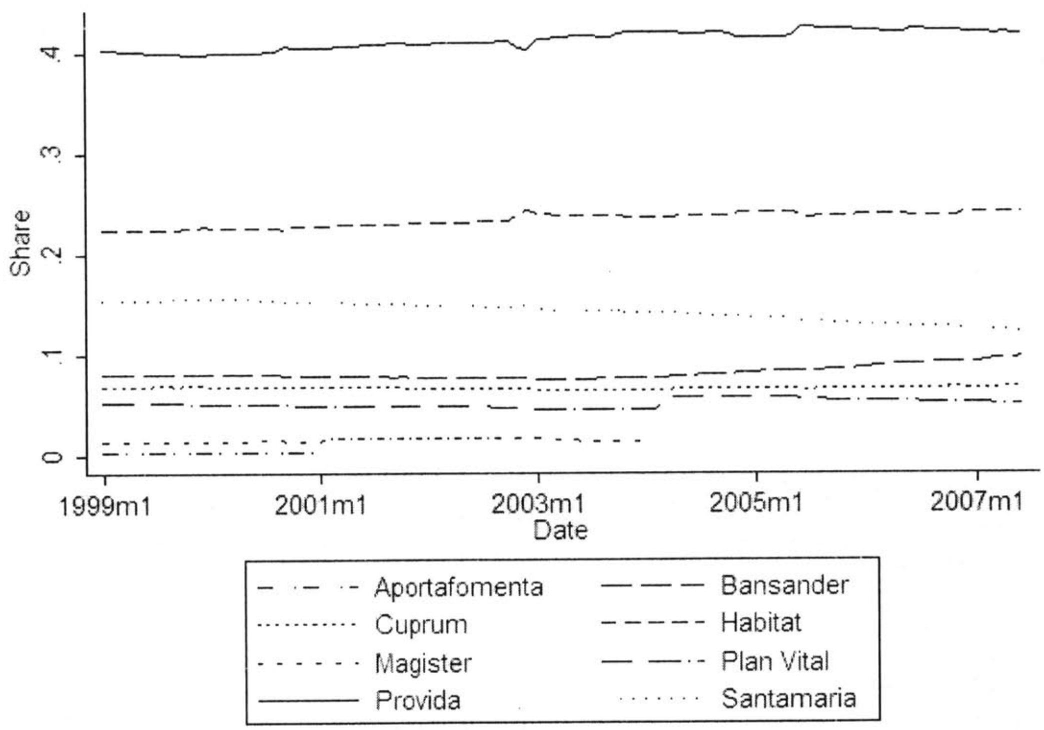


Figure 2.4: Market Shares, by Firm

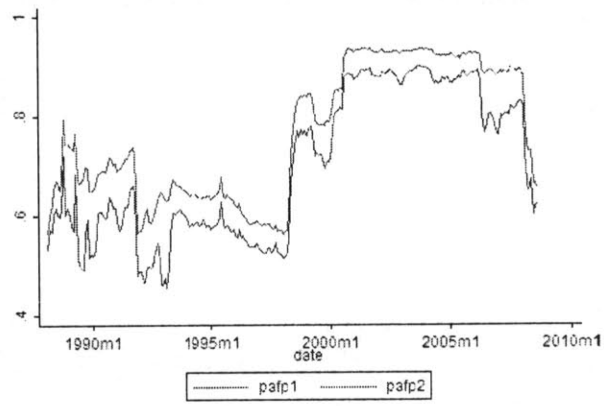


Figure 2.5: Matched Fraction of Individuals

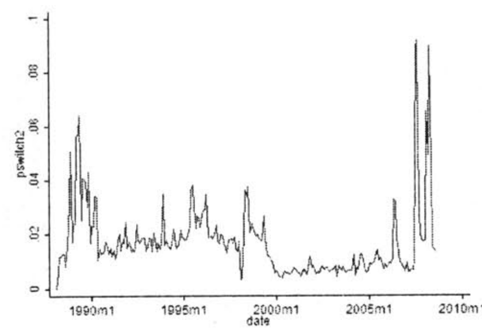


Figure 2.6: Fraction of Individuals who Switch

Tables

1988-2009 Sample		
Characteristic	Matched	Unmatched
Wage	233498.2	174647.5
	(250007.3)	(230582.9)
Age	36.8	37.8
	(10.6)	(14.1)
Female	0.363	0.379
	(.481)	(0.485)
Estimation Sample		
Characteristic	Matched	Unmatched
Wage	289521.6	178817.0
	(268020.6)	(258777.6)
Age	37.8	46.0
	(10.5)	(16.4)
Female	0.406	0.377
	(.491)	(0.484)

Table 2.1: Balance of Observable Characteristics for Matched and Unmatched Samples

	1	2	3	4	5	6	7	8
	Estimated Shares				Actual Shares			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
Fund A Return		20.22***		19.11***		21.17***		19.59***
		(2.370)		(2.208)		(2.979)		(2.783)
Fund B Return		-25.23***		-23.17***		-27.14***		-23.95***
		(3.739)		(3.510)		(4.699)		(4.425)
Fund C Return	0.0144	2.624	0.0647	2.476	0.539***	5.556**	0.634***	5.102**
	(0.267)	(2.044)	(0.262)	(1.891)	(0.164)	(2.569)	(0.158)	(2.383)
Fund D Return		-1.990		-2.852		-2.040		-3.518
		(3.750)		(3.483)		(4.713)		(4.391)
Fund E Return		-1.727		-1.574		-4.814		-4.315
		(2.393)		(2.213)		(3.007)		(2.789)
Variable Commission Rate	-546.8***	-540.0***	-621.1***	-559.9***	-549.8***	-492.0***	-605.3***	-519.2***
	(17.13)	(14.54)	(19.36)	(15.10)	(10.55)	(18.27)	(11.65)	(19.03)
Fixed Commission	-0.000674***	-0.00121***	-9.20e-06	-0.00119***	-0.000691***	-0.000735***	-0.000505***	-0.000705***
	(0.000105)	(6.92e-05)	(0.000122)	(7.14e-05)	(6.48e-05)	(8.69e-05)	(7.37e-05)	(9.00e-05)
Observations	696	349	681	334	696	349	681	334
R-squared	0.675	0.892	0.599	0.886	0.836	0.801	0.812	0.789
F Test of Excluded Instruments			190.5	138.5			190.5	138.5

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.2: Conditional Logit Estimates, Using Estimated Shares and Actual Shares

Estimated Shares	Chi Squared	P Value
Only Fund C	64.46	1.01e-14
All Funds	20.67	2.10e-03
Actual Shares	Chi Squared	P Value
Only Fund C	126.26	3.83e-28
All Funds	22.89	0.000834

Table 2.3: Hausman Tests of the Firm Level Conditional Logit Estimates

	1	2	3	4	5	6	7	8
Fund C Return	Aportafomenta	Bansander	Cuprum	Habitat	Magister	Plan Vital	Provida	Santamaria
Aportafomenta	0.003***	-0.003***	-0.003***	-0.002***	-0.003***	-0.003***	-0.002***	-0.003***
	(9.17e-07)	(7.52e-07)	(7.61e-07)	(3.89e-07)	(8.89e-07)	(8.46e-07)	(5.24e-07)	(6.99e-07)
Bansander		0.049***	-0.049***	-0.035***	-0.053***	-0.052***	-0.041***	-0.047***
		(2.16e-04)	(2.19e-04)	(1.11e-04)	(2.55e-04)	(2.43e-04)	(1.50e-04)	(2.01e-04)
Cuprum			0.046***	-0.032***	-0.049***	-0.048***	-0.038***	-0.044***
			(1.95e-04)	(9.90e-05)	(2.28e-04)	(2.17e-04)	(1.33e-04)	(1.79e-04)
Habitat				0.123***	-0.187***	-0.183***	-0.144***	-0.166***
				(0.001)	(0.003)	(0.003)	(0.002)	(0.003)
Magister					0.011***	-0.011***	-0.008***	-0.010***
					(1.10e-05)	(1.05e-05)	(6.50e-06)	(8.69e-06)
Plan Vital						0.022***	-0.017***	-0.020***
						(4.94e-05)	(3.06e-05)	(4.08e-05)
Provida							0.101***	-0.116***
							(0.001)	(0.001)
Santamaria								0.062***
								(3.51e-04)
Variable Commission Rate								
Aportafomenta	-3.240***	2.931***	2.955***	2.103***	3.191***	3.120***	2.449***	2.828***
	(0.02289)	(0.01808)	(0.01802)	(0.01314)	(0.02173)	(0.02063)	(0.01416)	(0.01710)
Bansander		-49.822***	50.228***	35.752***	54.244***	53.026***	41.627***	48.064***
		(0.38927)	(0.64645)	(0.03964)	(0.49366)	(0.44676)	(0.36847)	(0.33771)
Cuprum			-46.576***	33.152***	50.300***	49.170***	38.600***	44.569***
			(1.87340)	(1.47919)	(2.69948)	(2.73809)	(1.23016)	(2.28071)
Habitat				-125.848***	190.942***	186.655***	146.531***	169.188***
				(10.10245)	(41.84682)	(40.03678)	(22.87486)	(32.38274)
Magister					-11.288***	11.034***	8.662***	10.002***
					(0.06066)	(0.05628)	(0.04695)	(0.04683)
Plan Vital						-22.601***	17.743***	20.486***
						(0.38918)	(0.33037)	(0.32821)
Provida							-102.726***	118.610***
							(5.76895)	(10.63517)
Santamaria								-63.203***
								(1.02297)
Fixed Commission Rate								
Aportafomenta	-4.08e-06***	3.69e-06***	3.72e-06***	2.65e-06***	4.01e-06***	3.92e-06***	3.08e-06***	3.56e-06***
	(1.98e-13)	(1.61e-13)	(1.51e-13)	(8.96e-14)	(1.92e-13)	(7.41e-14)	(1.09e-13)	(1.51e-13)
Bansander		-0.00006***	0.00006***	0.00005***	0.00007***	0.00007***	0.00005***	0.00006***
		(2.84e-11)	(2.59e-11)	(1.58e-11)	(3.29e-11)	(3.21e-11)	(1.77e-11)	(2.66e-11)
Cuprum			-0.00006***	0.00004***	0.00006***	0.00006***	0.00005***	0.00006***
			(5.13e-11)	(2.96e-11)	(6.32e-11)	(6.13e-11)	(3.51e-11)	(5.06e-11)
Habitat				-0.00016***	0.00024***	0.00023***	0.00018***	0.00021***
				(2.09e-10)	(4.59e-10)	(4.49e-10)	(2.45e-10)	(3.70e-10)
Magister					-0.00001***	0.00001***	0.00001***	0.00001***
					(1.79e-12)	(1.74e-12)	(9.97e-13)	(1.44e-12)
Plan Vital						-0.00003***	0.00002***	0.00003***
						(4.46e-12)	(2.50e-12)	(3.72e-12)
Provida							-0.00013***	0.00015***
							(1.72e-10)	(2.54e-10)
Santamaria								-0.00008***
								(4.68e-11)

Standard Errors In Parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.4: Marginal Effects for the Conditional Logit Model, Using Actual Shares

	1	2	3	4	5	6	7	8
Fund C Return	Aportafomenta	Bansander	Cuprum	Habitat	Magister	Plan Vital	Provida	Santamaria
Aportafomenta	0.003***	-0.002***	-0.002***	-0.002***	-0.003***	-0.003***	-0.002***	-0.002***
	(4.19e-07)	(3.44e-07)	(3.59e-07)	(1.61e-07)	(4.08e-07)	(3.86e-07)	(2.41e-07)	(3.19e-07)
Bansander		0.056***	-0.057***	-0.038***	-0.061***	-0.059***	-0.047***	-0.054***
		(1.87e-04)	(1.96e-04)	(8.54e-05)	(2.22e-04)	(2.10e-04)	(1.31e-04)	(1.73e-04)
Cuprum			0.045***	-0.030***	-0.048***	-0.046***	-0.037***	-0.042***
			(1.26e-04)	(5.57e-05)	(1.43e-04)	(1.36e-04)	(8.39e-05)	(1.12e-04)
Habitat				0.151***	-0.242***	-0.236***	-0.186***	-0.214***
				(0.001)	(0.004)	(0.003)	(0.002)	(0.003)
Magister					0.011***	-0.010***	-0.008***	-0.009***
					(6.72e-06)	(6.36e-06)	(3.97e-06)	(5.25e-06)
Plan Vital						0.025***	-0.020***	-0.022***
						(4.17e-05)	(2.61e-05)	(3.44e-05)
Provida							0.117***	-0.134***
							(0.001)	(0.001)
Santamaria								0.073***
								(3.20e-04)
Variable Commission Rate								
Aportafomenta	-2.498***	2.263***	2.317***	1.535***	2.466***	2.406***	1.897***	2.178***
	(0.01893)	(0.01499)	(0.01535)	(0.00982)	(0.01812)	(0.01711)	(0.01135)	(0.01405)
Bansander		-53.379***	54.649***	36.191***	58.155***	56.748***	44.731***	51.359***
		(0.40489)	(0.71880)	(0.04123)	(0.51550)	(0.45572)	(0.46963)	(0.34794)
Cuprum			-42.647***	28.243***	45.383***	44.284***	34.907***	40.079***
			(2.19874)	(1.57461)	(2.97595)	(2.98689)	(1.38957)	(2.48011)
Habitat				-143.770***	231.022***	225.432***	177.694***	204.026***
				(12.38052)	(64.55674)	(61.75933)	(37.54166)	(50.23037)
Magister					-10.119***	9.874***	7.783***	8.937***
					(0.07420)	(0.06869)	(0.05230)	(0.05633)
Plan Vital						-23.685***	18.670***	21.436***
						(0.48371)	(0.42097)	(0.40265)
Provida							-111.632***	128.174***
							(6.50376)	(12.14881)
Santamaria								-69.298***
								(1.15236)
Fixed Commission Rate								
Aportafomenta	-2.08e-06***	1.89e-06***	1.93e-06***	1.28e-06***	2.06e-06***	2.01e-06***	1.58e-06***	1.82e-06***
	(1.02e-13)	(2.30e-14)	(2.31e-14)	(1.42e-14)	(2.80e-14)	(2.55e-14)	(1.71e-14)	(2.14e-14)
Bansander		-0.00004***	0.00005***	0.00003***	0.00005***	0.00005***	0.00004***	0.00004***
		(3.75e-11)	(3.69e-11)	(1.84e-11)	(4.39e-11)	(4.24e-11)	(2.44e-11)	(3.49e-11)
Cuprum			-0.00004***	0.00002***	0.00004***	0.00004***	0.00003***	0.00003***
			(4.17e-11)	(2.02e-11)	(4.90e-11)	(4.71e-11)	(2.77e-11)	(3.87e-11)
Habitat				-0.00012***	0.00019***	0.00019***	0.00015***	0.00017***
				(2.95e-10)	(7.21e-10)	(6.97e-10)	(3.99e-10)	(5.73e-10)
Magister					-8.44e-06***	8.23e-06***	6.49e-06***	7.45e-06***
					(1.58e-12)	(1.53e-12)	(8.93e-13)	(1.26e-12)
Plan Vital						-0.00002***	0.00002***	0.00002***
						(6.07e-12)	(3.51e-12)	(5.01e-12)
Provida							-0.00009***	0.00011***
							(2.08e-10)	(2.96e-10)
Santamaria								-0.00006***
								(6.39e-11)

Standard Errors In Parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.5: Marginal Effects for the Instrumental Variables Conditional Logit Model, Using Actual Shares

	1	2	3	4	5	6	7	8
Fund C Return	Aportafomenta	Bansander	Cuprum	Habitat	Magister	Plan Vital	Provida	Santamaria
Aportafomenta	0.00009***	-0.00008***	-0.00008***	-0.00006***	-0.00009***	-0.00008***	-0.00007***	-0.00008***
	(0.000003)	(0.000002)	(0.000002)	(0.000001)	(0.000002)	(0.000002)	(0.000001)	(0.000002)
Bansander		0.00132**	-0.00133**	-0.00095***	-0.00144**	-0.00141**	-0.00110***	-0.00127**
		(0.000596)	(0.000608)	(0.000308)	(0.000708)	(0.000680)	(0.000418)	(0.000555)
Cuprum			0.00122**	-0.00087***	-0.00132**	-0.00129**	-0.00101***	-0.00117**
			(0.000509)	(0.000258)	(0.000594)	(0.000570)	(0.000350)	(0.000465)
Habitat				0.00330	-0.00500	-0.00490	-0.00384	-0.00442
				(0.003723)	(0.008561)	(0.008215)	(0.005051)	(0.006711)
Magister					0.00030***	-0.00029***	-0.00023***	-0.00027***
					(0.000031)	(0.000030)	(0.000018)	(0.000024)
Plan Vital						0.00057***	-0.00045***	-0.00051***
						(0.000111)	(0.000068)	(0.000091)
Provida							0.00269	-0.00310
							(0.002476)	(0.003290)
Santamaria								0.00167*
								(0.000950)
Variable Commission Rate								
Aportafomenta	-3.281***	2.964***	2.993***	2.131***	3.231***	3.165***	2.482***	2.861***
	(0.062)	(0.049)	(0.049)	(0.035)	(0.058)	(0.056)	(0.038)	(0.046)
Bansander		-50.094***	50.578***	36.010***	54.599***	53.487***	41.938***	48.341***
		(1.052)	(1.743)	(0.107)	(1.337)	(1.211)	(0.993)	(0.913)
Cuprum			-46.278***	32.949***	49.958***	48.940***	38.373***	44.232***
			(4.881)	(3.849)	(7.025)	(7.133)	(3.208)	(5.930)
Habitat				-125.120***	189.711*	185.847*	145.719**	167.967**
				(26.578)	(109.720)	(105.107)	(60.049)	(84.746)
Magister					-11.375***	11.144***	8.737***	10.071***
					(0.161)	(0.151)	(0.125)	(0.124)
Plan Vital						-21.595***	16.932***	19.517***
						(0.942)	(0.793)	(0.786)
Provida							-102.034***	117.612***
							(15.146)	(27.784)
Santamaria								-63.227***
								(2.734)
Fixed Commission Rate								
Aportafomenta	-4.04e-06***	3.65e-06***	3.69e-06***	2.63e-06***	3.98e-06***	3.90e-06***	3.06e-06***	3.53e-06***
	(5.34e-13)	(4.34e-13)	(4.07e-13)	(2.42e-13)	(5.16e-13)	(4.70e-13)	(2.94e-13)	(4.06e-13)
Bansander		-0.00006***	0.00006***	0.00004***	0.00007***	0.00007***	0.00005***	0.00006***
		(7.68e-11)	(7.07e-11)	(4.28e-11)	(8.93e-11)	(8.75e-11)	(4.82e-11)	(7.20e-11)
Cuprum			-0.00006***	0.00004***	0.00006***	0.00006***	0.00005***	0.00005***
			(1.33e-10)	(7.67e-11)	(1.64e-10)	(1.60e-10)	(9.15e-11)	(1.31e-10)
Habitat				-0.00015***	0.00023***	0.00023***	0.00018***	0.00021***
				(5.51e-10)	(1.21e-09)	(1.19e-09)	(6.50e-10)	(9.75e-10)
Magister					-0.00001***	0.00001***	0.00001***	0.00001***
					(4.83e-12)	(4.71e-12)	(2.69e-12)	(3.88e-12)
Plan Vital						-0.00003***	0.00002***	0.00002***
						(1.09e-11)	(6.14e-12)	(9.08e-12)
Provida							-0.00013***	0.00014***
							(4.51e-10)	(6.61e-10)
Santamaria								-0.00008***
								(1.26e-10)

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.6: Marginal Effects for the Conditional Logit Model, Using Estimated Shares

	1	2	3	4	5	6	7	8
Fund C Return	Aportafomenta	Bansander	Cuprum	Habitat	Magister	Plan Vital	Provida	Santamaria
Aportafomenta	0.00024***	-0.00021***	-0.00022***	-0.00014***	-0.00023***	-0.00023***	-0.00018***	-0.00020***
	(0.000001)	(0.000001)	(0.000001)	(0.000000)	(0.000001)	(0.000001)	(0.000001)	(0.000001)
Bansander		0.00597***	-0.00627***	-0.00389***	-0.00654***	-0.00636***	-0.00517***	-0.00573***
		(0.000582)	(0.000643)	(0.000247)	(0.000700)	(0.000660)	(0.000437)	(0.000537)
Cuprum			0.00349***	-0.00216***	-0.00364***	-0.00353***	-0.00287***	-0.00318***
			(0.000199)	(0.000077)	(0.000217)	(0.000205)	(0.000135)	(0.000166)
Habitat				0.01572***	-0.02642**	-0.02568**	-0.02088***	-0.02314***
				(0.004054)	(0.011472)	(0.010829)	(0.007160)	(0.008800)
Magister					0.00104***	-0.00101***	-0.00082***	-0.00091***
					(0.000018)	(0.000017)	(0.000011)	(0.000013)
Plan Vital						0.00273***	-0.00222***	-0.00246***
						(0.000123)	(0.000081)	(0.000100)
Provida							0.01121***	-0.01242***
							(0.002062)	(0.002535)
Santamaria								0.00773***
								(0.000977)
Variable Commission Rate								
Aportafomenta	-2.265***	2.040***	2.144***	1.330***	2.237***	2.173***	1.767***	1.959***
	(0.045)	(0.035)	(0.039)	(0.022)	(0.043)	(0.040)	(0.028)	(0.033)
Bansander		-57.258***	60.170***	37.335***	62.776***	61.004***	49.602***	54.974***
		(0.968)	(1.661)	(0.128)	(1.222)	(1.081)	(1.266)	(0.826)
Cuprum			-33.444***	20.752***	34.893***	33.908***	27.570***	30.557***
			(4.236)	(2.503)	(5.244)	(5.218)	(2.626)	(4.257)
Habitat				-150.783***	253.526	246.368	200.321	222.019
				(32.311)	(197.892)	(189.750)	(122.117)	(153.633)
Magister					-9.942***	9.661***	7.855***	8.706***
					(0.213)	(0.194)	(0.152)	(0.158)
Plan Vital						-26.151***	21.263***	23.566***
						(1.518)	(1.415)	(1.270)
Provida							-107.523***	119.169***
							(16.503)	(28.874)
Santamaria								-74.120***
								(2.999)
Fixed Commission Rate								
Aportafomenta	-3.36e-08***	3.02e-08***	3.18e-08***	1.97e-08***	3.31e-08***	3.22e-08***	2.62e-08***	2.90e-08***
	(4.90e-15)	(4.32e-15)	(2.14e-15)	(3.19e-15)	(4.86e-15)	(2.61e-15)	(4.39e-15)	(4.18e-15)
Bansander		-8.48e-07***	8.92e-07***	5.53e-07***	9.30e-07***	9.04e-07***	7.35e-07***	8.15e-07***
		(1.27e-10)	(1.40e-10)	(5.41e-11)	(1.53e-10)	(1.44e-10)	(9.53e-11)	(1.17e-10)
Cuprum			-4.96e-07***		5.17e-07***		4.09e-07***	4.53e-07***
			(5.64e-11)		(2.76e-09)		(2.13e-09)	(2.74e-09)
Habitat				-2.23e-06***	3.76e-06***	3.65e-06***	2.97e-06***	3.29e-06***
				(8.83e-10)	(2.49e-09)	(2.36e-09)	(1.56e-09)	(1.91e-09)
Magister					-1.47e-07***	1.43e-07***	1.16e-07***	1.29e-07***
					(3.82e-12)	(3.61e-12)	(2.38e-12)	(2.93e-12)
Plan Vital						-3.88e-07***	3.15e-07***	3.49e-07***
						(2.64e-11)	(1.75e-11)	(2.15e-11)
Provida							-1.59e-06***	1.77e-06***
							(4.51e-10)	(5.54e-10)
Santamaria								-1.10e-06***
								(2.13e-10)

Standard Errors In Parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.7: Marginal Effects for the Instrumental Variables Conditional Logit Model, Using Estimated Shares

	1	2	3	4
Effective Commission	0.000335***	0.000311***	0.000357***	0.000334***
	(0.000003)	(0.000003)	(0.000004)	(0.000004)
Fund A Return			-2.328***	-2.490***
			(0.214)	(0.214)
Fund B Return			3.114***	3.042***
			(0.387)	(0.388)
Fund C Return	0.161***	0.150***	0.131	0.0735
	(0.016)	(0.016)	(0.241)	(0.241)
Fund D Return			0.389	0.851**
			(0.393)	(0.394)
Fund E Return			-0.164	-0.101
			(0.278)	(0.279)
Firm Specific Constant	Yes	Yes	Yes	Yes
Firm Specific Age Effect	No	Yes	No	Yes
Firm Specific Gender Effect	No	Yes	No	Yes
Observations	7,791,376	7,791,376	3,891,223	3,891,223

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.8: Conditional Logit Estimates with Individual Characteristics

	1	2	3	4
Effective Commission	-0.000118***	-0.000118***	-0.000171***	-0.000170***
	(0.000013)	(0.000013)	(0.000016)	(0.000016)
Fund A Return			-1.507*	-1.451
			(0.908)	(0.908)
Fund B Return			5.046***	4.964***
			(1.450)	(1.450)
Fund C Return	0.206**	0.205**	1.19	1.254
	(0.101)	(0.101)	(0.792)	(0.792)
Fund D Return			-7.338***	-7.356***
			(1.245)	(1.246)
Fund E Return			-1.218	-1.29
			(0.917)	(0.917)
Firm Specific Constant	Yes	Yes	Yes	Yes
Firm Specific Age Effect	No	Yes	No	Yes
Firm Specific Gender Effect	No	Yes	No	Yes
Observations	286,446	286,446	252,487	252,487

Standard Errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.9: Conditional Logit Estimates with Individual Characteristics, Restricted to Individuals earning more than 1MM a year

	1	2	3	4	5	6	7	8
Effective Commission	Aportafomenta	Bansander	Cuprum	Habitat	Magister	Plan Vital	Provida	Santamaria
Aportafomenta	1.16e-06***	-1.28e-07***	-8.20e-08***	-2.88e-07***	-1.47e-08***	-3.85e-08***	-4.73e-07***	-1.33e-07***
	(3.22e-08)	(3.69e-09)	(2.30e-09)	(8.01e-09)	(4.31e-10)	(1.09e-09)	(1.32e-08)	(3.71e-09)
Bansander		0.00003***	-2.34e-06***	-8.22e-06***	-4.20e-07***	-1.10e-06***	-1.35e-05***	-3.80e-06***
		(4.00e-07)	(3.25e-08)	(1.12e-07)	(6.20e-09)	(1.57e-08)	(1.85e-07)	(5.29e-08)
Cuprum			0.00002***	-5.28e-06***	-2.70e-07***	-7.06e-07***	-8.68e-06***	-2.44e-06***
			(1.96e-07)	(5.14e-08)	(3.28e-09)	(7.51e-09)	(8.47e-08)	(2.45e-08)
Habitat				0.00006***	-9.48e-07***	-2.48e-06***	-3.04e-05***	-8.56e-06***
					(1.12e-08)	(2.55e-08)	(2.86e-07)	(8.30e-08)
Magister					3.77e-06***	-1.27e-07***	-1.56e-06***	-4.38e-07***
						(1.60e-09)	(1.84e-08)	(5.28e-09)
Plan Vital						9.67e-06***	-4.07e-06***	-1.15e-06***
							(4.20e-08)	(1.21e-08)
Provida							0.00007***	-1.41e-05***
							(7.07e-07)	(1.37e-07)
Santamaria								0.00003***
								(3.02e-07)
Fund C Return								
Aportafomenta	0.00047***	-5.20e-05***	-3.34e-05***	-1.17e-04***	-5.99e-06***	-1.57e-05***	-1.93e-04***	-5.41e-05***
	(0.00006)	(0.00001)	(0.00000)	(0.00002)	(0.00000)	(0.00000)	(0.00003)	(0.00001)
Bansander		0.01202***	-9.54e-04***	-3.35e-03***	-1.71e-04***	-4.48e-04***	-5.50e-03***	-1.55e-03***
		(0.00162)	(0.00013)	(0.00045)	(0.00002)	(0.00006)	(0.00074)	(0.00021)
Cuprum			0.00806***	-2.15e-03***	-1.10e-04***	-2.88e-04***	-3.53e-03***	-9.94e-04***
			(0.00108)	(0.00029)	(0.00001)	(0.00004)	(0.00047)	(0.00013)
Habitat				0.02290***	-3.86e-04***	-1.01e-03***	-1.24e-02***	-3.49e-03***
				(0.00307)	(0.00005)	(0.00014)	(0.00166)	(0.00047)
Magister					0.00154***	-5.16e-05***	-6.34e-04***	-1.78e-04***
					(0.00021)	(0.00001)	(0.00009)	(0.00002)
Plan Vital						0.00394***	-1.66e-03***	-4.66e-04***
						(0.00053)	(0.00022)	(0.00006)
Provida							0.02965***	-5.73e-03***
							(0.00398)	(0.00077)
Santamaria								0.01245***
								(0.00167)

Bootstrap Standard Errors in Parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.10: Marginal Effects for the Conditional Logit with Individual Characteristics Model

	1	2	3	4	5	6	7	8
Effective Commission	Aportafomenta	Bansander	Cuprum	Habitat	Magister	Plan Vital	Provida	Santamaria
Aportafomenta	-1.12e-06***	1.48e-07***	2.96e-07***	3.01e-07***	1.82e-08***	1.68e-08***	2.33e-07***	1.04e-07***
	(3.09e-07)	(4.10e-08)	(8.19e-08)	(8.34e-08)	(5.21e-09)	(4.69e-09)	(6.46e-08)	(2.89e-08)
Bansander		-2.00e-05***	6.06e-06***	6.17e-06***	3.72e-07***	3.43e-07***	4.78e-06***	2.13e-06***
		(1.55e-06)	(4.78e-07)	(4.84e-07)	(4.04e-08)	(3.02e-08)	(3.77e-07)	(1.70e-07)
Cuprum			-3.39e-05***	0.00001***	7.44e-07***	6.86e-07***	9.55e-06***	4.26e-06***
			(2.78e-06)	(1.04e-06)	(8.42e-08)	(6.39e-08)	(8.07e-07)	(3.63e-07)
Habitat				-3.43e-05***	7.58e-07***	6.99e-07***	9.72e-06***	4.34e-06***
				(2.80e-06)	(8.55e-08)	(6.49e-08)	(8.19e-07)	(3.69e-07)
Magister					-2.78e-06***	4.22e-08***	5.87e-07***	2.62e-07***
					(3.10e-07)	(5.05e-09)	(6.64e-08)	(2.98e-08)
Plan Vital						-2.57e-06***	5.41e-07***	2.41e-07***
						(2.34e-07)	(5.04e-08)	(2.27e-08)
Provida							-2.88e-05***	3.36e-06***
							(2.36e-06)	(2.87e-07)
Santamaria								-1.47e-05***
								(1.22e-06)
Fund C Return								
Aportafomenta	0.00123*	-1.63e-04	-3.26e-04*	-3.32e-04*	-2.01e-05	-1.85e-05*	-2.57e-04*	-1.15e-04*
	(0.00075)	(0.00010)	(0.00020)	(0.00020)	(0.00001)	(0.00001)	(0.00016)	(0.00007)
Bansander		0.02206*	-6.68e-03*	-6.80e-03*	-4.11e-04*	-3.79e-04*	-5.27e-03*	-2.35e-03*
		(0.01169)	(0.00354)	(0.00361)	(0.00022)	(0.00020)	(0.00279)	(0.00125)
Cuprum			0.03741*	-1.36e-02*	-8.21e-04*	-7.57e-04*	-1.05e-02*	-4.70e-03*
			(0.01980)	(0.00719)	(0.00044)	(0.00040)	(0.00557)	(0.00249)
Habitat				0.03784*	-8.36e-04*	-7.71e-04*	-1.07e-02*	-4.78e-03*
				(0.02002)	(0.00044)	(0.00041)	(0.00567)	(0.00253)
Magister					0.00307*	-4.65e-05*	-6.48e-04*	-2.89e-04*
					(0.00163)	(0.00002)	(0.00034)	(0.00015)
Plan Vital						0.00284*	-5.97e-04*	-2.66e-04*
						(0.00149)	(0.00031)	(0.00014)
Provida							0.03174*	-3.71e-03*
							(0.01679)	(0.00196)
Santamaria								0.01621*
								(0.00858)

Bootstrap Standard Errors in Parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.11: Marginal Effects for the Conditional Logit with Individual Characteristics Model, Restricting Sample to Individuals Earning above 1MM

	1	2
	All Individuals	Over 1 MM
Effective Commission	.0003499***	-.0000551**
	(0.000007)	(0.000020)
Fund C Return	.1666514***	0.1383873
	(0.014)	(0.100255)
$\hat{v}$	-.0000148**	-.0000941***
	(0.000006)	(0.000024)
Firm Specific Constant	Yes	Yes
Firm Specific Age Effect	No	No
Firm Specific Gender Effect	No	No
Observations	7,791,376	286,446

Bootstrap Standard Errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.12: Instrumental Variables Conditional Logit Estimates with Individual Characteristics

Effective Commission	Aportafomenta	Bansander	Cuprum	Habitat	Magister	Planvital	Provida	Santamaria
Aportafomenta	5.01e-07***	-4.29e-08***	-3.65e-08***	-1.28e-07***	-5.57e-09***	-1.71e-08***	-2.11e-07***	-5.94e-08***
	(1.73e-08)	(1.50e-09)	(1.25e-09)	(4.42e-09)	(1.94e-10)	(5.73e-10)	(7.31e-09)	(2.08e-09)
Bansander		.0000273***	-2.18e-06***	-7.65e-06***	-3.32e-07***	-1.02e-06***	-.0000126***	-3.54e-06***
		(5.36e-07)	(4.21e-08)	(1.49e-07)	(6.44e-09)	(1.90e-08)	(2.51e-07)	(7.14e-08)
Cuprum			.0000236***	-6.52e-06***	-2.83e-07***	-8.69e-07***	-.0000107***	-3.02e-06***
			(4.41e-07)	(1.20e-07)	(5.21e-09)	(1.55e-08)	(2.04e-07)	(5.79e-08)
Habitat				.0000666***	-9.95e-07***	-3.05e-06***	-.0000377***	-.0000106***
				(1.26e-06)	(1.87e-08)	(5.45e-08)	(7.24e-07)	(2.06e-07)
Magister					3.84e-06***	-1.33e-07***	-1.64e-06***	-4.60e-07***
					(7.23e-08)	(2.37e-09)	(3.11e-08)	(8.94e-09)
Planvital						.0000115***	-5.02e-06***	-1.41e-06***
						(2.09e-07)	(9.32e-08)	(2.63e-08)
Provida							.0000853***	-.0000174***
							(1.64e-06)	(3.48e-07)
Santamaria								.0000365***
								(7.14e-07)
Fund C Return								
Aportafomenta	.0002385***	-.0000204***	-.0000174***	-.0000611***	-2.65e-06***	-8.15e-06***	-.0001005***	-.0000283***
	(2.17e-05)	(1.86e-06)	(1.57e-06)	(5.59e-06)	(2.41e-07)	(7.32e-07)	(9.19e-06)	(2.57e-06)
Bansander		.0130228***	-.0010369***	-.0036443***	-.0001582***	-.0004857***	-.0059922***	-.001685***
		(1.11e-03)	(8.78e-05)	(3.12e-04)	(1.34e-05)	(4.09e-05)	(5.13e-04)	(1.44e-04)
Cuprum			.0112514***	-.0031058***	-.0001348***	-.000414***	-.0051066***	-.001436***
			(9.58e-04)	(2.65e-04)	(1.14e-05)	(3.48e-05)	(4.36e-04)	(1.22e-04)
Habitat				.0317356***	-.0004737***	-.001455***	-.0179485***	-.0050472***
				(2.73e-03)	(4.05e-05)	(1.24e-04)	(1.55e-03)	(4.33e-04)
Magister					.0018305***	-.0000631***	-.0007789***	-.000219***
					(1.56e-04)	(5.31e-06)	(6.66e-05)	(1.86e-05)
Planvital						.005491***	-.0023923***	-.0006727***
						(4.65e-04)	(2.03e-04)	(5.68e-05)
Provida							.040618***	-.0082988***
							(3.49e-03)	(7.13e-04)
Santamaria								.0173871***
								(1.49e-03)

Bootstrap Standard Errors in Parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.13: Marginal Effects for the Instrumental Variables Conditional Logit with Individual Characteristics Model

	1	2	3	4	5	6	7	8
Effective Commission	Aportafomenta	Bansander	Cuprum	Habitat	Magister	Planvital	Provida	Santamaria
Aportafomenta	-3.80e-08*	5.56e-09*	9.94e-09*	1.02e-08*	3.62e-10*	5.65e-10*	7.88e-09*	3.51e-09*
	(1.80e-08)	(2.64e-09)	(4.72e-09)	(4.80e-09)	(1.80e-10)	(2.71e-10)	(3.73e-09)	(1.64e-09)
Bansander		-6.87e-06**	2.10e-06**	2.15e-06**	7.67e-08**	1.20e-07**	1.67e-06**	7.44e-07**
		(2.44e-06)	(7.46e-07)	(7.66e-07)	(2.79e-08)	(4.28e-08)	(5.95e-07)	(2.65e-07)
Cuprum			-.0000106**	3.85e-06**	1.37e-07**	2.14e-07**	2.98e-06**	1.33e-06**
			(3.75e-06)	(1.36e-06)	(4.94e-08)	(7.58e-08)	(1.05e-06)	(4.70e-07)
Habitat				-.0000108**	1.40e-07**	2.19e-07**	3.05e-06**	1.36e-06**
				(3.81e-06)	(5.07e-08)	(7.78e-08)	(1.08e-06)	(4.82e-07)
Magister					-5.20e-07**	7.80e-09**	1.09e-07**	4.85e-08**
					(1.88e-07)	(2.82e-09)	(3.94e-08)	(1.76e-08)
Planvital						-8.06e-07**	1.70e-07**	7.56e-08**
						(2.87e-07)	(6.05e-08)	(2.69e-08)
Provida							-9.05e-06**	1.05e-06**
							(3.20e-06)	(3.74e-07)
Santamaria								-4.62e-06**
								(1.64e-06)
Fund C Return								
Aportafomenta	0.0000955	-0.000014	-0.000025	-0.0000256	-9.11e-07	-1.42e-06	-0.0000198	-8.83e-06
	(8.36e-05)	(1.22e-05)	(2.20e-05)	(2.23e-05)	(8.42e-07)	(1.23e-06)	(1.73e-05)	(7.69e-06)
Bansander		0.0172662	-0.0052873	-0.0054133	-0.0001927	-0.0003004	-0.0041899	-0.0018686
		(1.25e-02)	(3.83e-03)	(3.92e-03)	(1.47e-04)	(2.15e-04)	(3.04e-03)	(1.35e-03)
Cuprum			0.0267237	-0.0096872	-0.0003448	-0.0005376	-0.0074978	-0.0033439
			(1.93e-02)	(7.01e-03)	(2.62e-04)	(3.84e-04)	(5.43e-03)	(2.42e-03)
Habitat				0.0271297	-0.000353	-0.0005504	-0.0076765	-0.0034236
				(1.97e-02)	(2.69e-04)	(3.94e-04)	(5.57e-03)	(2.48e-03)
Magister					0.0013061	-0.0000196	-0.0002732	-0.0001219
					(9.94e-04)	(1.47e-05)	(2.09e-04)	(9.28e-05)
Planvital						0.0020255	-0.000426	-0.00019
						(1.45e-03)	(3.05e-04)	(1.36e-04)
Provida							0.022733	-0.0026499
							(1.65e-02)	(1.92e-03)
Santamaria								0.0116068
								(8.41e-03)

Bootstrap Standard Errors in Parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.14: Marginal Effects for the Instrumental Variables Conditional Logit with Individual Characteristics Model, Restricting the Sample to those with Wages above 1MM

Chapter 3

Estimating the Effect of Potential Entry on Market Outcomes Using a Licensure Threshold

3.1 Introduction

[1]Understanding how firm entry affects competition is a central question in the Industrial Organization literature, and an important input for antitrust policy. The theoretical literature suggests that both the threat of entry and entry itself discipline firm behavior. This paper exploits a natural experiment in firm licensure from Washington state's 2012 privatization of liquor sales to identify causal effects of these entry forces. The key ingredient to our estimates is exogenous variation in the number of eligible licensees in local liquor markets, generated by a licensure threshold requirement on store size. Although privatized liquor markets in Washington state average fewer than five stores, we find that widening the pool of potential entrants has a small effect on pricing, but a significant effect on product offerings.

Until June 2012, Washington state held a local monopoly over all spirit sales, administered through

[1]We thank Glenn Ellison, Nancy Rose, and Paulo Somaini for their helpful suggestions and comments. This work uses data from the Kilts-Nielsen Data Center at the University of Chicago Booth School of Business. Information on availability and access to the data is available at research.ChicagoBooth.edu/nielsen. We gratefully acknowledge support from the George P. and Obie B. Shultz Dissertation Grant. Gaston Illanes also acknowledges support from the Lynde and Harry Bradley Foundation and from Conicyt Chile.

the Washington State Liquor Control Board (WSLCB).[2] The WSLCB oversaw approximately 320 liquor outlets, each with standardized inventory and uniform prices. This regime is similar to the infrastructure in other Alcohol Beverage Control (ABC) states.[3] In November 2011, voters approved a ballot initiative (I-1183) to privatize liquor sales. Within a year, the WSLCB sold its inventory and the rights to take over existing liquor outlets at auction. Apart from former state liquor outlets, establishments with at least 10,000 square feet of retail space were allowed to sell liquor.[4] We exploit this threshold rule to estimate the impact of potential entry on market outcomes, in the spirit of a regression discontinuity design. Comparisons of markets with existing supermarkets just above and below the 10,000 square foot cutoff allow us to recover the effect of potential entry on prices and product variety.

We find that grocery stores just above the threshold are 30% more likely to sell liquor than those just below. However, large supermarkets (12,000+ square feet) are less likely to enter in markets with a grocery store just above versus just below the cutoff. On net, these forces combine so that we cannot distinguish a positive effect of potential entry on realized entry in liquor markets. In terms of conduct, we find that shifting a store above the threshold leads to a 3% decrease in transacted liquor prices. This effect is driven by differences in the product mix across markets, as within-product price comparisons show no effect on prices. That is, markets with an additional grocery store above the 10,000 square foot cutoff exhibit a shift towards cheaper products, rather than lower prices for a fixed set of goods. Since eligible stores need not obtain licensure, we interpret these results as the net effect of potential entry. This effect combines two mechanisms: the effect of entry on market outcomes and the effect of deterrence on outcomes.

This paper complements the existing empirical literature on entry, which chiefly adopts a structural approach to tackle endogeneity concerns. Structural models allow the authors to back-out market and firm primitives from observed equilibrium outcomes. These primitives are then used to simulate the effect of entry on market outcomes. In a seminal paper, Bresnahan and Reiss [1991] develop a structural model to study the effects of entry on competitive conduct. Focusing on a cross-section of geographically segregated markets, they provide evidence of sharply diminishing effects of entry on price levels. In a similar vein, Berry and Waldfogel [1999] conclude that free entry leads to an excessive number of entrants in radio broadcasting. They find that marginal firms provide little

[2] Beer and wine less than 24% ABV were excluded.

[3] Alabama, Idaho, Maine, Maryland, Mississippi, Montana, New Hampshire, North Carolina, Ohio, Oregon, Pennsylvania, Utah, Vermont, and Virginia.

[4] http://liq.wa.gov/transition/retailers

variety in music genres, but incur large operating costs. On the other hand, Syverson [2004] finds that average production efficiency is higher in markets with more competitors. Berry [1992] argues that heterogeneity across entrants can explain the relationship between profitability and number of firms. Other papers that focus on heterogeneity across entrants include Ciliberto and Tamer [2009] and Jia [2008].

A recent empirical literature on entry deterrence adopts a less structured approach. As an example, Ellison and Ellison [2011] find evidence of strategic investment by testing predictions from a model of the pharmaceutical industry. Their test of entry deterrence boils down to a test of non-monotonicity in strategic investment as a function of market size. Goolsbee and Syverson [2008] study the incumbent price response to potential entry by Southwest Airlines. They construct an event study of rivals' responses to Southwest's incorporation of new cities into their flight network. They find that rivals lower prices substantially when Southwest expands into the two airports that define the incumbent's route, even when Southwest has not announced any intention of serving that route.

Our results are significantly smaller than those found by Ellison and Ellison [2011] and Goolsbee and Syverson [2008]. This discrepancy hints at the importance of barriers to entry in the airline and pharmaceutical industries compared to liquor markets. As an example, Goolsbee and Syverson [2008] point out that the airline industry is fraught with dynamic demand considerations (such as frequent flyer programs), which are absent from liquor markets and might make entry deterrence differentially profitable. However, our results are consistent with earlier work. In particular, if we ignore the entry deterrence mechanism, we can construct an 2SLS estimate for the effect of entry on prices. Our estimates suggest that a mid-sized entrant leads to a 10% decline in transacted prices. As before, this effect is driven by differences in the product mix, as we find no effect on prices in within-product comparisons. These results are consistent with Bresnahan and Reiss [1991]'s finding that the main impact of entry on pricing comes from moving from monopoly to duopoly, as the markets we consider average 4.3 firms. More work on entry is needed to understand where and when entry and strategic investment loom large.

This setting also offers an opportunity to investigate the operating goals of state liquor control boards. Seim and Waldfogel [2013] and Miravete et al. [2014] suggest the Pennsylvania Liquor Control Board (PSCLB) expressly tries to reduce alcohol consumption through outlet location decisions and by setting markups above the profit-maximizing level. Their conclusions are based on demand estimates coupled with structural models of profiting-maximizing monopoly behavior. Like the PSCLB, the

WSLCB chose store locations and set uniform markups, so that Washington's deregulation provides an event study we can use to benchmark their estimates. As an example, their results suggest that moving from the standard ABC uniform markup pricing rule[5] to a monopolist setting product-specific markups leads to high price increases for rum and gin compared to vodka, whiskey and tequila. The outcome of liberalization in Washington state was, in fact, the opposite: the price of tequila rose most dramatically. However, our findings are consistent with their determination that the returns to third-degree price discrimination (tailoring markups to local demand conditions) are small compared to second-degree discrimination (tailoring markups to products). They suggest that administrative costs might make such complex pricing strategies unprofitable, and we find that the average coefficient of variation across products is slim. In other words, there is little within-product variation in prices across markets. This result foreshadows our results on the returns to entry.

The remainder of the paper is structured as follows: Section 2 discusses the data used in this study, Section 3 provides an overview of the liberalization policy and its effects, Section 4 describes our empirical methodology and presents our main results, and Section 5 concludes.

3.2 Descriptive Evidence on Deregulation

3.2.1 Background on Liberalization

Washington privatized liquor sales on June 1, 2012. It is the first (and so far, only) control state to deregulate since the end of Prohibition.[6] Costco spent over \$20 million promoting the reform, which passed with 57% of the vote in a November 2011 referendum.[7] The reform was marketed as a win-win. Consumers expected lower prices and greater product variety, while the state levied new taxes to compensate for their forgone profits from selling liquor themselves: a 17% tax on spirit retailers and a 10% tax on spirit distributors. On net, the initiative was touted as a means to increase state revenue.

Before liberalization, the WSLCB operated 166 stores (called State Liquor Stores, or SLS) and licensed an additional 162 contract stores (Contract Liquor Stores, or CLS). Contract stores were run by private citizens, but their actions were tightly circumscribed by the WSLCB. In particular, all stores maintained the same prices and product selection. The state acted as distributor and retailer, and charged a uniform markup of 51.9% on all liquor products. In addition, it charged a 20.5% alcohol sales tax and a \$3.7708 per liter tax.

[5]Which Washington state also followed, albeit with a higher markup than Pennsylvania (51.9% vs 30%).
[6]Angel Gonzalez. June 30, 2014. "In Aftermath of Privatization, Spirits Everywhere, Not Cheap." *Seattle Times.*
[7]Melissa Allison. November 8, 2011. "Voters Kick State Out of Liquor Business." *The Seattle Times.*

Several reform attempts preceded I-1183, most notably I-1180 in 2010. Also funded by Costco, it was defeated at the polls, 53% to 47%. There are two key differences between I-1183 and the unsuccessful I-1180. First, initiative 1183 added a 10% tax on distributors and a 17% tax on retailers, aimed at bolstering government revenue.[8] Nonetheless, proponents of reform argued "some prices are expected to drop, though not as low as in California, because Washington will keep its high liquor taxes."[9] These proponents hoped that competition would drive markups down enough to compensate for the tax hikes. A second difference was the I-1183 size restriction for licensure, as I-1180 had no such restriction. One central argument against I-1180 was the fear it would allow convenience stores to sell liquor, increasing the availability of cheap products and "spark(ing) an increase in alcohol-related crime and underage drinking".[10] As a response, the 10,000 square foot requirement, aimed at excluding small retailers, was introduced in 2011. The next section presents an identification strategy based on this discontinuity.[11]

3.2.2 Liberalization and Prices

First, we investigate whether privatization led to higher prices for consumers. This descriptive evidence not only helps us understand what happened in Washington at regulation, but also contributes to the literature on the merits of state-owned enterprises versus deregulation. We propose several indices to measure changes in the overall price level, and then explore whether demographics help explain variation in these changes across the state and across product categories.

From a theoretical perspective, it is unclear whether liberalization leads to higher or lower prices, and the empirical evidence is mixed. As an example, in a case study of deregulation in Mexico, La Porta and López-De-Silanes [1999] document a 5% increase in prices. In our context, there are several forces whose combined effect on prices is ex ante ambiguous. First, if private firms hold market power, then deregulation might lead to price increases. If competition is strong, however, we would expect prices to fall. Indeed, proponents of reform in Washington argued the private firms would charge markups far below the WSLCB's 51.9% level. Second, the new taxes implemented at deregulation ought to increase

[8]See:
http://taxfoundation.org/blog/bottoms-and-prices-too-washingtons-liquor-privatization-scheme-tax-hike
http://www.oregonlive.com/opinion/index.ssf/2012/07/washington_states_liquor_lesso.html
http://liq.wa.gov/stores/liquor-pricing

[9]Melissa Allison. November 8, 2011. "Voters Kick State Out of Liquor Business." *The Seattle Times.*

[10]Melissa Allison. July 18, 2011. "Costco revamps liquor-sales initiative." *The Seattle Times.*

[11]Note that while the text of the law allows for exceptions to this rule in "under-served areas", with the definition of this concept left to the judgment of the WSLCB, as of 2015 no store with less than 10,000 square feet has received a liquor license.

prices. Finally, as the state monopolist, the WSLCB contracted directly with distillers (rather than purchasing from distributors), and might have paid lower acquisition prices than retailers in the new private system. Indeed, local papers are rife with small retailer complaints that they lose out to monopsonistic firms like Costco.

We utilize price indices to compare prices before and after the reform. One challenge in price comparisons is that the state and the private market offer different product selections. The WSLCB data contain prices for all products, but the scanner data contain prices only for transacted goods. If consumers substitute away from expensive products, then the missing prices are not randomly selected. A naive comparison would therefore understate increases in offered prices. To deal with these issues, we follow the discussion in Chevalier and Kashyap [2014] and employ the Törnqvist price index (Törnqvist [1936]) to measure changes in price levels.

The Törnqvist index formula for a comparison of prices between t and $t-1$ is

$$I_t = \prod_{j=1}^{N} \left(\frac{p_{j,t}}{p_{j,t-1}} \right)^{\frac{1}{2}\left(\frac{p_{jt}q_{jt}}{\Sigma_i^N p_{it}q_{it}} + \frac{p_{j,t-1}q_{j,t-1}}{\Sigma_i^N p_{i,t-1}q_{i,t-1}} \right)}. \tag{3.1}$$

This index is a weighted average of the relative prices across products, where the weights are the average expenditure shares across the two periods. If consumers substitute away from goods that have high price increases, those products still receive substantial weight in the index if they were frequently purchased in the pre-period. Caves et al. [1982] show that this index approximates the ideal price index for a representative consumer with homothetic utility.

Figure 3.1 shows the monthly Törnqvist for all products from October 2010 to December 2012, and it is easy to see the dramatic level shift in prices at privatization, as the first week of June 2012 saw a 13.2% price increase relative to May. To be clear, the figure shows the month-to-month change, so this spike reads as an immediate increase that is sustained through the end of our dataset.

We observe 1,220 products sold in the last month before liberalization, and 721 products sold in the first week after (the frequency of our data changes from monthly to weekly at liberalization). As a result, the previous price change includes the effect of the dramatic drop in product variety, as products that are transacted in period $t-1$ but are not transacted in period t will be included in the Törnqvist calculation. One might be interested in a calculation taking only into account products that are sold both before and after liberalization. Figure 3.2 repeats the previous exercise for products that are sold every week during our sample period. For these 354 products, which we call the "State-balanced Panel", prices increase by 19.6%.

Second, we examine whether this average price increase masks heterogeneity either across the state or across products. Washington's deregulation provides an event-study counterpoint to Miravete et al. [2014], who find that state uniform markup rules subsidize poorer clientele. They also find that wealthy consumers have less elastic demand for liquor, so that a monopolist ought to charge higher markups for products that are more demanded by wealthier individuals. As an example, cheaper goods have a higher consumption share among the poor in Pennsylvania, and Miravete et al. [2014] predict a large increase in the markup of these products if the PSLCB behaved as a monopolist. We find a similar trend for liquor consumption under the WSLCB. Following their lead, we categorize products as "cheap" ("expensive") if the product is priced below (above) median for the its category. Panel d in Figure 3.6 shows that the share of expensive products increases with income. However, we do not find that liberalization leads to to disproportionate price increases in poorer areas or for cheaper products. Figure 3.3 presents a scatter plot of median income of the zip code and the Törnqvist price index at liberalization, while Figure 3.4 presents the same plot for the Törnqvist price index from liberalization to the end of 2012. While there is significant heterogeneity across the state, these indices appear uncorrelated with median income. Figures B.1 and B.2 in the Appendix repeat this exercise using zip code population, and again find no relationship.

We find some heterogeneity in price changes across liquor categories, but they do not correspond to the predictions in Miravete et al. [2014]'s monopoly model. These authors simulate the transition from a uniform markup rule to a product-specific markup for Pennsylvania, and predict higher price increases for rum and gin, and lower price increases for vodka, whiskey and tequila. In Washington, we observe the change from uniform markups to liberalization, where markups are tied to demand elasticities for the first time. Figure 3.5 reports the observed price changes in Washington state, by liquor category, for the state balanced panel. Figure B.3 in the Appendix repeats this exercise for the unbalanced panel. We find that tequila experiences the highest price increase, while the remaining categories have roughly the same change.

We also document significantly less heterogeneity in markups across products than that predicted by Miravete et al. [2014]. We infer product-level marginal cost using pre-liberalization prices net of the WSLCB's 51.9% markup. These prices are the WSLCB's acquisition costs. While the WSLCB, as a state monopsonist, ostensibly had access to low wholesale prices, the chain supermarkets in our sample are also likely to wield substantial bargaining power in upstream markets. The WSLCB's costs therefore ought to be a useful proxy for the acquisition cost faced by this set of retailers. However, we adjust the WSLCB wholesale prices to account for the new 10% distributor tax at liberalization, which may be passed through to retailers. We bound post-liberalization markups under the no- and perfect-pass through cases, and document how these bounds vary across product categories.

Table 3.1 reports percentage markups for the period between June 2012 and December 2012, and shows that markups are remarkably consistent across product categories. For example, the average markup for Whiskey is between 38% and 44%, while the average markup for Rum is between 37% and 43%. As before, there is no significant difference in markups for products classified as "cheap" or "expensive". These figures are in the ballpark of what Miravete et al. [2014] simulate a profit maximizing monopolist would charge in Pennsylvania (reported in the last columns of Table 3.1), with two key exceptions: Tequila should be priced more competitively (a mere 27% markup); and cheaper products should have higher markups (67% vs. 26%, with an average of 42%).

The discrepancies between the predictions in Miravete et al. [2014] and the facts we document in Washington state hint at the differences between monopoly and competitive second degree price discrimination. Differences in demand between Washington and Pennsylvania may also contribute to these disparities. As an example, Figure 3.6 reports market shares at the zip code level for the period between January 2012 and May 2012, broken down by observable characteristics. The first sub-figure reports shares by product category as the percentage who are a minority increases. Unlike the patterns found by Miravete et al. [2014] in Pennsylvania, there does not seem to be a significant gradient for any category across this dimension. The second sub-figure shows market shares by product category as the percentage college educated increases. Here we observe that whisky and rum consumption are negatively correlated with college education, while vodka and gin consumption are positively correlated with college education. These patterns are consistent with those found for Pennsylvania, albeit with shallower slopes in all cases. The third sub-figure repeats the analysis as the percentage of individuals who earn more than $50,000 a year ("High-income") increases. Here we also find an increasing pattern of consumption of vodka and gin, and a weakly negative association for rum and whisky. However, we note that even when the underlying consumption patterns between the two states are are similar, market outcomes diverge from their simulations.

3.2.3 Liberalization and Number of Stores

While deregulation brought higher prices for consumers, it has also meant increased availability of spirits. Indeed, opponents of reform feared an overabundance of liquor outlets, leading to hikes in underage consumption and driving accidents. Figure 3.7 shows that these fears have been partially realized: the number of liquor outlets state-wide increased from approximately 360 to over 1,400 stores within the first six months of privatization. Chamberlain [2014] documents a rise in neighborhood crime associated with liquor availability following deregulation .

Figure 3.8 is a scatterplot of liquor entry versus the number of supermarkets at the zip code level. There is a strong, positive relationship between the number of stores in the TDLinx data and the number of licensees recorded by the WSLCB. On average, there are 2.6 liquor outlets in each zip code, and 1.3 of these are grocers, superettes, or convenience stores. Under state control, zip codes averaged a mere 0.6 liquor stores. the dramatic increase in liquor outlets following deregulation hint at a central finding in Seim and Waldfogel [2013]: the state monopolist restricted stores to curb alcohol consumption.

3.3 Data

3.3.1 Pre-liberalization: Price and Quantity Data from the WSLCB.

We obtained pre-liberalization data from the WSLCB's public records. The WSLCB published monthly price lists, and each list contains the retail price, liquor taxes, liquor type, size, brand name and proof for every product offered in that month. We use data from these price lists from October 2010 to May 2012, for a total of 45,948 product-month observations. During this period 1,916 products were sold, and the average after-tax price was $21.70. Although the state sold malt beverages and wine, we focus on sales of hard liquor for tractability. Quantity sold is reported at the establishment level on a monthly basis, both for State Liquor Stores (SLS) and Contract Liquor Stores (CLS).[12]

[12]We drop instances of negative sales based on conversations with Melissa Norton at the WSLCB. These seem to be inventory adjustments triggered by state audits, and happens for 0.5% of the observations.

3.3.2 Post-liberalization: Grocery and Convenience Store Sizes and Licensure

Data on grocery and convenience store square footage comes from TDLinx, a subsidiary of Nielsen. For each establishment in Washington state, TDLinx provides store name, addresses, and square footage in January 2011 and December 2012. This data is crucial for constructing the set of eligible licensees based on the WSLCB's threshold rule.

Figure 3.9 is a histogram of stores sizes near the threshold in 2011 and 2012. Importantly, the number of stores with reported square footage just above 10,000 square feet does not change across these periods. Since I-1183 passed in November 2011, stores had no incentive to manipulate square footage in January 2011. The stability of store sizes allays concerns that supermarkets might expand in order to gain licensure.

We match data on store sizes to data on liquor licensure from January 2013 using store name and addresses. The WSLCB maintains a list of off-premises licensees on their website. Historical licensure records are taken from theWayBackMachine. The licensure files contain information on licensee addresses, trade names, license type (beer, wine or spirit) and licensure date. Our final dataset is an establishment-level database of entry and square footage.

3.3.3 Post-liberalization: Grocery Store Liquor Prices

We collect data on post-reform price and quantities from the Nielsen Retail Scanner dataset, available through the Kilts Center at the University of Chicago's Booth School of Business. The Retail Scanner database tracks all transactions at a set of unnamed supermarkets across the United States. The data contain average weekly prices, quantities, sales information, and product descriptions for each anonymous establishment. Products are tracked at the UPC level. We focus on the 678 scanner stores in Washington that record at least one liquor transaction in 2012. The average after-tax price of a liquor product is \$25.40 based on the sale of 1,525 unique UPCs in the six months following liberalization.

The Retail Scanner data allows us to investigate the pricing and product variety decisions of large supermarkets. All of these stores enter the liquor market only after privatization. Unfortunately, while Nielsen records prices at the establishment level, it obfuscates store identities. Only the FIPS county code is reported to researchers. This obfuscation poses a difficulty in measuring our left-hand-side variable. We cannot conduct a store-level analysis by matching establishment sales to square footage of licensees. Instead, we match Nielsen Scanner establishments to zip codes, and average across stores to find zip code-level prices for each product, each week. Our markets are therefore zip codes, and our main results consider the effect of entry on zip code outcomes.

We match stores to zip codes using shopping patterns from the Nielsen Panel dataset. This dataset tracks a panel of consumers, rather than stores. Each included household records all purchases by shopping trip, including store identifiers we match to the Retail Scanner dataset. We infer store locations based on the zip codes of households who shop there most often. We count the number of trips originating in each zip code and culminating in each store, and then assign stores the modal zip code across trips. Importantly, we only count trips before privatization, so that household choice of grocers should be independent of local liquor market competition. Stores with fewer than 10 trips are excluded to reduce noise. Figure 3.10 is a histogram of trips to each store, by zip code rank. Across stores, the modal zip code (most popular) originates 33 trips, while the second most-popular zip code originates a mere 12.5. This sharp decay in trips suggests the panelist data is informative about likely store locations.

Since we are working with transactions data, any product that is offered in a store but that is not sold during a certain week will be missing from our dataset. In fact, 53.8% of the UPCs we observe post liberalization are not sold in any store in the state for at least one week. This would be a potential concern if we were estimating demand, for example, as it would imply that products with a low unobserved preference value are less likely to enter our data. However, since our question of interest regards the effects of market structure on transacted prices, the fact that we are missing prices for goods that are not transacted is irrelevant. More concerning is the fact that we do not observe prices in stores outside the scope of the Nielsen database, and we will miss any goods that are sold by specialty liquor stores and not by supermarkets. We interpret the results that follow as identifying the responses of grocery stores that sell liquor to market structure, and make no claims about external validity of these estimates to the response of other liquor retailers.

3.4 Empirical Strategy and Results

3.4.1 10,000 Square Foot Licensure Requirement on Entry: Store Level

Our main empirical strategy is to compare entry and prices in markets with stores in the neighborhood of the 10,000 square foot licensure threshold. First, we look for evidence on whether stores just above the cutoff are more likely to acquire a liquor license than stores just below. While I-1183 allowed the WSLCB to make exceptions to the size requirement in underserved areas, in the time period we consider, the board had yet to exploit this loophole.[13][14] In this analysis, therefore, we treat the threshold as a hard cutoff.

Our identification strategy requires that stores in the neighborhood of the threshold pose a threat of entry. If, for example, entry in liquor markets were blockaded (incumbents need not investment to deter entry), then this natural experiment would be uninformative about the effects of entry deterrence and entry on market outcomes. Our first test is therefore whether stores just above the threshold take advantage of their eligibility to sell liquor. We use the licensure requirement to construct a regression discontinuity design as in Imbens and Lemieux [2008], where s denotes store:

$$L_s = \beta_0 + \beta_1 E_s + \beta_2 E_s \times (SQFT_s - 10) + \beta_3 (1 - E_s) \times (10 - SQFT_s) + \delta' X_s + \epsilon_s \quad (3.2)$$

L_s takes a value of 1 if store s is licensed, $SQFT_s$ is square footage, and E_s is an indicator that the store is eligible for a license (has at least 10,000 square feet of space). We estimate (3.2) for three separate bandwidths: 2,000 square feet of the cutoff (73 stores), 5,000 square feet of the cutoff (246 stores), and the entire dataset (3,969 stores). Since there are relatively few stores near the threshold, we allow for linear trends in size only for the full sample. X_s includes zip code level control variables, meant to capture characteristics about store s's competitors. Our full specification includes the number of (licensed and unlicensed) supermarkets in the same zip code within, above, and below the bandwidth. We estimate the coefficients from (3.2) using OLS with heteroskedasticity-robust standard errors.

3.4.2 10,000 Square Foot Licensure Requirement on Entry and Prices: Market Level

The WSLCB square footage requirement is a treatment on individual supermarket eligibility to sell liquor, but also on the number of eligible liquor outlets at the market level. Some markets have more (fewer) potential entrants into liquor sales because the existing supermarkets were just above (below) the threshold. The identification assumption is that markets with the same number of mid-

[13] Melissa Allison. November 8, 2011. "Voters Kick State Out of Liquor Business." *The Seattle Times*.

[14] Jordan Schrader. May 9, 2013. "Liquor Board Votes to Allow More Small Stores to Sell Hard Liquor." *The Olympian*. [http://www.theolympian.com/2013/05/09/2538437/liquor-board-votes-to-allow-more.html]

sized stores (stores with square footage within a fixed bandwidth around the cutoff), but different store size distributions within that bandwidth, are otherwise similar. The number of liquor entrants in market m is a sum, across groceries, of individual establishment entry decisions. We aggregate (3.2) across the S_m stores in market m to model liquor entry at the market level:

$$\begin{aligned}
NL_m &= \sum_{s=1}^{S_m} L_s + \omega_m \\
&= \sum_{s=1}^{S_m} (\beta_0 + \beta_1 E_s + \beta_2 E_s \times (SQFT_s - 10) + \beta_3(1 - E_s) \times (10 - SQFT_s) + \delta' X_s + \epsilon_s) + \omega_m \\
&= \alpha_0 + \beta_0 S_m + \beta_1 S_m \sum_{s=1}^{S_m} E_s + \beta_2 \sum_{s=1}^{S_m} (E_s \times (SQFT_s - 10)) + \beta_3 \sum_{s=1}^{S_m} ((1 - E_s) \times (10 - SQFT_s)) \\
&\quad + \delta' X_m + \omega_m \qquad (3.3) \\
&= \alpha_0 + \beta_0 S_m + \beta_1 NE_m + \beta_2 TSQFTA_m + \beta_3 TSQFTB_m + \delta' X_m + \eta_m \qquad (3.4)
\end{aligned}$$

where NL_m is the number of liquor outlets in market m, S_m is the total number of grocery stores in the bandwidth, NE_m is the number of eligible groceries in the bandwidth, and $TSQFTA_m$ and $TSQFTB_m$ are total square footage above and below the cutoff, respectively. X_m includes market controls, such as the total number of groceries or the number of groceries above- and below the threshold, depending on the specification.

We extend this specification to examine how the licensure requirement affects liquor prices. We adopt (3.4) to a panel structure. The price of product j in market m in week t is modeled as:

$$LogPrice_{jmt} \quad = \quad \alpha_0 + \beta_0 S_m + \beta_1 NE_m + \delta' X_m + \gamma_j + \alpha_t + \nu_{jmt} \qquad (3.5)$$

where γ_j are product characteristics, including size or product-level fixed effects, and α_t are month fixed effects. We cluster standard errors at the market level, to account for correlation in pricing across products.

Finally, under additional assumptions, we can extend this strategy to identify the causal effect of entry on pricing. We estimate the following model by 2SLS:

$$\begin{aligned} LogPrice_{jmt} &= \gamma_0 + \gamma_1 NL_m + \lambda' X_m + \gamma_j + \alpha_t + \nu_{jmt} \\ NL_m &= \alpha_0 + \beta_0 S_m + \beta_1 NE_m + \delta' X_m + \omega_m \end{aligned} \tag{3.6}$$

As before, the identifying assumption requires that conditional on the number of stores in the bandwidth, the number of these which are eligible is uncorrelated with ν_{jmt}. So long as eligibility affects licensure (a robust first stage), then we can exploit the threshold to estimate the causal effect of entry on prices.

However, we are skeptical of this identification argument for the two-stage least square estimates. Eligibility may affect prices not only through entry, but also through the threat of entry. For example, with demand or cost uncertainty, firms may attempt to deter entry by signaling market unprofitability (Milgrom and Roberts [1982]). In a limit-pricing model, potential entry (the number of eligible stores) might affect prices directly, and the number of eligible firms would constitute an omitted variable in the second-stage pricing equation (3.6). Only if potential entry affects prices solely through realized entry will the two-stage least squares estimates of (3.6) be the causal effect of entry on prices.

3.5 Results

3.5.1 Liquor Licensure by Square Footage

Figure 3.11a shows the relationship between square footage and the probability of licensure for stores between 4,500 and 19,499 square feet. The probability of licensure jumps approximately 30% between 9,000 and 10,000 square feet. This estimate likely understates the impact of eligibility on entry for two reasons. First, the TDLinx data bins store square footage, so that stores characterized as 10,000 square feet range, in fact, from 9,500-10,499 square feet. This means the 10s bin contains stores ineligible for licensure. Second, square footage is clearly measured with error. Since the 10,000 square foot requirement is strict, the probability of entry for smaller stores should be zero. The incidence of licensed small stores in our data points to errors in recording square footage, confirmed by Google Maps estimates of true store sizes. We therefore interpret the discontinuity at 10,000 square feet as a lower bound for the impact of eligibility on licensure.

As a robustness check, we construct the same threshold comparison for beer licensure. Even before I-1183, the WSLCB licensed grocers to sell beer, without regard for store size. As a result, there should be no jump in the probability of licensure for beer at the 10,000 square foot mark. Figure 3.11b shows the standardized probability of entry by square footage for beer and liquor separately. If anything, at the threshold, the probability of beer licensure falls. These results suggest that the jump in liquor licensure is not driven by underlying differences in unobserved store characteristics.

Table 3.2 shows the store-level regressions corresponding to Figure 3.11a and model (3.2). The coefficient on the 10,000 square foot indicator is statistically significant and positive across all specifications. Adding controls for the market configuration, such as the number of other stores within the market and their sizes, affect neither coefficient magnitudes nor significance. Widening the bandwidth from 2,000 to 5,000 or to all square footage increases the point estimates from approximately 30% to 40%. This is consistent with larger supermarkets being more likely to sell liquor, even apart from the 10,000 foot threshold.

3.5.2 Effect of Licensure on Entry at the Market Level

Table 3.3 reports the estimates of licensure on entry at the market level, which corresponds to the model in equation (3.4). Columns (1) through (5) employ a 2,000 square foot bandwidth around the 10,000 square foot cutoff. The coefficient on the number of firms in the bandwidth, but just above the threshold, is large, positive, and statistically significant. The estimates imply an additional store just above the threshold corresponds to 1/3 more liquor licensees (within the bandwidth) at the market-level. This result is consistent with the store-level regression results in Table 3.2, and is robust to including controls for the composition of other supermarkets in the market.

We also consider the effect of eligibility on the number of small (less than 8,000 square feet) and large (12,000+ square feet) liquor licensees (columns (3) and (4)). Stores below 8,000 square feet are ineligible for licensure, so there should be no effect of entry by mid-sized stores on the number of small licensed stores. Column (4) confirms a null effect; the point estimate is small and statistically insignificant. Column (3) considers the effect of mid-sized store eligibility on large store licensure. If mid-sized stores enter, they may crowd-out larger stores from liquor markets. The point estimate is negative, but statistically insignificant. Reassuringly, the coefficient on the number of large supermarkets is large, positive and statistically significant, implying that almost all large supermarkets (80.8%) choose to sell liquor. Column (5) reports the coefficient of mid-sized store eligibility on the total number of licensed grocers. The coefficient is positive (on the order of 15%) but statistically insignificant. This attenuation (compared to columns (1) and (2)) comes from the crowd-out effect reported in column

(3). Indeed, the coefficient estimate in column (5) is the sum of the estimates in (2)-(4).

Since there are relatively few stores within the 2,000 square foot bandwidth, we estimate the effect of licensure at the market-level using all stores, but controlling for linear trends in square footage. Columns (6) and (7) report the coefficient on the number of groceries above 10,000 square feet in size on the total number of licensed grocers and total number of licensees, respectively. The estimates imply that moving a grocery from just below to just above the 10,000 square foot threshold leads to another .5 liquor-selling grocers, on average. The point estimate in column (7) suggests a similar increase in the total number of liquor outlets, but the standard errors are too large to reject a null effect. As a whole, these results suggest that markets with a grocery just above, rather than just below the threshold have 0.3 more licensed grocers.

3.5.3 Effect of Licensure on Prices

In this section, we present results on the effect of eligibility on liquor prices. In this specification, each observation is a UPC - week - zip code combination, and we cluster standard errors at the zip code level. Column (1) in Table 3.4 reports the baseline regression; an additional eligible supermarket reduces prices approximately 3%. The effect is marginally significant (the p-value is .054). Column (2) reports the coefficient estimates when product (UPC) fixed effects are included. The coefficient on eligibility loses both economic and statistical significance. While the fixed effects specification uses only within-product variation, the baseline specification constructs cross-product comparisons. These results suggest that pricing differences across markets with different numbers of eligible licensees are driven by differences in product offerings, rather than by differences in prices. In particular, supermarkets offer cheaper products in markets with more eligible licensees, rather than cheaper prices for common products.

To test this theory, we divide the sample according to product popularity. Estimates for popular products – those sold most widely across the state – should be invariant to the inclusion of fixed effects. Columns (3) and (4) report effects for the 10% most popular products, with and without fixed effects respectively. The estimates confirm our interpretation of the coefficients reported in columns (1) and (2); there is no discernible effect of eligibility on prices in either specification. In contrast, column (5) shows an effect for less popular products, which is eviscerated by including fixed effects.

As a second test, we consider the effect of eligibility on product variety directly. We calculate the number of products (unique UPC codes) sold in each market during the six months following deregulation. Results, presented in column 5 of Table 3.6, suggest that an additional eligible firm leads to an increase of 60 products available in the market. This amounts to a 7% increase in product variety. If consumers value variety, then within-product comparisons understate the returns to competition.

Table 3.5 shows the effect of eligibility across product categories. We find the largest effects for vodka and rum, and no evidence that eligibility affects prices for gin. Miravete et al. [2014] suggest that demand elasticities differ substantially across these product categories, and that profit-maximizing markups ought to be highest for products with low-income, low-education clientele. While the consumption patterns across demographics in Pennsylvania do not mirror those in Washington, we find substantial evidence of heterogeneity across product categories. As an example, eligibility has a large effect on vodka prices and a null effect for gin prices, although both of these track with college education.

We also consider the effects of entry on price dispersion and quantity sold. We calculate the standard deviation of prices for each UPC in each week after privatization, and test whether markets with more eligible grocers exhibit higher price dispersion. The results, presented in table 3.6, show no effect of potential entry on this measure of dispersion. Second, we test whether the average sales per store decline with potential entry. If entrants engage chiefly in business-stealing, then we would expect that quantities decline at chain stores in markets with additional potential entrants. However, we find no affect on sales, even including product fixed effects. Since we have already established that prices for fixed products do not decline with potential entry, this suggests that the additional products do not cannibalize sales of staple products.

If the effect of eligibility on prices operates only through the entry channel, then we can construct IV estimates for the effect of entry on prices. This interpretation precludes forces such as entry deterrence. Our estimates imply that an additional entrant leads to a decline in transacted prices on the order of 10%. IV estimates of entry on average price are presented in Table 3.7. Column (1) presents the OLS estimates of log prices on the number of mid-sized groceries selling liquor at the market-level. The coefficient estimate is statistically significant, and suggests that entry leads to higher prices. Of course, a central concern in the OLS estimates is that entry is endogenous, so that markets with more firms have higher demand. This omitted variable might drive the positive correlation between entry and prices. Column (2) presents the baseline IV estimate, using the 2,000 square foot bandwidth. The coefficient on the number of licensed grocers is now large and negative (suggesting an additional entrant leads to a 10% decline in price), but the standard errors are wide.

A null effect of entry on prices for a fixed set of products is consistent with Bresnahan and Reiss [1991]'s non-experimental evidence of entry on tire prices. They cannot reject that markets with three-, four-, and five- firms have the same prices. On average, markets with grocers just below the threshold have 4.15 liquor outlets, so that an additional liquor outlet constitutes a shift from a four- to five- firm market. However, our results hint that entry affects product variety and therefore average prices. We find that markets with an additional eligible firm offer consumers more and cheaper choices.

3.6 Conclusion

Proponents of liquor deregulation in Washington state confronted a canonical challenge in designing institutions for a new private market: how to harness competition to improve market outcomes. Concern about alcohol-related crime prompted regulators to institute a 10,000 square foot licensure requirement to curtail entry, implicitly selecting the number and type of potential entrants into these fledgling markets across the state. We exploit this threshold rule to estimate causal effects of potential entry on market outcomes in the six months following privatization. Our findings that suggest an additional potential entrant lowers average prices by 3%. This effect represents a shift in the product mix towards cheaper goods; markets with an additional potential entrant have roughly 60 more products transacted. However, we find that prices for a fixed set of goods do not change, and that there are no effects of potential entry on price dispersion or on rivals' average sales.

These results contrast with the recent empirical literature on potential entry, which find larger effects. As an example, Goolsbee and Syverson [2008] document large price changes in airlines or and Ellison and Ellison [2011] find large effects on advertising decisions in pharmaceuticals. Our results point to the importance of interactions between potential entry and other market features. As an example, liquor retail sales involve smaller fixed costs and have fewer dynamic considerations than these two comparison industries.

Our results also provide policy implications from Washington state's experience with privatization of liquor retail. First, we find that the 10,000 square foot regulation appears binding (stores just above the cutoff are more likely to enter than those just below), but it does not significantly affect the overall number of liquor outlets within each market. Large supermarkets adjust their entry decisions depending on the eligibility of their mid-sized neighbors. Therefore, it appears the regulation chiefly affected the composition of liquor outlets in the market. While these findings suggest that extending the liquor franchise to smaller supermarkets would not dramatically increase liquor availability, the behavior of very small stores (for example, gas stations) need not conform to this result. Second, we find that markets with an additional potential entrant shift their product mix towards cheaper products. This confirms concerns that competition in liquor markets leads to greater availability of cheap alcohol, and suggests that regulation has an effect in limiting the availability of those types of products.

Exploiting credible exogenous variation to separately identify the effects of realized entry and entry deterrence is a important next step in unpacking the effects of potential entry we explore here. As an example, to test whether firms deter entry using limit pricing, researchers should study markets where (for exogenous reasons) potential entrants are differentially informed about market conditions. In a setting like ours, this might involve a comparison of chain versus independent stores. Researchers might also employ variation similar to ours to test the robustness of structural entry models, by comparing their predictions using pre-liberalization data to actual market outcomes. Such work would complement the growing entry games literature, in a fashion analogous to Peters [2006] and the merger simulation literature.

Figures

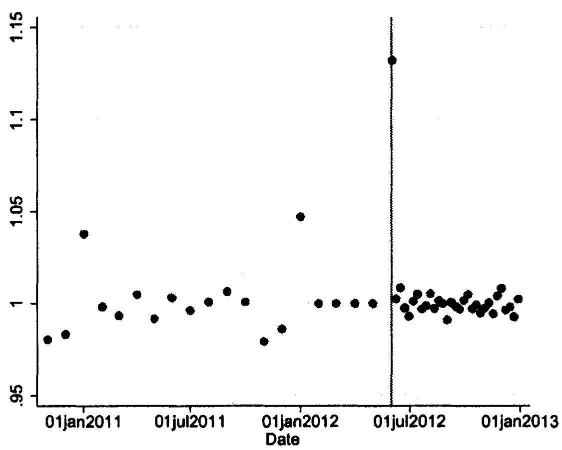


Figure 3.1: Törnqvist Price Index, Unbalanced Panel

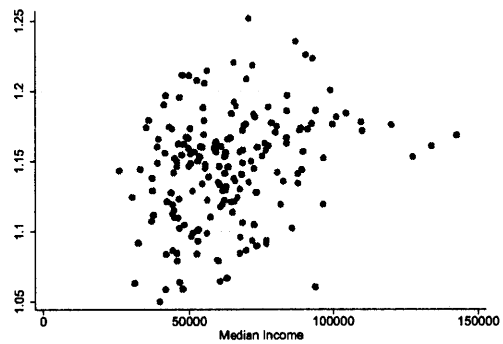


Figure 3.3: Törnqvist Price Index Change at Liberalization and Zip 5 Median Income, State-balanced Panel

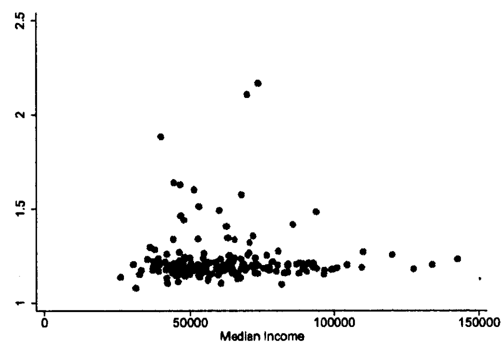


Figure 3.4: Törnqvist Price Index Change from Liberalization to End of 2012 and Zip 5 Median Income, State-balanced Panel

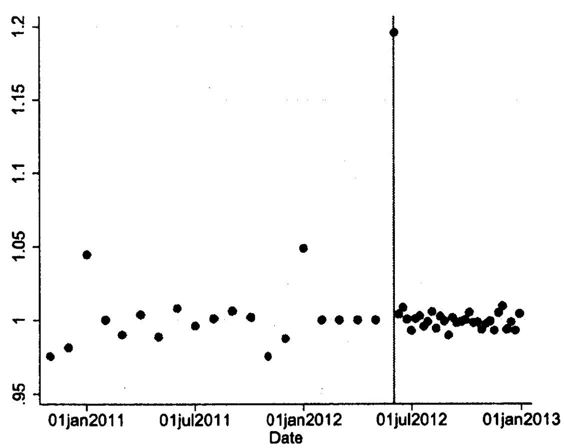


Figure 3.2: Törnqvist Price Index, State-balanced Panel

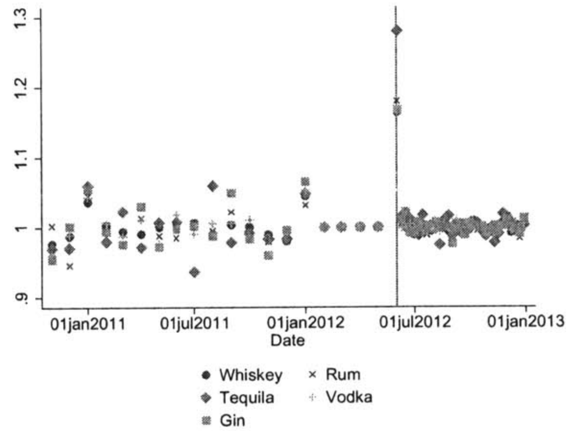
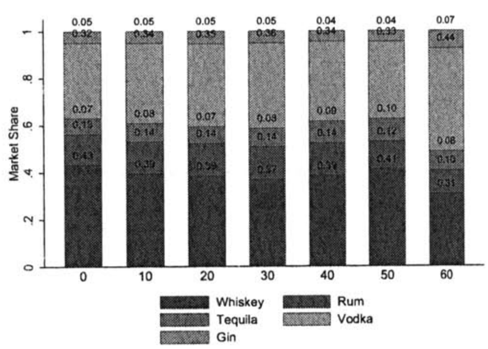
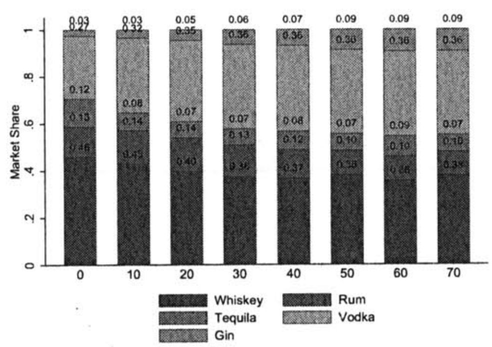


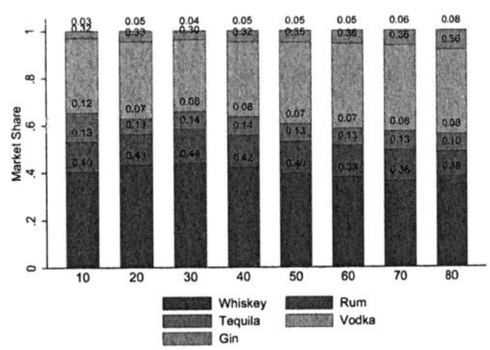
Figure 3.5: Törnqvist Price Index for Liquor Categories, State-balanced Panel



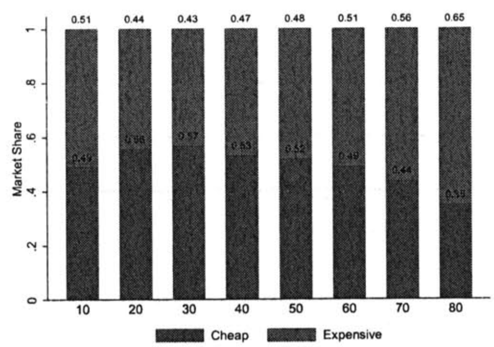
(a) % Minority and Shares by Product Category



(b) % College Educated and Shares by Product Category



(c) % High-Income and Shares by Product Category



(d) % High-Income and Shares by Product Type

Figure 3.6: Shares by Product Category/Type and Zip Code Demographics

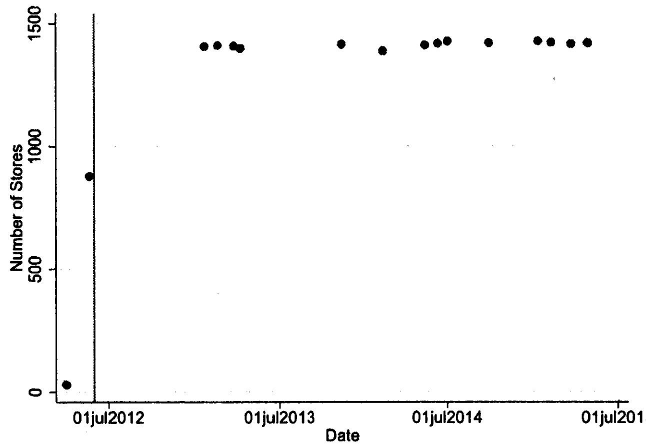


Figure 3.7: Number of Licensees over Time

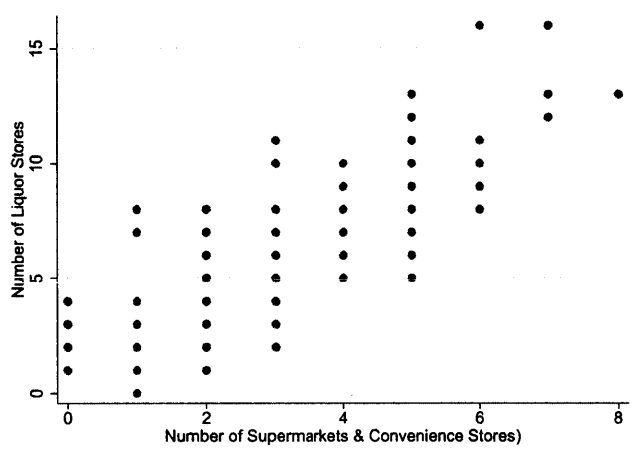


Figure 3.8: Liquor Licensees vs. Supermarkets by Zip Code

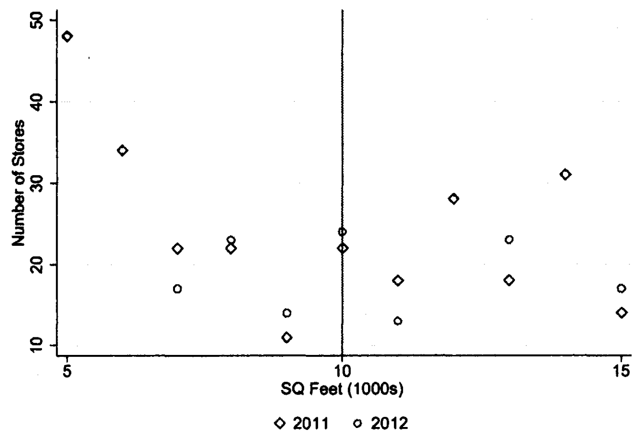


Figure 3.9: Number of Stores by Size, Before and After Privatization

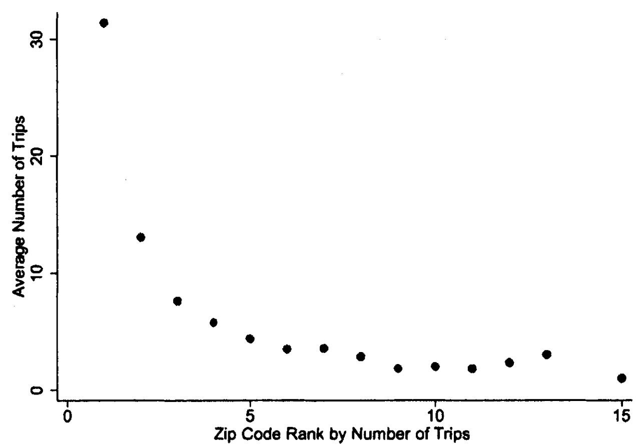
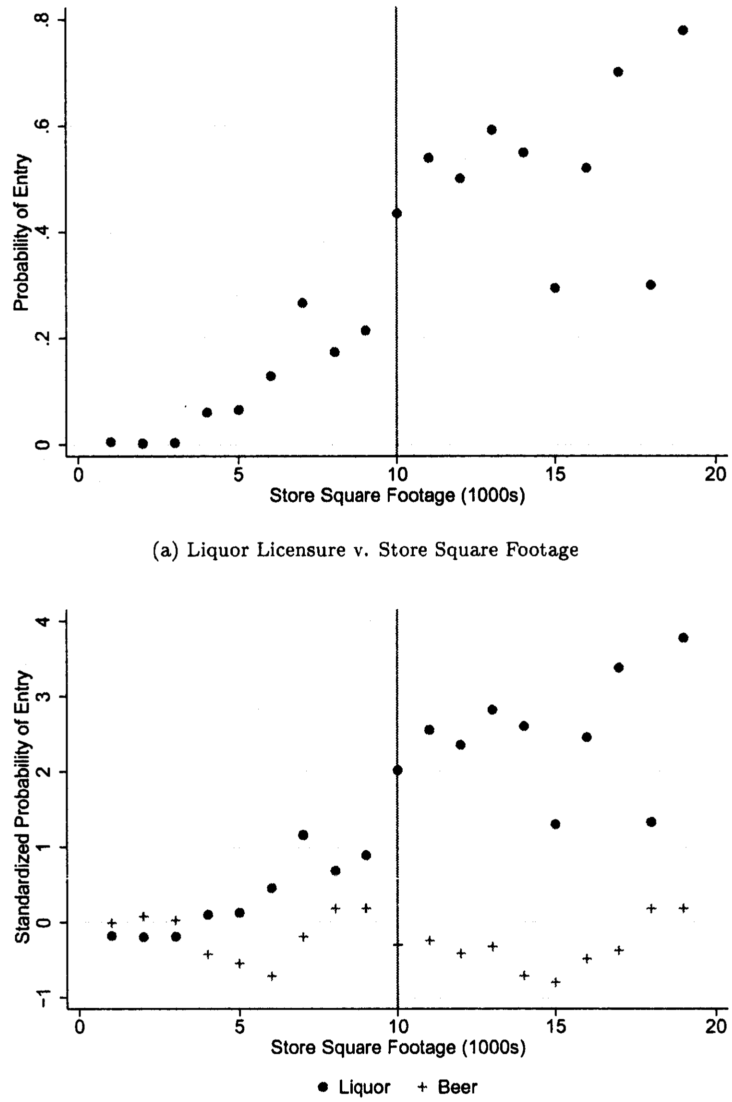


Figure 3.10: Zip Code Fuzzy Match Algorithm



(a) Liquor Licensure v. Store Square Footage

(b) Standardized Liquor and Beer Licensure

Figure 3.11: Liquor Licensure at the 10,000 Square Foot Threshold

Tables

	No Pass-Through		Perfect Pass-Through		MST (2015)	
	Mean	SD	Mean	SD	Mean	SD
Full Sample	44.0%	8.5%	38.4%	9.3%	42.7%	30.2%
Product Category						
Whiskey	43.9%	8.5%	38.2%	9.3%	39.9%	31.3%
Rum	42.7%	9.7%	37.0%	10.6%	59.6%	23.5%
Tequila	46.9%	7.3%	41.6%	8.0%	21.8%	15.1%
Vodka	43.9%	7.6%	38.3%	8.3%	40.7%	27.2%
Gin	42.7%	7.4%	37.0%	8.2%	50.6%	41.2%
Product Type						
Expensive	44.7%	8.7%	39.1%	9.6%	26.3%	13.3%
Cheap	42.7%	7.7%	37.0%	8.4%	67.4%	31.7%

Notes: This table reports summary statistics for estimated post-liberalization percentage markups between June 2012 and December 2012. All statistics are unweighted and calculated using data at the week-store-upc level. Products are categorized as "Expensive" ("Cheap") if their WSLCB price exceeds the median price charged by the WSLCB for that category. All results infer marginal cost from pre-liberalization data. Columns labelled "No Pass-Through" assume that the 10% distributor tax is not passed through to retailers, while columns labelled "Perfect Pass-Through" assume that the entirety of the tax is passed through to retailers. Columns under the header "MST (2015)" report the predicted markups from Table 11 in Miravete, Seim and Thurk (2015).

Table 3.1: Post-Liberalization Markups by Product Category and Type

Sample	8,000-12,000 Square Feet			5,000-14,000 Square Feet			Full		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
SQFT ≥ 10	0.283***	0.286***	0.268**	0.382***	0.384***	0.370***	0.425***	0.424***	0.427***
	(0.107)	(0.106)	(0.109)	(0.056)	(0.056)	(0.057)	(0.049)	(0.049)	(0.049)
(SQFT < 10) × feet from cutoff (1000s)							-0.023***	-0.023***	-0.023***
							(0.005)	(0.005)	(0.005)
(SQFT ≥ 10)× feet from cutoff (1000s)							0.009***	0.009***	0.009***
							(0.001)	(0.001)	(0.001)
Constant	0.189***	0.106	0.031	0.140***	0.176***	0.200***	0.190***	0.202***	0.204***
	(0.065)	(0.093)	(0.131)	(0.031)	(0.046)	(0.053)	(0.040)	(0.041)	(0.041)
Controls for Competitors in Zip Code									
# Groceries & Convenience Stores		✓	✓		✓	✓		✓	✓
Flexible Controls for Competitor Sizes			✓			✓			✓
N	73	73	73	246	246	246	3969	3969	3969

Notes: This tables shows the increase in the likelihood a grocery/convenience store obtains a liquor license at the 10,000 Square foot threshhold. Heteroskedasticity-robust standard errors in parentheses. Coefficients are signficant at the *10, **5%, and ***1% level. SQFT is the square footage of the store, measured in thousands. Sample is from December, 2012 TDLinx and January, 2013 WSLCB licensure data. Flexible controls for store size include: number of stores in bandwidth (8-12, 5-14, all, respectively), number of stores above bandwidth (12+,15+,10+), and number of stores below bandwidth (0-7,0-4,0-9), within Zip code.

Table 3.2: Effect of Store Size on Liquor Licensure

	# Licensed Groceries					# Licensed Groceries	Total # Liquor Licensees
	12 > SQFT ≥ 8	12 > SQFT ≥ 8	SQFT ≥ 12	8>SQFT	Total		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
# Groceries by Size:							
12 > SQFT ≥ 10	0.299***	0.299***	-0.142	0.013	0.170		
	(0.114)	(0.114)	(0.135)	(0.055)	(0.181)		
12 > SQFT ≥ 8	0.208***	0.204***	-0.019	-0.014	0.170		
	(0.070)	(0.068)	(0.099)	(0.042)	(0.142)		
8 > SQFT		0.000	-0.000	0.009**	0.009		
		(0.003)	(0.008)	(0.004)	(0.010)		
SQFT ≥ 12		0.005	0.827***	-0.024**	0.808***		
		(0.008)	(0.027)	(0.010)	(0.030)		
SQFT ≥ 10						0.468***	0.387
						(0.128)	(0.338)
Total						0.121	0.570*
						(0.109)	(0.303)
Other Variables							
Total SQFT Above Cutoff						0.008***	0.011***
						(0.001)	(0.004)
Total SQFT Below Cutoff						-0.014	-0.059
						(0.014)	(0.038)
N	512	512	512	512	512	512	512

Notes: Each observation is a five-digit zip code in Washington state. Data on the number and size of groceries (including convenience stores and Superettes) is from TDLinx. Data on liquor licensure is from the WSLCB. Total SQFT Above Cutoff is the sum, across all groceries above 10,000 square feet in the zip code, of the square footage. Total SQFT Below Cutoff is defined analogously. Heteroskedasticity-robust standard errors reported in parentheses. Coefficients are statistically significant at the *10%, **5%, and ***1% level.

Table 3.3: Number of Stores that Sell Liquor vs. Stock of Potential Entrants

sample	Full		Top 10% Carried		Bottom 90% Carried	
	(1)	(2)	(3)	(4)	(5)	(6)
# Grocers	-0.029*	-0.006	-0.002	-0.006	-0.039*	-0.006
12 > SQFT ≥10	(0.015)	(0.005)	(0.005)	(0.004)	(0.023)	(0.005)
# Grocers	0.032***	0.002	0.008*	0.004	0.042**	0.001
12 > SQFT ≥8	(0.011)	(0.003)	(0.004)	(0.004)	(0.017)	(0.004)
UPC Fixed Effects		✓		✓		✓
N	926013	926013	328322	328322	597691	597691

Notes: Observations are at the zip code - week - UPC level. Standard errors clustered at the zip code level. Coefficients are statistically significant at the *10%, **5%, *1% level. Controls include month of the year and product size (liters).

Table 3.4: Effect of Potential Entry on Price, by Liquor Type

	Whiskey		Rum		Tequila		Vodka		Gin	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
# Grocers 12 > SQFT ≥10	-0.022 (0.023)	-0.008 (0.006)	-0.025** (0.012)	-0.004 (0.005)	-0.019 (0.015)	-0.007 (0.006)	-0.041*** (0.012)	-0.005 (0.004)	0.001 (0.018)	-0.003 (0.004)
# Grocers 12 > SQFT ≥8	0.030 (0.019)	0.002 (0.004)	0.019** (0.008)	0.002 (0.003)	0.018** (0.009)	0.004 (0.005)	0.040*** (0.009)	0.003 (0.003)	0.016 (0.012)	0.003 (0.004)
UPC Fixed Effects		✓		✓		✓		✓		✓
N	264110	264110	162097	162097	101892	101892	343223	343223	54691	54691

Notes: Observations are at the zip code - week - UPC level. Standard errors clustered at the zip code level. Coefficients are statistically significant at the *10%, **5%, *1% level. Controls include month of the year and product size (liters).

Table 3.5: Effect of Potential Entry on Price, by Liquor Category

	Log Quantity Per Store		Standard Deviation of Price		Number of Products
	(1)	(2)	(3)	(4)	(5)
# Grocers 12 > SQFT ≥10	-0.004	0.003	0.823	0.597	58.090*
	(0.048)	(0.053)	(2.239)	(1.162)	(33.411)
# Grocers 12 > SQFT ≥8	0.002	0.001	1.512	0.168	-52.576*
	(0.039)	(0.041)	(2.530)	(1.237)	(28.644)
UPC Fixed Effects		✓		✓	
N	1185504	1185504	277404	277404	184

Notes: Observations are at the zip code - week - UPC level in columns (1) - (4). Observations are at the zip code level in column (5). Standard errors clustered at the zip code level. Coefficients are statistically significant at the *10%, **5%, *1% level. Controls in columns (1)-(5) include month of the year and number of supermarkets. Controls in columns (1)-(5) also include product size (liters). Quantity per store in the average quanity sold in Nielsen Scanner stores matched to the five-digit zip code. Standard deviation of price is calculated at the zip code - week - UPC level. Number of products is the number of unique UPCs sold in the zip code in the first size months of privatization.

Table 3.6: Effect of Store Sizes on Market Outcomes

	OLS	2SLS					
	Full Sample		Whiskey	Rum	Tequila	Vodka	Gin
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
# Licensed Grocers	0.026**	-0.149	-0.109	-0.151	-0.086	-0.212	-0.008
12 > SQFT ≥8	(0.011)	(0.179)	(0.160)	(0.201)	(0.125)	(0.221)	(0.084)
# Grocers		0.086	0.071	0.074	0.046	0.115	0.028
12 > SQFT ≥8		(0.077)	(0.072)	(0.085)	(0.052)	(0.097)	(0.038)
N	926013	926013	264110	162097	101892	343223	54691

Notes: Observations are at the zip code - week - UPC level. Standard errors clustered at the zip code level. Coefficients are statistically significant at the *10%, **5%, *1% level. Controls include month of the year, number of grocery stores, month, and product size (liters).

Table 3.7: 2SLS Estimates of Entry on Log Prices

Bibliography

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. Ackerberg. 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`.

Philippe Aghion and Patrick Bolton. Contracts as a Barrier to Entry. *American Economic Review*, 77(3):388–401, 1987.

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`.

Joshua D. Angrist and Alan B Krueger. The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples. *Journal of the American Statistical Association*, 87(418):328–336, 1990.

Joe S. Bain. *Barriers to new competition: their character and consequences in manufacturing industries.* Harvard University Press, 1956.

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`.

Solange Bernstein and Carolina Cabrita. Determinantes de la elecci6n de afp en chile: Nueva evidencia a partir de datos individuales. 2006.

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`.

Steven T Berry. Estimation of a Model of Entry in the Airline Industry. *Econometrica*, 60(4):889–917, 1992.

Steven T Berry. Estimating discrete choice models of product differentiation. *RAND Journal of Economics*, 25(2):242–262, Summer 1994.

Steven T. Berry and Joel Waldfogel. Free Entry and Social Inefficinecy in Radio Broadcasting. *RAND Journal of Economics*, 30(3):397–420, 1999.

Timothy F. Bresnahan and Peter C. Reiss. Entry and Competition in Concentrated Markets. *Journal of Political Economy*, 99(5):977–1009, 1991.

Luis Cabral. Switching costs and equilibrium prices. 2012.

D.W. Caves, L.R. Christensen, and W.E. Diewert. Multilateral comparisons of output, input, and productivity using superlative index numbers. *The Economic Journal*, 92(365):73–86, 1982. ISSN 00130133. doi: 10.2307/2232257.

Andrew Chamberlain. Urban Crime and Spatial Proximity to Liquor: Evidence from a Quasi-Experiment in Seattle. 2014.

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.

Judith A Chevalier and Anil Kashyap. Best Prices: Price Discrimination and Consumer Substitution. 2014.

Federico Ciliberto and Elie Tamer. Market Structure and Multiple Equilibria in Airline Markets. *Econometrica*, 77(6):1791–1828, 2009. ISSN 0012-9682. doi: 10.3982/ECTA5368.

Christopher T. Conlon. The empirical likelihood mpec approach to demand estimation. December 2013. URL `http://ssrn.com/abstract=2331548`.

Christopher T Conlon and Nirupama S Rao. Wholesale Prices, Retail Prices and the Lumpy Pass-Through of Alcohol Taxes. 2015a.

Christopher T Conlon and Nirupama S Rao. The Price of Liquor is Too Damn High: Alcohol Taxation and Market Structure. 2015b.

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.

Avinash Dixit. The Role of Investment in Entry-Deterrence. *The Economic Journal*, 90(537):95–106, 1980.

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.

Glenn Ellison and Sara Fisher Ellison. Strategic entry deterrence and the behavior of pharmaceutical incumbents prior to patent expiration. *American Economic Journal: Microeconomics*, 3(1):1–36, 2011. ISSN 19457669. doi: 10.1257/mic.3.1.1.

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`.

Christopher Gedge, James W Roberts, and Andrew Sweeting. A Model of Dynamic Limit Pricing with an Application to the Airline Industry. 2014.

Austan Goolsbee and Chad Syverson. How do incumbents respond to the threat of entry? Evidence from the major airlines. *Quarterly Journal of Economics*, 123(November):1611–1633, 2008. ISSN 0033-5533. doi: 10.1162/qjec.2008.123.4.1611.

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 Tejeda-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 Hortaçsu, 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`.

Jerry Hausman. Specification tests in econometrics. *Econometrica*, 46(6):1251–1271, Nov 1978.

Jerry Hausman and Daniel McFadden. Specification tests for the multinomial logit model. *Econometrica, Vol. 52, No. 5 (Sep., 1984), pp. 1219-1240.*, 52(5):1219–1240, Sep 1984.

Jerry Hausman and David. A Wise. A conditional probit model for qualitative choice. *Econometrica*, 46(2):403–426., Mar 1978.

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. *AAmerican Economic Review (forthcoming)*, 2014.

Guido W. Imbens and Thomas Lemieux. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635, 2008. ISSN 03044076. doi: 10.1016/j.jeconom.2007.05.001.

Atsushi Inoue and Gary Solon. Two-Sample Instrumental variables estimators. *The Review of Economics and Statistics*, 92(3):557–561, 2010. ISSN 03044076. doi: 10.1016/0304-4076(90)90061-W.

Panle Jia. What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry. *Econometrica*, 76(6):1263–1316, 2008. ISSN 0012-9682. doi: 10.3982/ECTA6649.

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`.

Rafael La Porta and Florencio Lόpez-De-Silanes. The Benefits of Privatization: Evidence from Mexico. *The Quarterly Journal of Economics*, 114(4): 1193–1242, 1999. ISSN 0033-5533. doi: 10.1162/003355399556250. URL `http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=2630203&site=ehost-live&scope=site`.

Fernando Luco. *Switching Costs and Competition in Retirement Investment*. PhD thesis, Northwestern University, 2014.

N. Gregory Mankiw and Michael D. Whinston. Free Entry and Social Inefficiency. *The RAND Journal of Economics*, 17(1):48–58, 1986.

Eric S. Maskin. Uncertainty and Entry Deterrence. *Economic Theory*, 14:429–437, 1999.

D. McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York, 1974.

Paul Milgrom and John Roberts. Limit Pricing and Entry under Incomplete Information: An Equilibrium Analysis. *Econometrica*, 50(2):443–459, 1982.

Eugenio J Miravete, Katja Seim, and Jeff Thurk. Complexity, Efficiency, and Fairness of Multi-Product Monopoly Pricing. 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.

Craig Peters. Evaluating the Performance of Merger Simulation: Evidence from the US Airline Industry. *Journal of Law and Economics*, 49(2):627–649, 2006.

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`.

SC Salop. Monopolistic Competition with Outside Goods. *The Bell Journal of Economics*, 10(1):141–156, 1979. ISSN 0361915X. doi: 10.2307/3003323. URL `http://www.jstor.org/stable/3003323`.

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`.

Robert C. Seamans. Threat of Entry, Asymmetric Information, and Pricing. *Strategic Management Journal*, 34:426–444, 2013. ISSN 01432095. doi: 10.1002/smj.

Katja Seim and Joel Waldfogel. Public Monopoly and Economic Efficiency: Evidence from the Pennsylvania Liquor Control Board's Entry Decisions. *American Economic Review*, 103(2):831–862, April 2013. ISSN 0002-8282. doi: 10.1257/aer.103.2.831. URL `http://pubs.aeaweb.org/doi/abs/10.1257/aer.103.2.831`.

Chad Syverson. Market Structure and Productivity: A Concrete Example. *Journal of Political Economy*, 112(6):1181–1222, 2004.

Joseph V. Terza, Anirban Basu, and Paul J. Rathouz. Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics*, 27(3):531–543, May 2008.

Jean Tirole. *The theory of industrial organization*. MIT Press, 1988.

Leo Törnqvist. The Bank of Finland's Consumption Price Index. *Bank of Finland Monthly Bulletin*, 10:1–8, 1936.

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.

E Glen Weyl and Michal Fabinger. Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition. *Journal of Political Economy*, 121(3):528–583, 2013. ISSN 00223808. doi: 10.1086/670401. URL `http://www.jstor.org/stable/10.1086/670401`.

Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, MA, 2002.

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.

Appendix A

Appendices for Chapter 1

A.1 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 1.12 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 period[1]. However, one can certainly find more expensive alternatives in the US system.

A.2 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 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 1.7shows monthly returns, by fund, between November 2002

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

and July 2012, while Table 1.13 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 1.15 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:

$$r_{ift} = \alpha_0 + \alpha_1 p_{it} + \alpha_2 r_{if,t-1} + \delta_t + \gamma_f + \epsilon_{ift} \tag{A.2}$$

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 E[2]. 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 SCOMP[3]. 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 1.8 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

[2]Obviously, these regressions do not have fund effects.

[3]Sistema de Consultas y Ofertas de Montos de Pension.

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.

A.3 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 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 1.14 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 auction[4] 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 existence[5]. 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 aggresively 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.

[4] Recall 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.

[5] To 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.

A.4 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:

$$\begin{aligned} &u\left(j, j', X_{ijt}, \epsilon_{ijt}\right) - u\left(j', j', X_{ij't}, \epsilon_{ij't}\right) \\ &\geq \beta \cdot \left(E\left[V_{i,t+1}\left(j', B_{it}\right)|\Omega_{it}\right] - E\left[V_{i,t+1}\left(j, B_{it}\right)|\Omega_{it}\right]\right) \end{aligned} \tag{A.3}$$

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

$$\begin{aligned} &E\left[V_{i,t+1}\left(j', B_{it}\right)|\Omega_{it}\right] - E\left[V_{i,t+1}\left(j, B_{it}\right)|\Omega_{it}\right] \geq \\ &E\left[u\left(j^*_{t+1}, j', X_{ij,t+1}, \epsilon_{ij,t+1}\right) - u\left(j^*_{t+1}, j, X_{ij,t+1}, \epsilon_{ij,t+1}\right)|\Omega_{it}\right] \\ &+\beta^{T_i-t-1} \cdot E\left[B_{iT}\left(\{j^*_\tau\}_{\tau=t+1}^{T_i}, B_{i,t+1}\left(j'\right)\right) - B_{iT}\left(\{j^*_\tau\}_{\tau=t+1}^{T_i}, B_{i,t+1}\left(j\right)\right)|\Omega_{it}\right] \\ &\geq -\gamma + \beta^{T_i-t-1} \cdot B_{it} \cdot E\left[\left(r_{j't} - r_{jt}\right)\textstyle\prod_{\tau=t+1}^{T}\left(1 + r_{j^*\tau}\right)|\Omega_{it}\right] \end{aligned} \tag{A.4}$$

Which means we can re-write A.3 as:

$$\begin{aligned} &\alpha w_{it}\left(p_{jt} - p_{j't}\right) + \delta\left(-1 + \beta\right) \\ &+\beta^{T_i-t} \cdot B_{it} \cdot E\left[\left(r_{j't} - r_{jt}\right)\textstyle\prod_{\tau=t+1}^{T}\left(1 + r_{j^*\tau}\right)|\Omega_{it}\right] \\ &\geq \epsilon_{ij't} - \epsilon_{ijt} \end{aligned} \tag{A.5}$$

In the second case, we get that:

$$\begin{aligned} &u\left(j, j', X_{ijt}, \epsilon_{ijt}\right) - u\left(j'', j', X_{ij't}, \epsilon_{ij't}\right) \\ &\geq \beta \cdot \left(E\left[V_{i,t+1}\left(j'', B_{it}\right)|\Omega_{it}\right] - E\left[V_{i,t+1}\left(j, B_{it}\right)|\Omega_{it}\right]\right) \end{aligned} \tag{A.6}$$

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

$$
\begin{aligned}
&E\left[V_{i,t+1}\left(j'', B_{it}\right)|\Omega_{it}\right] - E\left[V_{i,t+1}\left(j, B_{it}\right)|\Omega_{it}\right] \geq \\
&E\left[u\left(j^*_{t+1}, j'', X_{ij,t+1}, \epsilon_{ij,t+1}\right) - u\left(j^*_{t+1}, j, X_{ij,t+1}, \epsilon_{ij,t+1}\right)|\Omega_{it}\right] \\
&+\beta^{T_i-t-1} \cdot E\left[B_{iT}\left(\{j^*_\tau\}_{\tau=t+1}^{T_i}, B_{i,t+1}\left(j''\right)\right) - B_{iT}\left(\{j^*_\tau\}_{\tau=t+1}^{T_i}, B_{i,t+1}\left(j\right)\right)|\Omega_{it}\right] \\
&\geq -\delta + \beta^{T_i-t-1} \cdot B_{it} \cdot E\left[\left(r_{j''t} - r_{jt}\right)\prod_{\tau=t+1}^{T}\left(1 + r_{j^*\tau}\right)|\Omega_{it}\right]
\end{aligned}
\tag{A.7}
$$

Which means we can re-write A.3 as:

$$
\begin{aligned}
&\alpha w_{it}\left(p_{jt} - p_{j't}\right) + \delta\beta \\
&+\beta^{T_i-t} \cdot B_{it} \cdot E\left[\left(r_{j't} - r_{jt}\right)\prod_{\tau=t+1}^{T}\left(1 + r_{j^*\tau}\right)|\Omega_{it}\right] \\
&\geq \epsilon_{ij't} - \epsilon_{ijt}
\end{aligned}
\tag{A.8}
$$

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:

$$
\begin{aligned}
E\left[V_{it}\left(\emptyset, B_{it}\right)|\Omega_{it}\right] \equiv \quad & \max_{\{j_\tau \in \mathcal{J}_\tau\}_{\tau=t}^{T_i}} \left\{u\left(j_t, \emptyset, X_{ijt}, \epsilon_{ijt}\right) + \sum_{\tau=t+1}^{T_i} \beta^{\tau-t} \cdot E\left[u\left(j_\tau, j_{\tau-1}, X_{ij\tau}, \epsilon_{ij\tau}\right)|\Omega_{it}\right]\right. \\
& \left. +\beta^{T_i-t} \cdot E\left[B_{iT_i}\left(\{j_\tau\}_{\tau=t}^{T}, B_{it}\right)|\Omega_{it}\right]\right\} \\
& = \max_{j_t} \; u\left(j_t, \emptyset, X_{ijt}, \epsilon_{ijt}\right) + \beta \cdot E\left[V_{i,t+1}\left(j_t, B_{it}\right)|\Omega_{it}\right]
\end{aligned}
\tag{A.9}
$$

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:

$$
\begin{aligned}
&u\left(j, \emptyset, X_{ijt}, \epsilon_{ijt}\right) - u\left(j', \emptyset, X_{ij't}, \epsilon_{ij't}\right) \\
&\geq \beta \cdot \left(E\left[V_{i,t+1}\left(j', B_{it}\right)|\Omega_{it}\right] - E\left[V_{i,t+1}\left(j, B_{it}\right)|\Omega_{it}\right]\right)
\end{aligned}
\tag{A.10}
$$

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

$$\begin{gathered} \alpha w_{it}\left(p_{jt}-p_{j't}\right)+\delta\beta \\ +\beta^{T_i-t}\cdot B_{it}\cdot E\left[\left(r_{j't}-r_{jt}\right)\prod_{\tau=t+1}^{T}\left(1+r_{j^*\tau}\right)|\Omega_{it}\right] \\ \geq \epsilon_{ij't}-\epsilon_{ijt} \end{gathered} \tag{A.11}$$

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

$$E\left[V_{i,T_i-1}\left(j,B_{i,T_i-1}\right)|\Omega_{i,T_i-1}\right]\equiv\max_{j_{T_i-1}}u\left(j_{T_i-1},j,X_{ijt},\epsilon_{ijt}\right)+\beta\cdot E\left[B_{iT_i}\left(j_{T_i-1},B_{i,T_i-1}\right)|\Omega_{i,T_i-1}\right] \tag{A.12}$$

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

$$\begin{gathered} u\left(j,j,X_{ij,T_i-1},\epsilon_{ij,T_i-1}\right)-u\left(j',j,X_{ij',T_i-1},\epsilon_{ij',T_i-1}\right) \\ \geq\beta\left(E\left[B_{iT_i}\left(j',B_{i,T_i-1}\right)|\Omega_{i,T_i-1}\right]-E\left[B_{iT_i}\left(j,B_{i,T_i-1}\right)|\Omega_{i,T_i-1}\right]\right) \end{gathered} \tag{A.13}$$

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

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

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

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

as well as:

$$
\begin{aligned}
& u\left(j, j', X_{ij,T_i-1}, \epsilon_{ij,T_i-1}\right) - u\left(j'', j', X_{ij',T_i-1}, \epsilon_{ij',T_i-1}\right) \\
& \geq \beta\left(E\left[B_{iT_i}\left(j', B_{i,T_i-1}\right)|\Omega_{i,T_i-1}\right] - E\left[B_{iT_i}\left(j, B_{i,T_i-1}\right)|\Omega_{i,T_i-1}\right]\right)
\end{aligned}
\tag{A.17}
$$

$$
\begin{gathered}
\alpha w_{i,T_i-1}\left(p_{j,T_i-1} - p_{j',T_i-1}\right) \\
+\beta \cdot B_{i,T_i-1} \cdot E\left[r_j - r_{j'}|\Omega_{i,T_i-1}\right] \\
\geq \epsilon_{ij't} - \epsilon_{ijt}
\end{gathered}
\tag{A.18}
$$

This defines all the cases in equation 1.14.

A.5 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 $X_{ijj't} \equiv w_{it}\left(p_{jt} - p_{j't}\right)$, $\epsilon_{ijj't} \equiv \epsilon_{ijt} - \epsilon_{ij't}$, and ignore returns differences across firms. Then:

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

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

$\Pr\left[X < x, Y = 1|j, z\right] = \Pr\left[\epsilon < X - \delta\left(1+\beta\right) < x - \delta\left(1+\beta\right)|j, z\right]$

$\leq \Pr\left[\epsilon < x - \delta\left(1+\beta\right)|j, z\right] = F\left(x - \delta\left(1+\beta\right)|j, z\right)$

$\Pr\left[X > x, Y = 0|j, z\right] = \Pr\left[\epsilon > X - \delta\left(1-\beta\right)|j, z\right]$

$\leq \Pr\left[\epsilon > x - \delta\left(1-\beta\right)|j, z\right] = 1 - F\left(x - \delta\left(1-\beta\right)|j, z\right)$

$\Pr\left[X > x, Y = 1|j', z\right] = \Pr\left[x + \delta\left(1+\beta\right) < X + \delta\left(1+\beta\right) < \epsilon|j', z\right]$

$\leq \Pr\left[x + \delta\left(1+\beta\right) < \epsilon|j', z\right] = 1 - F\left(x + \delta\left(1+\beta\right)|j', z\right)$

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

$\leq \Pr\left[x + \delta\left(1-\beta\right) > \epsilon|j', z\right] = F\left(x + \delta\left(1-\beta\right)|j', z\right)$

Then:

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

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

Combining the two inequalities:

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

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

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

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

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

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

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

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

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

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

Proof. If: □

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

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

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 $\forall x \in \left(\max\left\{ \bar{X}^{1,j} - \delta(1+\beta), \bar{X}^{0,j} - \delta(1-\beta) \right\}, \min\left\{ \underline{X}^{0,j'} + \delta(1-\beta), \underline{X}^{1,j'} + \delta(1+\beta) \right\} \right)$ (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 $\delta > \bar{\delta}$, $\forall x \in \left(\max\left\{ \bar{X}^{1,j} - \delta(1+\beta), \bar{X}^{0,j} - \delta(1-\beta) \right\}, \min\left\{ \underline{X}^{0,j'} + \delta(1-\beta), \underline{X}^{1,j'} + \delta(1+\beta) \right\} \right)$ 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 $\min\left\{\underline{\mathrm{X}}^{0,j'}+\delta(1-\beta), \underline{\mathrm{X}}^{1,j'}+\delta(1+\beta)\right\} < x''(\delta) < \max\left\{\underline{\mathrm{X}}^{0,j'}+\delta(1-\beta), \underline{\mathrm{X}}^{1,j'}+\delta(1+\beta)\right\}$, we can reject δ at x''. Since $\underline{\mathrm{X}}^{0,j'}+\delta_0(1-\beta)<x'<\underline{\mathrm{X}}^{1,j'}+\delta_0(1+\beta)$, we also have that:

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

Therefore, if for some $\delta^* > \max\left[\bar{\delta}, \frac{\underline{\mathrm{X}}^{0,j'}-\underline{\mathrm{X}}^{1,j'}}{2\beta}\right]$ there exists $x'\in\left(\underline{\mathrm{X}}^{0,j'}+\delta^*(1-\beta), \underline{\mathrm{X}}^{1,j'}+\delta^*(1+\beta)\right)$ 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:

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

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

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

Then we have that:

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

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

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:

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

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 $\Pr[X - \delta^*(1-\beta) > x^*, Y = 0|j,z_{j'}]$, then we can reject δ $\forall\delta > \delta^*$.

Figure 1.9 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.

Appendix B

Appendices for Chapter 3

B.1 Appendix Figures

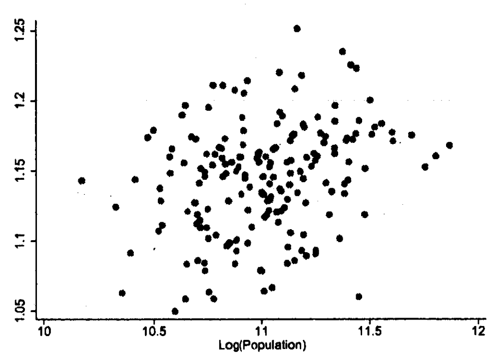


Figure B.1: Törnqvist Price Index Change at Liberalization and Zip 5 Population, State-balanced Panel

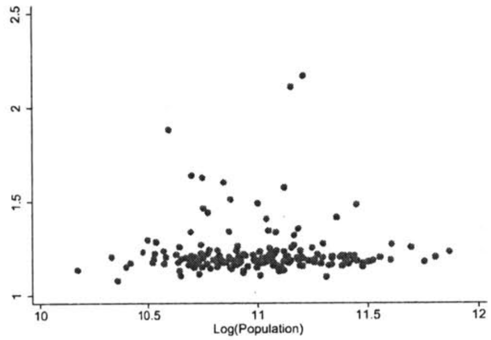


Figure B.2: Törnqvist Price Index Change from Liberalization to End of 2012 and Zip 5 Population, State-balanced Panel

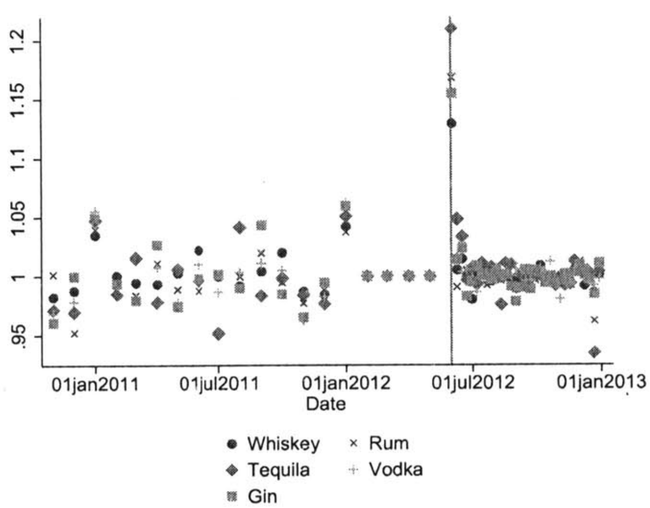


Figure B.3: Törnqvist Price Index for Liquor Categories, Unbalanced Panel