Quantifying Search and Switching Costs in the U.S. Auto Insurance Industry

Elisabeth Honka*

October 3, 2010

PRELIMINARY AND INCOMPLETE

FOR AN UPDATED VERSION, PLEASE CHECK: http://home.uchicago.edu/~ehonka/

Abstract

Consumers making purchases in service categories such as insurance potentially face search and switching costs. On the one hand, incomplete information about the various alternatives necessitates search behavior which could lead to consumers switching from the current provider to a different provider. On the other hand, brand loyalty, the prospect of dealing with a new provider, and other psychological factors result in switching costs that introduce frictions in the market. Both search and switching have been studied in the empirical literature albeit separately. In this paper, our objective is to distinguish empirically between search and switching costs in the context of consumers' choices in the U.S. auto insurance industry. Our data contain information on respondents' search behavior in terms of their consideration set of insurance providers; the corresponding premia for those providers; the actual provider chosen; the identity of the previous insurer and the demographic and psychographic characteristics of the consumers and their policies. The unique aspects of this information enable us to identify separately both search and switching costs. To quantify the magnitudes of these costs, we develop an integrated utility-maximizing model of consumers' decisions of how many companies to search; the identities of these companies and from which company to purchase. Our modeling approach rationalizes the observed consideration set as being the outcome of a search process by a consumer. Using this approach we learn about the levels of search and switching costs, and their relative magnitudes. Taking our model to the data, we find that

*Booth School of Business, University of Chicago. Email: ehonka@ChicagoBooth.edu. We wish to thank Pradeep Chintagunta, Jean-Pierre Dubé, Ali Hortacsu, Chad Syverson, Pranav Jindal, and Navdeep Sahni. We thank the participants of the 2010 IIOC conference and the seminar participants at the Illinois Institute for Technology. We are grateful to an anonymous company and Highline Data Inc. for providing us with the data. The usual disclaimer applies.

there are substantial search and switching costs in the U.S. auto insurance industry. The cost of collecting a price quote through an online channel are about \$45, the cost of collecting a price quote offline vary from \$70 to \$110. The average switching costs a consumer incurs are about \$85. We find consumer search costs and customer satisfaction to be the two main drivers of the high retention rate observed in the U.S. auto insurance industry. To support managers in their decision making about investments in measures to increase customer satisfaction, we calculate additional profits resulting from a maximization of customer satisfaction. And finally, we study whether insurance companies should increase or decrease insurance premia to new and existing customers. We find insurance companies lose money by offering discounts to new customers and increase their profits by increasing premia to existing customers. Our results are consistent with the "invest-harvest" strategy found in the switching cost literature.

1 Introduction

The U.S. auto insurance industry wrote insurance premia for about $160 billion and spent nearly $3 billion on advertising in 2007. This is an industry that is characterized by a very high retention rate. According to industry sources around 90% of customers stay with their insurance provider after a one year contract period and the percentage increases with tenure (this pattern is also described by Israel (2005a)). Further, industry sources state that consumers only collect 2.4 quotes on average before making a purchase decision. At the same time, insurance companies advertise potential annual savings of around $400 when switching insurance providers. For example, The Hartford claims on its website (November 13th, 2009) annual savings of approximately $402 upon switching. Another insurer, Esurance, introduced a "Switch & Save Discount" where new customers can save an additional 5% on their first year premia[1]. One explanation for such a high retention rate and for the level of advertised savings is the presence of consumer search and switching costs. Here, we define search costs as the cost to the consumer of conducting one search and switching costs as the cost a consumer incurs when switching insurance providers after the search has been completed[2].

High retention rates as observed in the U.S. auto insurance industry can be problematic if they are not driven by customer satisfaction, but by market rigidities such as search and switching costs. High retention rates increase companies' market power and can lead to higher prices for consumers[3]. To determine the importance of these rigidities, it is necessary to quantify search and switching costs. These two types of costs can play important roles in managers' decision-making regarding factors like pricing or advertising. *Separately* quantifying both types of costs is crucial to managers in order to determine whether to increase or decrease insurance premia and whether to charge different prices to new and existing customers. For example, if search costs are the primary source of market frictions and switching costs are small, offering discounts to new customers might not attract many because new customers would not know about the discount unless the company is in their consideration set. And getting into new customers' consideration sets is difficult when search costs are large as, in that case, consumers only consider very few companies. Instead, offering discounts to existing customers might prevent many from switching to a different insurance provider. If switching costs are large compared to search costs, the opposite strategy, i.e. increasing insurance premia to existing customers, might be optimal for insurance companies.

Public policy makers may be better positioned to directly influence consumer search and switching costs. Portals on state insurance commissioners websites which give a consumer price quotes for all auto insurance companies operating in his state at once could substantially reduce consumer search costs. Making it more

[1] We are using the terms "prices" and "premia" interchangeably.
[2] We are using the terms "conduct a search", "consider" and "get/ collect quotes from" interchangeably.
[3] Researchers have found switching costs to make markets more competitive in certain contexts (Dubé et al. (2009)).

difficult for insurance companies to decline to cover the cost of an accident by requiring to justify it in writing to the state insurance commissioner's office or offering last-resort guarantees to customers similar to the banking industry could decrease consumer switching costs. Separately quantifying both types of costs allows public policy makers to decide whether they should focus on decreasing search or switching costs as it is unclear a priori whether, given the current level of search costs, reducing switching costs or, given the current level of switching costs, reducing search costs will lower retention rates to a larger extent. Furthermore, it is unclear whether reducing either type of cost will have company-neutral effects or favor one company over another. And finally, knowing the consequences of reducing these costs allows managers and public policy makers to conduct cost-benefit analyses that can guide decision making with regards to investments that reduce search and switching costs.

Most of the research on measuring search and switching costs has been confined to frequently purchased products such as margarine and laundry detergent[4]. Given panel data on repeated choices by consumers, researchers have been able to estimate switching costs by exploiting consumer inertia in choice behavior (e.g. Dubé et al. (2010) for margarine). However, when it comes to search costs, researchers have essentially depended on choice data with the identification of search costs coming from the structure imposed by assumptions on the search process (e.g. Mehta et al. (2003) for laundry detergents). Hence, identifying search and switching costs just from panel data on consumer choices is a challenging task. By contrast, in this paper, we have access to data both on whether a consumer stays with the existing insurance provider or moves to a new provider and the set of companies he considers prior to making the choice. This unique aspect of the data wherein consideration sets are observed allows us to separately identify search and switching costs.

We describe a consumer's auto insurance policy purchase decision after he has decided on the various policy characteristics he would like (i.e. coverage levels etc.). Thus, we assume that consumers want to buy the same or a very similar insurance policy from each available insurance company. While these assumptions are somewhat restrictive, we observe over 93% of consumers in our data stating that they kept their coverage choice the same during the last shopping occasion. We also assume that when searching for auto insurance the main source of a consumer's uncertainty is the price, i.e. the premium charged by each provider. Thus the only search dimension is price. This assumption is motivated by 47% of the consumers in our data reporting that learning about prices was their most important reason to shop[5]. Further, the presence of advertising campaigns that promise consumers savings on insurance premia as Geico's "15 minutes could save you 15% or more" and the previously mentioned advertisements on insurers' websites reinforce the notion of potential savings and price dispersion in the auto insurance industry.

[4]Note that there is also an extensive literature in marketing on quantifying the returns to search that does not explicitly estimate search costs (e.g. Ratchford and Srinivasan (1993), Gauri et al. (2008)).

[5]"Buying or selling a vehicle" was the second most important shopping prompter with 7.5%.

To quantify the search and switching costs consumers incur while shopping for auto insurance we develop an integrated utility-maximizing model that describes consumers' consideration and purchase decisions, i.e. how many and which companies to consider and which company to sign up with. Before engaging in search we assume that consumers know all firms' characteristics except premia. We assume consumers have expectations about prices which are based on the premia they and other consumers have paid in the past. Further, we assume that consumers know the premium their previous insurer is going to charge them to renew their insurance policy. Consumers must engage in search to learn about actual prices charged by all companies except their previous insurer. Since search is costly, consumers only search a limited number of firms which together with the quote from their previous insurer make up their consideration sets. Explicitly characterizing considerations sets requires an assumption on the type of search method (simultaneous versus sequential) that the consumer follows. Here we assume a simultaneous search approach[6]. Once consumers have decided which and how many companies to consider, they collect quotes from all companies in their consideration sets and all uncertainty regarding prices is resolved. Finally, consumers choose an insurer from their consideration set. We jointly estimate both parts (consideration and choice) of our model. As we detail later, we need to overcome several challenges in implementing our model. For example, premia are a challenge to work with when using data from the insurance industry. Unlike CPG or most other goods, consumer and policy characteristics are inherently built into prices.

Previous literature has found search costs in the auto insurance industry to lie between CAD $127 - CAD $554 (adjusted for inflation; Dahlby and West (1986)) and switching costs in the auto insurance industry to equal 20% of the premium paid (Cummins et al. (1974)) or CAD $180 - CAD $370 (adjusted for inflation; Berger et al. (1989)). Our results indicate that average switching costs are about $85; the cost of one price search over the internet is about $45, while the cost of a price search through local agents, mail, and calling centers are $83, $110, and $70, respectively. When studying how search and switching costs, customer satisfaction, and consumer preferences other than those associated with inertia contribute to the level of retention found in the auto insurance industry, we find consumer search costs and customer satisfaction to be the two main drivers. These search and switching costs together with customer satisfaction provide an explanation for the very high retention rate observed in the auto insurance industry and appear to support the high potential savings advertised by the industry. To support managers in their decision making about investments in measures to increase customer satisfaction, we calculate additional profits resulting from a maximization of customer satisfaction. Assuming a 15% (25%) profit margin, we find maximizing customer satisfaction to increase total company profits across all insurance types by 7% (10%) for Allstate, 6% (10%) for Geico, 14% (24%) for Safeco, and 3% (5%) for Travelers. This can be compared to the costs of

[6]Later in the paper, we discuss our choice of simultaneous versus sequential search (see Section 5.1).

increasing satisfaction to decide whether to invest in increasing satisfaction. And finally, we study whether insurance companies should increase or decrease insurance premia to new and existing customers. We find insurance companies to lose money by offering discounts to new customers and increase their profits by increasing premia to existing customers. Our results are consistent with the "invest-harvest" theme found in the switching cost literature.

The remainder of this paper is organized as follows: First, we describe some institutional details of the U.S. auto insurance industry, then we discuss the relevant literature. In Section 4, we present our data. We introduce our model, estimation approach, and discuss identification in the following three sections. In Section 8, we present our results and study three counterfactuals in the following section. We conduct robustness checks in Section 10 and make suggestions for future research and conclude in the last two sections.

2 The U.S. Auto Insurance Industry

The net premia written in the property & casualty insurance in the U.S. equaled $447.9 billion (after reinsurance transactions, excluding state funds) in 2007 according to the National Association of Insurance Commissioners (NAIC). About 500,000 people were employed by direct property & casualty insurers in 2008 according to the U.S. Bureau of Labor Statistics. The largest line of business in the property & casualty insurance sector accounting for more than one third of the net premia written ($159.7 billion in 2007) is private passenger auto insurance. The national average expenditure for auto insurance in 2006 was $817, down 1.7% from 2005.

Several features of the U.S. private passenger auto insurance industry are worth mentioning: First, it is very fragmented. In 2007, the Top 5 auto insurance groups had a market share of less than 50%. During the last decade the "old" market leaders – State Farm and Allstate – have lost market share, while "new", internet focused insurance groups – Progressive and Berkshire Hathaway (i.e. Geico) – have roughly doubled their market shares. The market concentration as a whole has increased as the combined market share of the Top 10 insurance groups has increased by 6.4% between 1998 and 2007. Second, it is a highly regulated industry. All states except Wisconsin require auto insurance companies to file all form and rate changes with the state's insurance commissioner, i.e. all price changes fall under the Freedom of Information Act (FOIA) and are publicly known. Third, all premia and premium changes have to be argued on the basis of cost and are subject to approval by the state insurance commissioners. Fourth, all premia or premium changes have to be based on allowed observable customer characteristics. Both which and how high customer characteristics can be priced in the pricing schedule are also subject to approval. Fifth, only the largest auto insurers (and not even all of them) offer auto insurance in all 50 states and DC. Most insurance companies only offer coverage in a number of states, e.g. American Family in 18 and Mercury in 13 states. Sixth, following from the previous point there are big differences in the competitive landscape in each state. While the Top 5 companies are present in all states and DC (except Progressive in Massachusetts) they face a different composition of competitors in each state and have different market shares. Similar to Bronnenberg et al. (2009) who found geographic preferences for CPG brands, we find the same for auto insurance companies. And seventh, not only do market leaders vary by state, but also the number of companies that consumers can choose from is different. For example, Hawaiians can pick from 16, consumers in Wyoming from 31, consumers in New York from 47, and consumers in Indiana from 61 of the Top 100 insurance companies. We believe the Top 100 companies to be a good estimate of the total number of insurance companies since they had a market share of 96.2% in 2007.

We will take the characteristics of the U.S. private passenger auto insurance industry into account in our modeling approach. The observation that it is a very fragmented industry indicates that it is not enough to model demand for the Top 5 insurance companies, instead a much larger set of companies needs to be included. We will model demand for 16 of the Top 20 insurance groups[7]. The strict regulation of insurance premia will be important in our approach to insurance prices. And finally, the observation that the set of companies offering auto insurance varies widely from state to state (by size and composition) motivates our approach to let consumers choose from state-specific sets of companies.

3 Relevant Literature

This paper is related to four streams of literature - those on search costs, on switching costs, on consideration sets, and on auto insurance. We refer to Baye et al. (2007) for an overview of the literature on consumer search and discuss previous literature that our paper is most closely related to in this section. Dahlby and West (1986) estimate search costs in the auto insurance industry in Alberta (Canada) in the seventies. They use the model developed by Carlson and McAfee (1983) which is a general equilibrium model assuming consumers have heterogenous search costs and firms have heterogeous production costs, consumers search sequentially and the number of firms over which consumers search randomly is finite. Using this model, Dahlby and West (1986) estimate search costs for different risk groups separately and find them to lie between CAD $127 and CAD $554 (adjusted for inflation). Recently, there has been a lot of interest in estimating search costs distributions using data on price only. Both Hong and Shum (2006) and Moraga-Gonzalez and Wildenbeest (2008) develop two different methods to estimate heterogeneous search costs for homogenous products using price data only in a simultaneous and sequential search model. Neither paper

[7]One insurance group (AIG) had two private passenger auto insurance companies at the time of our survey (21st Century and AIG). We model demand for these two companies separately.

observes the search process or its outcome. These papers assume that prices are the only factor that matters to consumers; that companies are identical in their characteristics, and that consumers randomly pick the companies they search. Mehta et al. (2003) and Kim et al. (2009) explicitly acknowledge and model search as the way consumers form their consideration sets. Neither Mehta et al. (2003) nor Kim et al. (2009) observe the search process, but the latter paper observes the outcome of search at the aggregate level. Both papers estimate the size and composition of consideration sets. In contrast to the previous literature, we observe consideration sets and purchase decisions at the individual level. We explicitly model consideration sets as the outcomes of a search process that results in which and how many companies the consumers' consideration sets consist of, as well as the purchase decision given consideration. Thus identification of search and switching costs is facilitated with fewer assumptions than the previous literature.

Because of the rarity of observing consumer search behavior, a stream of marketing literature estimated monetary returns to search instead (Ratchford and Srinivasan (1993), Ratchford et al. (2003), Brynjolfsson et al. (2004), Gauri et al. (2008)). Returns to search provide an upper bound on search costs. Some of the previous literature (e.g. Ratchford and Srinivasan (1993), Gauri et al. (2008)) has used a non-structural approach in estimating benefits of search. Such an approach does not permit to predict changes in consumer and market place behavior as a consequence of changes in the underlying search cost distribution. Predicting those changes is important as the internet, shopbots, and regulatory measures can change consumer search costs.

There is another stream of marketing literature on consideration sets in which studies have either infered or observed consideration sets. Among the research that has estimated consideration sets from choice data alone, researchers have - for the most part - not assumed that consumers engage in search behavior to form their consideration sets. Consideration sets have been directly estimated from choice data (e.g. Siddarth et al. (1995), Chiang et al. (1999)) or predicted based on past purchase (e.g. Mehta et al. (2003)) or advertising (e.g. Allenby and Ginter (1995), Van Nierop et al. (2010)) together with choice data. Van Nierop et al. (2010) validate this approach through an experimental design. In the experiment, they observe advertising, consideration, and choice and find that consideration decisions can be predicted well by advertising. Among the studies that have observed consideration sets are Draganska and Klapper (2009) and Terui et al. (2009). While Draganska and Klapper (2009) observe consideration at the individual level, they only have information on purchase behavior at the aggregate level. They do not model consideration sets as the outcome of a search process. Draganska and Klapper's (2009) goal is to study how the individual-level information on consideration can be best included in aggregate demand estimation. Terui et al. (2009) focus on the question whether advertising has a stronger effect consideration and choice and do not explore the question how consumers form their consideration sets.

Klemperer (1995) provides an excellent overview of the empirical literature on switching costs. Recently, there has been a lot of interest in quantifying switching costs (e.g. Dubé et al. (2010), Handel (2010)). Additionally, there are (at least) two theoretical papers which include both search and switching costs in their analysis: Schlesinger and von der Schulenberg (1991) and Wilson (2009). Wilson (2009) has shown theoretically that in markets where both types of cost are present, cost estimates based on models with only one type of cost are biased upwards. While the theoretical papers provide us with important insights on competition and welfare, quantifying search and switching costs will enable us to see how much competition and welfare are actually affected by these costs.

Previous literature on auto insurance has mostly focused on questions related to asymmetric information. Puelz and Snow (1994) and Dionne et al. (2001) tested for the existence of adverse selection; Cummins and Tennyson (1996), Abbring et al. (2003), and Israel (2007) studied whether evidence for moral hazard in the auto insurance industry can be found. Chiappori and Salanié (2000), Israel (2006a), and Chiappori et al. (2006) suggested tests for the existence of asymmetric information in contractual relationships in the auto insurance industry. In addition to the previously mentioned paper by Dahlby and West (1986), there is one noticeable departure from this line of research on the auto insurance industry. Israel (2005b) studies whether consumers learn from product experience using data on customer retention after not-at-fault car accidents. He finds two groups of customers. Members of the first group defect from their insurance provider shortly after an accident and Israel (2005b) attributes it to customers learning about the service quality in the light of the not-at-fault accident, not being satisfied, and leaving the insurance provider. Members of the second group do not defect from the insurance provider after the not-at-fault accident and also have a lower probability of doing so in the future. Israel (2005b) attributes this to the "lock-in" effect: Consumers either do not leave because they have learned about the good quality of their insurance provider or because they have switching costs that are so high that they prevent them from switching. Because of the nature of his data, Israel (2005b) is not able to differentiate between both explanations. In contrast, in this paper we will be able to disentangle the part of customer inertia that stems from customer satifaction from the part that comes from consumer switching costs since we observe customers' satisfaction levels and their switching behavior. Since almost all data used by previous literature are comprised of auto insurance purchase histories from *one* insurance company, the question how consumers make their purchase decisions could not be studied. Because of the unique characteristics of our data in that we observe which companies consumers considered and which premia they were quoted, we are able to fill this gap in the literature.

4 Data

Our data come from an insurance shopping study conducted in 2006 and 2007 by a large market research company. We observe which companies consumers got quotes from, which they signed up with as well as how much they paid. This gives us information on the number and identity of companies searched and the switching decision. In addition, we observe a nearly complete customer profile containing information on demographics, drivers, cars, location, past claims history, and other insurance products (see Table 1). Even though we observe the most important variables we do not have complete information about the respondent and his insurance policy. We made assumptions about the unobserved variables (e.g. age difference between spouses or annual mileage) which are all listed in Table 2. As information on these variables that we do not observe is required to get a quote, we would not have been able to recover coverage choices and collect the missing premiums without making them. We made assumptions for the unobserved variables in a way that they were consistent within a respondent and across insurance companies as well as across all respondents. This procedure ensures that we did not introduce any heterogeneity in consumer characteristics that did not actually exist in the original data. We also assess the robustness of our results to these assumptions in Section 10.

We observe from which companies each consumer got an insurance quote, but do not have information on the quoted premiums other than the company finally chosen. In order to be able to estimate the model we needed to reconstruct these competitive quotes. Since the data are collected between April 01, 2006 and April 30, 2007 we needed to be able to translate current quotes (that we can collect) into those that are likely to have prevailed at the time the consumer was seeking the quote. The quote reconstruction and time adjustment were done the following way: First, we collected all rate change filings for all insurance companies in all states between April 01, 2006 and December 2008. Rate change filings reflect the average percentage change in premium for customers as a consequence of a change in a company's pricing schedule. Second, we adjusted the premium the consumer paid with the chosen provider by the rate percentage change since purchase time. The underlying assumption behind this is that the consumer's premium would have changed in a manner similar to the average premium. Then, we infered the consumer's policy coverage levels by requesting a new quote that matches the adjusted cited premium (within +/- $10). And finally, using the consumer profile and policy coverage levels we requested the competitive quotes. We did so by going to the websites of all insurance companies that a consumer listed as having collected a quote from, completed the online quote request forms with the consumer's personal information and coverage choices and recorded the quoted premium. To get a clean picture we constrain all respondents to be looking for auto insurance only, own 3 or fewer vehicles, and have 4 or fewer drivers insured under the policy. Table 3 shows reconstructed premia for three consumers. For consumers 005353 and 006963, their previous insurer offers the lowest premium; for consumer 012672, a newly searched company offers the lowest premium. The range of insurance premia, given a consumer's decision to collect two or three quotes in addition to the renewal offer from the previous insurer, is $224.47, $294.62, and $234.36, respectively. These premium range estimates show that, in the data reconstruction process, we were able to recover competitive quotes that are reasonable.

The original data set contained around 2,950 respondents from all US states except Massachusetts. We were able to get the rate change filings from 32 out of the original 50 states. These 32 states contain the biggest and most important auto insurance markets (California, Florida, Texas, Michigan, Pennsylvania, New Jersey, New York, Ohio, Illinois) and also over 70% of the respondents (2,281) from the original data set. Our final data set is considerably smaller for the following reasons: First, recreating the competitive quotes involved recovering the respondent's policy's coverage levels. If the respondent was insured with an auto insurance company that required a social security number in order to get a quote (e.g. AAA) or being a member or a relative of a member of the US military (USAA), it was not possible to recover the coverage levels and reconstruct the competitive quotes. This is by far the most important reason for the shrinkage in data set size. Second, if the respondent was insured with a rather small insurance company which does not offer online quotes or if the respondent was insured with an insurance company that does not offer online insurance quotes in this specific state (e.g. Nationwide does not offer online quotes in California or Florida, but does so in many other states), it was not possible to recover coverage levels and reconstruct competitive quotes. Third, only respondents are included in the final data set for whom we were able to recover the coverage levels and all competitive quotes. Fourth, we excluded 37 respondents who were outliers based on their characteristics (e.g. 95 years old) or reported prices that were not credible (e.g. $200 for a 6-months policy with State Farm in Florida with four drivers and three cars).

Our final data set has 945 respondents. Even though we are only able to use about one third of the original data set because of the aforementioned reasons, our final data are very similar to the original data in almost all characteristics. Table 4 shows a detailed comparison of the two data sets. Besides the slightly higher average number of quotes in the original data, the two data sets are very similar in all other aspects. The smaller average number of quotes is probably due to our data reconstruction as it is easier to reconstruct all competitive prices for a smaller number than a larger number of quotes.

To ensure the representativeness of our data set vis-a-vis the population of consumers purchasing auto insurance, we use representativeness weights provided by the market research company to summarize our data and to make predictions. Representativeness weights are weights assigned to each consumer by the market research company with the goal of creating a data set that is representative of the population with respect to demographics. These weights vary from .45 to 2.06 and in the empirical section, our results do not dramatically change when these weights are ignored giving us confidence that our sample is indeed representative.

Table 5 contains descriptive statistics. Consumers get on average 2.96 quotes (including one from their previous insurer) and collect information on auto insurance from 2.12 sources. They use two different methods to collect quotes (e.g. insurer website, agent, etc.) and need about 1.5 weeks each to gather price quotes and then decide on their insurance provider. Consumers have been with their auto insurance provider, on average, for seven years, and have 1.64 drivers on the policy who driver 1.58 cars. The average premium with a consumer's previous insurer was $764.41, while it is $554.49 for the same group of consumers[8] with the current insurer. Across all consumers in our data the average premium with the current insurer is $592.97. Table 6 shows the mean characteristics for each respondent type (No search/ No switch, Search/ No switch, Search/ Switch). Consumers who neither search nor switch obviously get only one quote - the one that their previous insurer sends to them. Consumers who search, but decide not to switch collect 2.89 quotes and consumers who search and switch gather 3.51 quotes. While there is not much difference in the number of information sources and online quote sites visited across the three respondent types, consumers who search use more methods to obtain quotes. Consumers who neither search nor switch pay the highest average premium, followed by consumers who search, but decide to stay with their previous insurer. The three respondent types are similar in terms of number of vehicles on the policy, number of drivers, household size, hours online per week, vehicle and respondent age. Figure 1 shows the distribution of consideration set sizes. The majority of consumers collect two or three quotes. This is consistent with previous literature on consideration sets which has found average consideration set sizes to lie between two and three (e.g. Mehta et al. (2003), Chiang et al. (1999)). Table 7 contains information about the composition of consideration sets. The left column shows that more than half of the respondents request a quote from Geico, followed by Progressive, Allstate, and State Farm. The right column accounts for limited availability, i.e. the fact that not all auto insurance companies operate in all states. It shows the probability that a respondent considers an insurer if this insurer is offering policies in the respondent's state. The largest differences are seen for American Family, Erie, and Mercury. These are companies which only operate in a few states, but have a relatively strong presence in these states.

Table 8 shows the market shares, retention and conversion rates, i.e. given that a company is being considered, the probability that it will be chosen. Geico, Allstate, Progressive, and State Farm have the largest market shares with about 12% to 19%. Safeco and GMAC have the smallest market shares. The average retention rate is 74%. Allstate and State Farm have higher retention rates than Geico and Progressive

[8]Some consumers did not provide the premium they paid with their previous insurer.

(70% and 72% versus 64% and 63%). As noted previously, a question is whether switching costs stem from satisfaction or market rigidities. In our data, we observe consumer satisfaction with their previous insurer on several dimensions. We will explore the question how switching costs vary with satisfaction levels in the empirical section. The companies with the highest conversion rates are, surprisingly, Erie and Travelers. Among the companies with the lowest conversion rates are not only the companies with the lowest market shares, Safeco and GMAC, but also two companies which have the highest and third-highest market shares, namely Geico and Progressive. It seems that these two companies are good at getting people to request quotes from them, but not good at converting these quotes into actual purchases.

We also have data on radio and TV advertising spending by company and month in 2006 (see Table 9). Radio and TV advertising represents over 90% of an insurance company's total advertising budget. Geico spent the most with an average of nearly $46 million per month, followed by Allstate and Progressive which spent an average of $27 million and $29.5 million respectively and State Farm which spent an average of $19 million. Generally, there is a lot variation in advertising spending as some companies barely advertise or do not do so at all (e.g. Erie, GMAC, Safeco, Travelers).

There is one more piece of data from the shopping survey that we use in our model. Study participants were asked whether they recalled having seen advertising from a specific company during the last 12 months. Table 10 shows advertising recall by company. Geico has the highest recall rate with about 90%, followed by Allstate, Progressive, and State Farm with nearly 80%. Companies with some of the smallest market shares in our data - Erie, Safeco and Mercury - also have the lowest recall rates. Previous literature has found advertising to affect both choice and consideration (e.g. Terui et al. (2009)).

We will now describe how we will use the data in the model formulation. Data on prices, advertising spending and advertising recall will be used to set up a consumer's utility function. We will interact advertising spending with advertising recall to get individual-specific measures of recalled advertising. As mentioned before, our model will describe a consumer's consideration and purchase decisions. We will use information of the number and identity of considered companies to describe a consumer's consideration decision and the identity of the chosen company to describe a consumer's purchase decision.

5 Model

We will first discuss the issues related to the consumer's search method, i.e. simultaneous versus sequential search, and then describe our model.

5.1 Simultaneous versus Sequential Search

In quantifying search costs, it is necessary to first make an assumption about the search process. Baye et al. (2007), describe the two main types of search - simultaneous and sequential. These are defined as follows: Under simultaneous search based on their search costs, consumers commit to conducting a given number of searches before the beginning of the actual search. Even if they get a very desirable search outcome early on, they continue to make all searches they initially committed to. Under sequential search, consumers decide after each search whether they want to continue searching or not. Both search methods have their strengths and weaknesses and Morgan and Manning (1985) have shown from a theoretical perspective that either method can be optimal under certain circumstances. A main advantage of sequential search is that it allows a consumer to stop searching after each price quote. If a consumer gets a good enough quote early in the search process, his expected gains from additional searches might be so small that they do not outweigh the cost of additional searches and he stops searching. The main advantage of the simultaneous search model is also the primary disadvantage of the sequential search model, namely that prices can be gathered quickly. So in a situation where a consumer is trying to obtain quotes from agents and if each agent takes a day or two to provide a quote, then obtaining quotes sequentially could consume more time than a fixed sample search.

Empirical research has used both types of search processes. The sequential approach has been adopted by e.g. Kim et al. (2009) and the simultaneous search approach appears in Mehta et al. (2003). In a recent working paper, De los Santos et al. (2010) estimate both simultaneous and sequential search models using data on online book purchases and find the simultaneous search model to better fit the data. Note that even though we observe the outcome of a consumer's search, i.e. which and how many companies a consumer searched, this data are not enough to identify whether the consumer used a simultaneous or sequential search method. In order to do so, the sequence of searches would have to be observed. For our purposes here, we assume a simultaneous search process by consumers. This assumption is largely driven by our ability to characterize consumer behavior under simultaneous search where there is a large number of alternatives. We are working on extending the approach to sequential search as well (see Section 9.6).

5.2 Model Description

As described in the data section, the outcomes that we observe for each individual are the following: starting with the final decision and working backwards from there, first, we observe a consumer's choice of a specific insurance provider from the subset of providers that the consumer considers. This is a choice problem conditional on the consideration set. Second is the choice of the consideration set itself from the set of all available providers in a consumer's geographic region (state). The choice of a consideration set is comprised of two aspects which the consumer has to choose: the size and composition of the consideration set. Assuming a simultaneous search process, the consumer makes a decision about these two aspects at the same time and before starting the search, i.e. requesting prices. So the integrated utility-maximizing model we develop reflects the following decisions of consumers: (1) which companies to search, (2) how many companies to search, and (3) which company to choose.

We assume there is a set of companies that offer auto insurance in a consumer's state. Consumers know all company characteristics except prices. Searching for prices is costly; therefore consumers only collect price information from a limited number of companies. Consumers do not search their previous insurer as insurance companies automatically send their customers renewal notices and thus consumers receive a free quote from their previous insurer[9, 10]. So the quote from the previous insurer is going to be a consumer's "fall-back option"[11].

Before starting any search, i.e. collecting price information, consumers have to decide which companies to search. Even though consumers do not know prices at this point, we assume - as common in the literature on search - they know the company-specific price distributions. For the sequential search model, Weitzman (1979) showed that the optimal company selection strategy, i.e. the answer to the question which companies to search, is to calculate each company's reservation utility, rank them in decreasing order, start with the top-ranked company, and work your way down. Such a general result does not exist for the simultaneous search model. In the working paper version of their paper (Chade and Smith (2006)), Chade and Smith (2005) show that a simple optimal selection strategy for simultaneous search exists if there is first order stochastic dominance among the company-specific distributions over which the consumer is searching. In that case, the consumer's optimal selection strategy is to calculate the expected utilities, rank them in decreasing order, start with the top-ranked company, and work his way down. For a second special case, namely second order stochastic dominance among the company-specific distributions, Vishwanath (1992) shows that - identical to Weitzman's (1979) result for the sequential search model - the optimal strategy for simultaneous search is for the consumer to rank the options according to his reservation utility and work his way down the list of options. Note that the ordering in Weitzman (1979) and Vishwanath (1992) is according to reservation utilities, while it is according to expected utilities in Chade and Smith (2005).

We are not aware of a strategy which describes the optimal selection strategy in the absence of first

[9]For the purpose of this paper, we define "search" as active search where the consumer has to take action to request a quote and pays a cost of doing so.

[10]In our data set, all consumers get automatic renewal offers from their previous insurers. Even if we observed some consumers not getting renewal notices, i.e. the insurance company refusing to renew a policy, again, that would be the company's and not the consumer's decision.

[11]We call a consumer's free quote from his previous insurer a "fall-back option" and not an "outside option" as outside options usually express the no-purchase option which does not exist in the auto insurance market.

or second order stochastic dominance for the simultaneous search model. Here is some intuition why a general solution is so difficult to find: Picking a set of companies to search is similar to a portfolio problem where each option is characterized by a mean and variance. Suppose one has to make a decision between two options: The first option has a lower mean, but a higher variance than to the second option, i.e. the expected value (utility) from the first option is lower, but there is also a higher chance of getting an outlier with a high value (utility) than with the second option. How should a consumer weigh the lower mean against the higher variance? To circumvent this problem, we assume first order stochastic dominance among the price belief distributions and use the optimal selection strategy suggested by Chade and Smith (2005). We assume a specific form of first order stochastic dominance, namely, that the price belief distributions have company-specific means, but the same variance across all companies.

The optimal number of searches is the solution to an optimization problem where the consumer picks the number of searches that maximizes the net benefit of search, i.e. expected benefit of search minus cost of search. After consumers have searched the optimal set of companies, they know the prices charged by these companies and all uncertainty has been resolved. Given their consideration sets, consumers then decide to purchase from the company with the highest utility. This can be viewed as a choice model conditional on consideration sets.

Formally, there are $i = 1, \ldots, N$ consumers who live in $m = 1, \ldots, M$ states with a different number and composition of insurance companies in each market. We view each state as a market. Across all markets there are $j = 1, \ldots, J$ insurance companies. The set of available auto insurers in state m is denoted by Ψ_m. Consumer search costs are denoted by c and the number of searches that a consumer conducts is denoted by k. We assume that search and switching costs are homogeneous across consumers. This is because we do not have access to panel data. In the empirical section, we make these costs a function of observable characteristics to account for this source of heterogeneity. The set of companies a consumer searches is denoted by $S_i = \{1, \ldots, k\}$, the set of companies a consumer does not search is denoted by $\overline{S}_i = \{1, \ldots, \Psi_m - k - 1\}$. The companies a consumer searches and his previous insurer form the consumer's consideration set C_i, i.e. $C_i = S_i \cup \{j_{PI}\}$ so $C_i \subset \Psi_m$ and $C_i = \left\{1, \ldots, \tilde{k}\right\}$ where $I_{ij,t-1}$ denoted the previous insurer and $\tilde{k} = k + 1$. Note that we can exhaustively decompose the set of available auto insurers in state m into the set of searched companies, the set of not searched companies, and the previous insurer, i.e. $\Psi_m = S_i \cup \overline{S}_i \cup \{j_{PI}\}$.

Consumer $i's$ indirect utility for company j is given by

$$u_{ij} = \alpha_{ij} + \beta_i I_{ij,t-1} + \gamma p_{ij} + \epsilon_{ij} \tag{1}$$

with

$$\alpha_{ij} = \tilde{\alpha}_j + X_{ij}\rho$$

and

$$\beta_i = \tilde{\beta} + Z_i\kappa$$

ϵ_{ij} captures a component of the utility that is observed by the consumer, but not the researcher. α_{ij} are consumer-specific brand intercepts. While we are not able to allow for unobserved heterogeneity in brand intercepts, we control for a rich set of observed heterogeneity. X_{ij} is a vector of the product- and consumer-specific attributes among them an interaction term between advertising spending and recall. "Recall" captures consumer $i's$ self-reported recall of having seen advertisements from insurer j during the insurance shopping period. By interacting company-specific advertising spending with consumer-specific advertising recall, we ensure that only advertising consumer i is aware of enters his utility function. Additionally, X_{ij} also contains demographic factors (e.g. age, marital status), psychographic factors (e.g. loyalty, interest in finance, etc.), quoting channels (e.g. agent, insurer website) and regional fixed effects (e.g. new england, mid-atlantic). These variables are common across j. So while they drop out of the conditional choice decision, they may play a role in the search and consideration decisions. Finally, X_{ij} also contains $X_{i,I_{ij,t-1}}$, the satisfaction with the previous insurer. This variable is also common across all j and will drop out of the conditional choice decision. However, it influences search and consideration in that ρ, the coefficient of this variable, is negative. It explains that consumers will search less if they are are highly satisfied with the current insurance provider. β_i describes consumer inertia. $I_{ij,t-1}$ is an dummy variable indicating whether company j is consumer $i's$ previous insurer. Again, we control for observable heterogeneity in consumer inertia β_i . Z_i contains demographic factors (e.g. age, gender), psychographic factors (e.g. loyalty, interest in finance), and customer satisfaction with the previous insurer (e.g. with claims process, pricing). Note that since this is interacted with $I_{ij,t-1}$, it plays a role in the conditional choice decisions. In interpreting our results, we will attribute the part of customer inertia associated with demographic and psychographic factors to customer switching costs and the part of customer inertia associated with customer satisfaction to customer satisfaction. γ captures a consumer's price sensitivity and p_{ij} denotes the price charged by company j. Note that - in contrast to the consumer packaged goods industry - in the auto insurance industry, prices depend on consumer characteristics. This has implications for the estimation that we will discuss in detail at the end of this section.

As common in the literature on search, we assume that - even though consumers do not know the prices each specific company would charge them - they know the distributions of prices present in the market. We assume prices p_j have an EV Type I distribution with location parameter η_j and scale parameter μ (as in Mehta et al. (2003))[12]. As consumers know the distribution of prices in the market, they know η_j and μ. Notice that these price distributions are common across consumers. We describe in detail at the end of this section how we estimate these common price distributions conditional on consumer characteristics.

We will now describe the model from the perspective of a consumer. In the previous section, we established that a consumer's decision which companies to search depends on the expected indirect utilities (EIU; Chade and Smith (2005)) where the expectation is taken with respect to the characteristic the consumer is searching for - these are prices in our case. So consumer $i's$ EIU (expectation with respect to price) is given by

$$E\left[u_{ij}\right] = \alpha_{ij} + \beta_i I_{ij,t-1} + \gamma E\left[p_j\right] + \epsilon_{ij} \tag{2}$$

Consumer i observes these EIUs for every brand in his market (including ϵ_{ij}). To decide which companies to search, consumer i ranks all companies according to their EIUs (Chade and Smith 2005)[13] and then picks the top k companies to search them. Notice that the expectation of prices is common across consumers because we assume there is a common price distribution across individuals.

To decide on the number of companies k a consumer searches, the consumer calculates the net benefit of all possible search sets *given the ranking of EIUs*, i.e. if there are N companies in the market, the consumer can choose among $N-1$ search *sets* (as one free quote comes from the PI). A consumer's benefit of a searched *set* is, similarly, given by the expected *maximum* utility among the searched brands. Given the EV distribution of prices, the maximum utility has also an EV distribution (see Technical Appendix A)

$$\max_{j \in S} u_{ij} \sim EV\left(\frac{1}{b}\ln \sum_{j \in S} \exp\left(b a_{ij}\right), b\right) \tag{3}$$

with $a_{ij} = \alpha_{ij} + \beta_i I_{ij,t-1} + \epsilon_{ij} + \gamma\eta_j$ and $b = \frac{\mu}{\gamma}$. Let us further define $\tilde{a} = \frac{1}{b}\ln\sum_{j\in S}\exp\left(b a_{ij}\right)$. Then the benefit of a searched set is given by

$$E\left[\max_{j \in S} u_{ij}\right] = \tilde{a}_S + \frac{e_c}{b} \tag{4}$$

where e_c denotes the Euler constant. The consumer picks the size of S denoted by k which maximizes his net benefit of searching denoted by $\Gamma_{i,k+1}$, i.e. expected maximum utility among the considered companies minus the cost of search

[12] $f(x) = \mu \exp\left(-\mu\left(x-\eta\right)\right)\exp\left(-\exp\left(-\mu\left(x-\eta\right)\right)\right)$ and $F(x) = \exp\left(-\exp\left(-\mu\left(x-\eta\right)\right)\right)$ with location parameter η and scale parameter μ. Mean is $\eta + \frac{e_c}{\mu}$ and variance $\frac{\pi^2}{6\mu^2}$ where e_c is Euler constant (Ben-Akiva and Lerman (1985), p. 104).

[13] Note that the EIU for the PI will not be part of the company ranking as consumers get a free renewal offer from the PI and do not need to search for PI's price.

$$\Gamma_{i,k+1} = E\left[\max_{j \in S \cup \{j_{PI}\}} u_{ij}\right] - kc \tag{5}$$

Note that in this case, any consideration set contains at least the quote from the previous insurer. The consumer picks the number of searches k which maximizes his net benefit of search. If a consumer decides to search k companies, he pays kc on search cost, but will have $k+1$ companies in his consideration set.

This concludes our description of how a consumer forms his consideration set. Once a consumer has formed his consideration set and received all price quotes he requested, all price uncertainty is resolved. Both the consumer and the researcher observe prices. The consumer then picks the company with the highest utility among the considered companies, i.e.

$$j = \arg\max_{j \in C_i} u\left(X_{ij}, Z_i, I_{ij,t-1}, p_{ij}, \epsilon_{ij}; \alpha_{ij}, \beta_i, \gamma\right) \tag{6}$$

where p_{ij} are now the quoted price for consumer i by company j.

6 Estimation

We start out by pointing out the crucial differences between what the consumer observes and what the researcher observes:

1. While the consumer knows the distribution of prices in the market, the researcher *estimates* the parameters defining the price distributions.

2. While the consumer knows each company's position in the EIU ranking, the researcher only partially observes the ranking by observing which companies are being searched and which ones are not being searched.

3. In contrast to the consumer, the researcher does not observe ϵ_{ij}.

To address the first issue that the researcher estimates the parameters defining the price distributions, we will integrate over the empirical distributions of these parameters given sampling error associated with estimating the parameters of the price distributions (McFadden (1986))[14]. To address the second issue, we point out that partially observing the ranking contains information that allows us to estimate the composition of consideration sets. Since the consumer ranks the companies according to their EIU and only searches the highest ranked companies, the researcher knows from observing which companies are searched that the EIUs

[14]For now, we disregard this issue in the estimation and treat the parameters of the price distribution as if the researcher observed them.

among all the searched companies have to be larger than the EIUs of the non-searched companies or, to put it differently, that the minimum EIU among the searched companies has to be larger than the maximum EIU among the non-searched companies, i.e.

$$\min_{j \in S} \left(E\left[u_{ij}\right]\right) \geq \max_{j' \notin S} \left(E\left[u_{ij'}\right]\right) \tag{7}$$

Since a consumer decides simultaneously which and how many companies to search, the following condition has to hold for any searched set

$$\min_{j \in S} \left(E\left[u_{ij}\right]\right) \geq \max_{j' \notin S} \left(E\left[u_{ij'}\right]\right) \quad \cap \quad \Gamma_{ik} \geq \Gamma_{ik'} \ \ \forall k \neq k' \tag{8}$$

i.e. the minimum EIU among the searched brands is larger than the maximum EIU among the non-searched brands *and* the net benefit of the chosen searched set of size k is larger than the net benefit of any other search set of size k'.

And finally, we account for the fact that the researcher does not observe ϵ_{ij} by assuming ϵ_j has an EV Type I distribution with location parameter 0 and scale parameter 1 and integrate over its distribution. Then probability that a consumer picks a consideration set Υ is

$$\Pr\left(C_i = \Upsilon \mid X_{ij}, Z_i, I_{ij,t-1}, \eta_j, \mu; \theta\right) = \Pr\left(\min_{j \in S} \left(E\left[u_{ij}\right]\right) \geq \max_{j' \notin S} \left(E\left[u_{ij'}\right]\right) \quad \cap \quad \Gamma_{i,k+1} \geq \Gamma_{i,k'+1} \ \ \forall k \neq k'\right) \tag{9}$$

with $\theta = \{\alpha_{ij}, \beta_i, \gamma, c\}$. Note that the quote from the previous insurer directly influences consumer's choice of the size of a consideration set. A consumer is going to renew his insurance policy with his previous provider if the utility of doing so is larger than the expected net benefit $\Gamma_{i,k+1}$ of any number of searches.

Let us now turn to the purchase decision given consideration. Let $\bar{j}$ be the base brand for consumer i. Then the consumer's choice probability conditional on his consideration set is

$$\Pr\left(y_i = j \mid X_{ij}, Z_i, I_{ij,t-1}, p_{ij}, C_i; \theta\right) = \left(u_{ij} \geq u_{ij'} \quad \forall j \neq j', \quad j, j' \in C_i\right) \tag{10}$$

where y_{ij} is a binary variable indicating whether this brand was chosen, p_{ij} are now the quoted prices, and $\alpha_{\bar{j}}$ is the base brand intercept for which we use the brand intercept for Geico. Note that there is a selection issue: Given a consumer's search decision, ϵ_{ij} do not follow an EV Type I distribution and the conditional choice probabilities do not have a logit form. The consumer's unconditional choice probability is given by

$$\Pr\left(y_i = j \mid X_{ij}, Z_i, I_{ij,t-1}, p_{ij}, C_i; \theta\right) = \Pr\left(C_i = \Upsilon \mid X_{ij}, Z_i, I_{ij,t-1}, \eta_j, \mu; \theta\right) \Pr\left(y_i = j \mid X_{ij}, Z_i, I_{ij,t-1}, p_{ij}, C_i; \theta\right) \tag{11}$$

In summary, the researcher estimates the price distributions, only partially observes the utility rankings, and does not observe ϵ_{ij} in the consumer's utility function. Accounting for these differences compared to the consumer, we derived an estimable model with consideration set probability given by (9) and the conditional and unconditional purchase probabilities given by (10) and (11).

We maximize the joint likelihood of consideration set and purchase. The likelihood of our model is given by

$$L = \prod_{i=1}^{N} \prod_{l=1}^{L} \prod_{j=1}^{J} \Pr\left(C_i = \Upsilon \mid X_{ij}, Z_i, I_{ij,t-1}, \eta_j, \mu; \theta\right)^{\vartheta_{il}} \Pr\left(y_i = j \mid X_{ij}, Z_i, I_{ij,t-1}, p_{ij}, C_i; \theta\right)^{\delta_{ij}} \tag{12}$$

where ϑ_{il} indicates the chosen consideration set and δ_{ij} the purchased company. $\theta = \{\alpha_{ij}, \beta_i, \gamma, c\}$ is the set of parameters to be estimated. Neither the consideration set probability $\Pr\left(C_i = \Upsilon \mid X_{ij}, I_{ij,t-1}, \eta_j, \mu; \theta\right)$ are shown is equation (9) not the conditional purchase probability $\Pr\left(y_i = j \mid X_{ij}, I_{ij,t-1}, p_{ij}, C_i; \theta\right)$ as shown in equation (10) have a closed-form solution. We therefore use a simulation approach to calculate them. In particular, we simulate from the distribution of the unobserved component of the utility ϵ_{ij}. Calculating the proportion of times that the optimality condition in equation (9) or (10) is satisfied would result in lumpy probabilities which cannot be optimized using common optimization routines and would require non-gradient based methods or a very large number of draws[15]. Therefore we smooth the probabilities using a multivariate scaled logistic CDF. We describe our estimation approach in detail in Technical Appendix B. There we also describe a simulation study whose purpose is to evaluate the properties of our estimation approach and show that it recovers the true model parameters.

6.1 Comparison with Previous Approaches

Our approach of modeling the probability of observing a consideration set is different from the ones suggested by previous literature. Siddarth et al. (1995) and Chiang et al. (1999) estimate choice and consideration set probabilities from choice data alone. Siddarth et al. (1995) estimate consideration sets based on past purchase, while Chiang et al. (1999) enumerate all possible consideration sets (given knowledge of the

[15]Given the number of companies in our market, a very large number of draws would be required to achieve smoothness of the probabilities.

purchased brand) and estimate the probabilities of each of them. Given a consideration set, the probability of choosing a brand is given by the multinomial logit probability. Neither Siddarth et al. (1995) nor Chiang et al. (1999) model the *process* that leads to the formation of consideration sets. By contrast, in our case, the consideration set is the outcome of a consumer's explicit trade-off between search costs and a set of alternatives that provide consumer utility. Since search costs are a model primitive, our approach allows us to look at changes in consideration set sizes as a consequence of changes in search costs. In contrast to the aforementioned papers and similarly to our approach, Mehta et al. (2003) explicitly model the search process that leads to the formation of consideration sets although they also only use choice data. Intuitively speaking, they use a choice model to describe a consumer's decision of consideration set. They enumarate all possible consideration sets (by size and composition) and assume that the consumer picks the one which provides him with the largest benefit (net of cost) by fitting the choice model. This approach is only feasible for a small set of alternatives since the number of consideration sets dramatically increases with the number of alternatives[16]. With our data where consumers can choose from up to 17 companies, it is infeasible to implement their approach. By using the theory based on Chade and Smith (2005) and the associated EIU ranking, our approach effectively reduces the number of search sets from $2^N - 1$ to N and thereby allows us to estimate consideration set probabilities in categories with a large number of brands. Note that equation (9) is flexible enough to estimate probabilities of all possible combinations of consideration set sizes and compositions. Suppose there are four brands in a market denoted by A, B, C, and D; a consumer considers brands A and B and purchases brand A. Assuming the considered brands are observed and using the approach suggested by Mehta et al. (2003) we would fit a choice model with 4 possible outcomes[17] where the consumer chooses one of these outcomes[18]. Using our approach, we would calculate the *joint* probability that the minimum EIU among the considered brands is larger than the maximum EIU among the non-considered brands and that the consideration set of size two (which contains the two brands with the largest EIUs as all brands are ordered according to their EIUs) provides the consumer with a lrger net benefit than a consideration set of any other size. Thus our approach is able to predict any consideration set size and composition combination and to do so independently of the number of alternatives by describing the probability of observing a consideration set as the joint probability of *two* events. The additional flexibility of our approach does come at some cost. Despite the EV Type I distribution of ϵ_j there exists no closed-form solution for the expression in equation (9). We solve this problem by using a smoothed simulated MLE approach which we describe in detail in Section 7.

[16]Given knowledge of the purchased brand, some of the combinations can be eliminated as the purchased brand always has to be part of the consideration set.

[17]The possible outcomes are AB, ABC, ABD, and ABCD.

[18]If consideration sets are not observed and we use the approach by Mehta et al. (2003), eight outcomes are possible(A, AB, AC, AD, ABC, ABD, ACD, ABCD)

6.2 Empirical Challenges with Insurance Data

Insurance prices comprise - in contrast to CPG or most other product prices - personal consumer and policy characteristics, e.g. coverage choices, payment arragements, etc. Prices as given cannot be used because they not only reflect the inherent price variation we need to estimate the price coefficient, but also these personal and policy characteristics. For example, consumer 1 and consumer 2 can be equally responsive to prices, but consumer 1 needs an auto insurance for himself and his one car, while consumer 2 needs a policy for four cars and three drivers one of whom is under 25 years. Both consumers get a quote from the same company; consumer 1 for \$500 and consumer 2 for \$2,000. If we now used these quoted prices, our model would incorrectly predict that consumer 2 is less responsive to prices than consumer 1. To solve this issue we use an approach used by the literature on quality adjusted prices. We first regress prices on personal and policy characteristics and then use the residuals from this hedonic pricing regression in the model estimation[19]. The "quality" measures, i.e. personal and policy characteristics, that define insurance prices fall under the Freedom of Information Act (FOIA) and are publicly known. As described previously we have data on almost all these characteristics.

Recall that in our model formulation prices are needed at both the consideration stage and the choice conditional on consideration stage. We now describe how we obtain these prices. To get the prices necessary to estimate the latter part of the model, i.e. choice given consideration, as previously mentioned we assume prices have an EV Type I distribution and regress all quoted prices on personal and policy characteristics, i.e.

$$p_{ij} \sim EV\left(\xi_{ij};\mu\right) \quad with\ \xi_{ij} = D_i\gamma \tag{13}$$

where D_i contains personal characteristics as well as coverage choices and γ is a coefficient vector. We then calculate the price residuals $p_{ij}^{res} = p_{ij} - \hat{p}_{ij}$. We use the residuals p_{ij}^{res} as prices in the consumer choice problem.

The model states that before starting collecting quotes, prices are unknown to the consumer. But consumers know the company-specific *distributions* of prices present in the market[20]. We estimate the parameters of these distributions using prices from our data. To reflect consumer's knowledge of the prices present in the market before the search, we only use the prices charged by the previous insurers (PI) of consumers in our sample. Again, we assume prices have an EV Type I distribution and regress all prices charged by previous insurers (pooled across consumers) on personal and policy characteristics, i.e.

[19]This approach has been previously used when estimating demand for pharmaceuticals (e.g. Berndt et al. (2003)) or cars.
[20]Note that we check the robustness of this assumption in a later counterfactual.

$$p_{ij}^{PI} \sim EV\left(\xi_{ij}^{PI};\mu^{PI}\right) \quad with\ \xi_{ij}^{PI} = D_i\varsigma \tag{14}$$

where ς is a coefficient vector. We then calculate the price residuals $p_{ij}^{PI,res} = p_{ij}^{PI} - \hat{p}_{ij}^{PI}$ and group them by company. These residuals also have an EV Type I distribution with company-specific means and the same variance across all companies

$p_j^{PI,res} \sim EV\left(\eta_j^{PI,res};\mu^{PI}\right)$

We use these price residual distributions as the company-specific price distributions known by the consumer at the consideration stage.

Because of these empirical challenges associated with insurance prices, we estimate our model in two steps. First, we estimate the two pricing regressions as described in equations (13) and (14). We then estimate our model as described previously. To do so, we have to modify the loglikelihood (equation (12)) to account for the two pricing regressions. The modified likelihood of our model is given by

$$L = \prod_{i=1}^{N}\prod_{l=1}^{L}\prod_{j=1}^{J} \Pr\left(C_i = \Upsilon \mid X_{ij}, PI_{ij}, \eta_j^{PI,res}, \mu^{PI};\theta\right)^{\vartheta_{il}} \Pr\left(y_i = j \mid X_{ij}, PI_{ij}, p_{ij}^{res}, C_i;\theta\right)^{\delta_{ij}} \tag{15}$$

where $\eta_j^{PI,res}$ and μ^{PI} denote the parameters estimated based on the pricing regression shown in equation (14) and p_{ij}^{res} denotes the residuals from the pricing regression shown in equation (13). As noted previously, since $\eta_j^{PI,res}$ and μ^{PI} are estimated quantities, we need to account for sampling error when computing standard errors.

7 Identification

As noted previously, the set of parameters to be estimated θ is given by $\{\alpha_{ij}, \beta_i, \gamma, c\}$. For the purpose of this discussion, let us split θ into $\theta_1 = \{\alpha_{i1} - \alpha_{iJ}, \ldots, \alpha_{i,J-1} - \alpha_{iJ}, \gamma, \beta_i\}$ and $\theta_2 = \{\alpha_{iJ}, c\}$. θ_1 contains all preference parameters up to the base brand intercept, θ_2 contains the base brand intercept and the search cost parameter. Using choice data alone, θ_1 could be identified. Observing a consumer's consideration set allows us to additionally identify θ_2.

We first discuss the identification of the θ_1 parameters. With a data set where the researcher observes consumers making a choice among several alternatives, the differences between the brand intercepts and the base brand intercept are identified together with preferences for other attributes and switching costs. Recall from the introduction that previous researchers identified switching costs from within consumer persistence in choice over time after controlling for consumer heterogeneity. In our case, we only observe information on the previous and current insurer. Thus we are not able to account for all unobserved sources of heterogeneity. A valid concern in estimating the switching cost parameter β_i is that it reflects these sources of heterogeneity rather than switching costs. We note that since we observe a comprehensive set of consumer characteristics we are able to account for a wide range of observable heterogeneity among consumers. Our identification of switching costs rests on our ability to control for heterogeneity via these sources. Additionally, we also control for a variety of observed factors that allow for switching costs to vary across consumers including factors such as satisfaction with the previous insurance provider. In our data, the fact that we observe consumers who switch and consumers who do not switch is crucial in identifying switching costs.

Next, we turn to the identification of θ_2. Intuitively speaking, the size of the consideration set will pin down search costs. We can only identify a range of search costs as it is utility-maximizing for all consumers with search costs in that range to search a specific number of times. For example, suppose it is optimal for a consumer to search once if his search costs lie between three and five, twice if his search costs lie between one and three and three times if his search costs lie between zero and one. Then by observing a consumer searching twice, we know that his search costs must lie between one and three. Beyond the fact that a consumer's search cost lie within this range which rationalizes searching a specific number of times, the variation in our data does not identify a point estimate for search costs. The search cost point estimate will be identified by the functional form of the utility function and the EV Type I distribution of the unobserved part of the utility.

The base brand intercept and search costs do not intuitively appear to be separately identified as they both only appear together in describing the benefit of a consideration set of a certain size (equation (4)). However, identification is facilitated due to the specific functional form assumption made and to the nonlinearity of that expression[21].

8 Results

8.1 Pricing Regressions

We start by discussing the results from the two pricing regressions. Recall that the pricing regression is required for two purposes: (a) to characterize consumers' knowledge about prices in the market and (b) for the price measure to be used in the estimation. We use the pricing regressions (equations (13) and (14)) for (a) and the residuals from these pricing regressions for (b). The results are shown in Table 11. The middle columns show the parameter estimates for the previous insurer prices, the columns on the right

[21] Stated differently, if there was a "fixed" component of the search cost that did not vary by the number of searches, this fixed cost would not be separetely identified from the base brand intercept.

show the parameter estimates for all quoted prices. We allow the location parameter of the EV Type I distribution to vary with various relevant characteristics. The characteristics we use describe a consumer's personal characteristics (age, gender, marital status, etc.), the other drivers on this policy, the insured cars (model age, make, and class), past tickets and accidents, coverage choices, discounts he might receive from having multiple insurance policies with his current auto insurer, and his location (i.e. state and community type). Note that most coefficients have expected signs and reasonable magnitudes. For example, insuring a second vehicle costs about \$130[22] for a 6-month policy and having two or more accidents makes an insurance policy about \$200 more expensive[23]. The coefficient estimates across these two price regressions are also quite similar in magnitude and lie within two standard errors of each other suggesting some internal consistency. For example, the results from the previous insurer (PI) price distribution suggest that having home owner insurance with your auto insurance provider saves \$26.15 on the auto insurance premium, while the results from the all quotes price regression suggest that having home owner insurance with your auto insurance provider saves \$21.58. We conducted likelihood ratio tests and can reject the null models where only location and scale parameters are estimated (and the location parameter is not a function of the observable characteristics) at $p < .001$ for both the pricing regression using only prices charged by previous insurers and the pricing regression using all quoted premia.

As issue that arises is that if characteristics completely explain prices, consumers have no need to search as they will know the prices each insurance company will charge them. If there is some unexplained proportion of prices, it will motivate consumer search[24]. Table 12 describes the means and variances of the company-specific residual distributions. First, the company-specific means intuitively make sense as holding everything else constant the same person expects to pay more for the same policy with State Farm than with Geico (holds for residuals from both pricing regressions). Second, across the two pricing regressions, the company-specific residual means are directionally the same in most cases and, if we ranked the companies according to their residual means, companies which belong to the group with high residual means in the previous insurer price regression would remain in that group based on the results from the all quotes price regression even if their exact rank might change slightly. The same applies to the companies with medium and low residual means. Third, the standard deviations are substantial. This implies that there remain factors that we do not observe that might change individual prices and consumers have to search to learn the premium an insurance company's is going to charge them.

[22]\$134.34 based on the PI prices and \$128.04 based on the all quoted prices.

[23]\$190.80 based on the PI prices and \$196.69 based on all quoted prices.

[24]An example of characteristic that might be related to the unexplained proportion of prices is a consumer's Social Security Number. Insurance companies use them to learn a consumer's credit score and set insurance prices accordingly. Consumers might not know their credit score and therefore not know the price an insurance company is going to charge them.

8.2 Results from the Proposed Model

Next, we turn to the parameter estimates of our model of search and purchase. Table 14 shows the results for two different model specifications. In Model 0, we assume consumers only have search, but no switching costs and in Model 1, we allow for both types of costs. All paramters are significant in Models 0 and 1. Both AIC and BIC clearly support the addition of switching costs to the model. Wilson (2009) predicts a potential upward bias in cost estimates for markets in which both search and switching costs are present, but only either search or switching costs are estimated. To translate search costs from utils into dollars, we divide the search cost parameter by the price coefficient[25]. We find search costs of $48.24 in Model 0 where we do not estimate switching costs and search costs of $48.29 in Model 1 where we estimate both search and switching costs. We conclude that in our model, we do not find evidence for an upward bias in the search cost estimate when switching costs are ignored. One explanation is that the upward bias Wilson (2009) found was due to him not observing the extent of consumer search. While Wilson (2009) only used the proportion of consumers who searched at the aggregate level, we observe and model the amount of each consumer's search behavior at the individual level.

Tables 15 - 17 show two model specifications where we control for demographic and psychographic factors, quoting channels, regional fixed effects, and customer satisfaction levels in the utility function and for observed heterogeneity in both search costs and in customer inertia. In Model 2, we control for demographic and psychographic factors[26], quoting channels[27], regional fixed effects, and customer satisfaction levels in the utility function. In Model 3, we additionally let search costs vary with quoting channels and customer inertia vary with age, gender, income, education, psychographic factors, and satisfaction levels. AIC and BIC support the inclusion of all these variables. In interpreting our results, we will concentrate on Model 3[28].

The upper half of Table 16 shows how search costs vary with observable consumer characteristics. The search cost constant now denotes the cost of collecting a quote through an agent. Search costs are 45 - 50% lower when consumers collect their quotes through online channels (insurer website, online quoting service) compared to collecting quotes through agents. The cost of collecting a price quote by mail are 31% higher than through agents and the cost of collecting a price quote through an insurer calling center are nearly 20% lower than through agents. The lower half of the Table 16 shows how customer inertia varies with observable characteristics. These observable characteristics can be divided into three groups: The first group consists

[25] Note that we estimate search costs $c = \exp(\tau)$ and report τ in Tables 14 - 17.

[26] The psychographic factors are Attitude towards Auto Insurance Shopping & Switching, New Technology Adoption, Technology Usage, Loyalty, and Interest in Finance (see Table 18).

[27] The quoting channels are local agents, mail, insurer websites, online quoting services, and calling centers.

[28] Std. errors for the coefficients in Model 3 are currently being calculated.

of the intercept and demographic and psychographic factors that are uncorrelated our observed measures of customer satisfaction, i.e. gender, education, and "New Technology Adoption"; the second group of demographic and psychographic factors that are correlated with customer satisfaction, i.e. age, income, and the remaining four psychographic factors; and the third group consisting of the customer satisfaction measures. We will attribute the part of customer inertia associated with the first group of variables to customer switching costs and the part of customer inertia associated with the third group of variables to satisfaction-based inertia or loyalty. The second group of variables, i.e. those demographic and psychographic factors that are correlated with customer satisfaction measures, can be attributed to either customer switching costs or satisfaction based inertia. Here, we will attribute it to customer switching costs and report estimates under the alternative of attributing it to satisfaction-based inertia in the robustness checks. We find no significant effects of gender, but significantly increasing switching costs with age, income and education levels. Several psychographic factors also significantly influence switching costs. The more consumers dislike the auto insurance shopping and switching process, the higher their switching costs. While loyalty increases switching costs, Interest in Finance decreases them. Interest in Finance captures a consumer's interest in financial markets and publications. This factor is likely to also capture a consumer's financial savviness which decreases his switching costs as he is more willing to switch to a new auto insurance provider if this is advantageous to him. And finally, we also look at how customer inertia varies with satisfaction in several areas. We find three significant effects: Customer inertia increases with satisfaction with prices, policy offerings, and overall satisfaction. Satisfaction in the other areas (billing and payment process, interaction with insurer, and claims service) has no significant effects.

8.3 Model Fit and Predictive Ability

Before discussing the levels of search and switching costs, we first want to see how well our model predicts consideration and purchase. The average size of a consideration set is 2.96 in our data, while our model predicts an average size of 2.78. Table 23 shows how the predicted composition of consideration sets, the purchase market shares, and the retention rates compared to the ones observed in the data. Our model predicts the composition of consideration sets within two percentage point for all companies with the exception of Geico, Progressive, and State Farm[29]. The purchase market shares are predicted within the same range for all companies with the exception of State Farm[30]. Our model underpredicts retention rates by about 2 percentage points[31]. The largest differences between predicted and actual retention rates

[29]The differences between predicted and actual consideration set shares for Geico, Progressive, and State Farm are 2.06, 2.50, and 2.37 percentage points, respectively.

[30]The difference in actual and predicted purchase market shares is 2.60 percentage points.

[31]The actual average retention rate is 72%; our estimates predict an average retention rate of 70%

are found for Geico, GMAC, MetLife, and Progressive. We conclude that our model is able to describe the patterns observed in the data.

8.4 Search and Switching Costs

After establishing the validity of our results, we will now discuss the levels of search and switching costs in terms of dollars. We find search costs (per search) of \$83.42 through agents, \$109.84 through mail, \$68.01 through calling centers, \$47.06 through insurer websites, and \$41.19 through online quoting services. Dahlby and West (1986) estimated search costs to be between CAD \$127.65 - CAD \$553.61 (adjusted for inflation) for auto insurance in the seventies in Canada. Our search cost estimates are lower. We attribute this to the data being collected partially in rural Canada where the density of agents was lower, to the introduction of calling centers and to the internet during the last decades which lowered consumer search costs. Similar to search costs, we divide the switching cost parameter by the price coefficient to calculate a dollar value for switching costs[32]. Average switching costs are \$84.19, but there is a lot of variation in switching costs based on observable characteristics as shown in Figure 2. Cummins et al. (1974) found switching costs to equal 20% of the premium paid which would be \$118.59 with our data. Berger et al. (1989) found switching costs to lie between CAD \$178.86 and CAD \$370.49 (adjusted for inflation) so that their switching cost estimate in today's dollar is somewhat higher. We conclude that our switching cost estimates are smaller than those found by the previous literature. We attribute this to the fact that previous literature was not able to differentiate between consumer inertia due to customer satisfaction and consumer inertia due to switching costs and attributed all observed inertia to switching costs. If we follow this approach, we find the median value of consumer inertia to be \$284.30 which is similar to the estimates reported by Berger et al. (1989). An interesting question that was raised in the introduction is to what extent the high retention rate that we observe in the auto insurance industry is caused by consumer satisfaction and to what extent by market ridigities. We will explore this question in detail in a counterfactual (Section 10.1).

8.5 Price Elasticities and Implied Mark-ups

Table 19 shows the implied own-price elasticities. The mean price elasticity is -2.10 with a minimum of -2.80 and a maximum of -1.01. The mark-up implied by the average price elasticity is 32.26%. Mark-ups in the insurance industry are calculated as premiums earned minus losses (claims) and administrative expenses. Administrative expenses have to be filed with the pricing schedules, but we have no information on them. Two additional difficulties in calculating industry mark-ups are that insurance companies do not report

[32] Similar to the search cost coefficient, we estimate the coefficient capturing consumer inertia $\beta_i = \exp\left(\tilde{\beta}' + Z_i\kappa'\right)$ and report $\tilde{\beta}'$ and κ' in Tables 14 - 17.

profits and profit margins by business line and that earned premiums are re-invested and not all insurance companies report this income separately. But we were able to collect data on premiums earned and losses for private passenger auto insurance for 2007. Using this data, we calculated the implied mark-ups as shown in Table 22. In interpreting these mark-ups and comparing them to those implied by our model, we have to be careful as they do neither include administrative expenses which would push them down nor premium re-investments which would push them up. We conclude that the mark-ups implied by our model are similar to those implied by industry data. Thus it provides some check of the validity of our estimation results.

To summarize, we find substantial search and switching costs present in the auto insurance industry with a large amount of heterogeneity based on observable characteristics. The estimated price elasticities suggest mark-ups similar to those implied by industry data.

9 Counterfactuals

In our model, customer retention can come from four sources: search costs, switching costs, customer satisfaction, and consumer preferences other than those associated with inertia. In the first counterfactual, we study how much each of these factors contributes to the observed high retention rate the in U.S. auto insurance industry. In the second counterfactual, we evaluate the value of customer satisfaction to insurance companies. And in the last counterfactual, we study whether insurance companies should increase or decrease their premia and whether they should do so uniformly for all customers or differentiate between new and existing customers.

9.1 Decomposition of Customer Retention

We study how each of the four potential sources - search costs, switching costs, customer satisfaction, and consumer preferences other than those associated with inertia - contribute to the high observed consumer retention rate. The results are shown on the left side of Table 27. The retention rate predicted by our model under current consumer behavior is 70.60%. When search costs are eliminated, the retention rate decreases to 43.41%. When customer satisfaction is also additionally eliminated, the retention rate decreases to 14.42%. Finally, when switching costs are also eliminated, the retention rate drops to 7.16%. We conclude that there are two main reasons for the high observed retention rate in the U.S. auto insurance industry: search costs and customer satisfaction. 38% of the retention in this industry is due to search costs and 41% due to customer satisfaction. After accounting for customer satisfaction, we find switching costs to drive only 11% of the high customer retention.

9.2 Value of Customer Satisfaction

Insurance companies invest heavily in measures aimed at increasing customer satisfaction. The cost of such measures are relatively easy to quantify; quantifying the benefits is more difficult. Increasing customer satisfaction can have three effect on consumer behavior: (1) It can prevent consumers from requesting price quotes from other insurance companies; (2) given consumers' consideration sets, it can prevent consumers from switching to a different insurance provider; and (3) it can increase a company's brand value. This first two effects occur immediately, the third effect has long-term character. In this counterfactual, we will concentrate on the immediate two effects, i.e. on consumer search behavior and purchase behavior given consideration. We do so by conducting four counterfactuals. In each of them, we pick one company and maximize the satisfaction of existing customers with this company (holding everything else constant). The four companies we study are Allstate, Geico, Safeco, and Travelers. We picked them as they differ on two dimensions: company size and the level of the average premium.

Results are shown in Table 28. All four companies benefit from an increase in customer satisfaction by raising their consideration and purchase market shares as well as retention rates. Comparing the companies on the two dimensions company size and level of average premium, the smaller companies (Safeco and Travelers) benefit more than larger companies (Allstate and Geico) and the less expensive companies (Geico and Safeco) benefit more than the more expensive companies (Allstate and Travelers). On average, insurance companies are able to increase their share among the considered brands by about 5%, their share among the purchased brands by about 15%, and their retention rates by about 19%.

To put a dollar value on the benefits of increasing customer satisfaction, we calculate the additional profits associated with the higher market shares and rescale them to the actual size of the U.S. auto insurance industry of about $160 billion. As shown in Table 19 and discussed in the Results section, gross profit margins lie between 20% and 30% in the U.S. auto insurance industry. We deduct 5% from these gross profit margins to account for insurance companies' administrative expenses and calculate the increase in profits assuming net profit margins of 15% and 25%. Results are shown in Table 29. All four insurance companies can increase their profits between $124 and $658 million (assuming a profit margin of 15%) and $194 million and $1,108 million (assuming a profit margin of 25%) from maximizing their customers' satisfaction.

9.3 Pricing

While insurance companies can influence consumer search and switching costs only to a very limited extent, in this counterfactual, we explore how insurance companies should set prices for new and existing customers in the presence of current levels of search and switching costs. Here, we look at the effects of a company offering a 10% discount to new customers holding everything else constant and at the effects of a company increasing premia by 10% to new customers holding everything else constant. We do the same in the case of a discount and premium increase for existing customers. The two companies that we study in this counterfactual are Allstate and Geico. The results are shown in Table 30. Not surprisingly, we find that offering discount to either new or existing customers increases a company's marketing share and increasing the level of an insurance premium to either new or existing customers decreases a company's market share. Note that the more expensive company Allstate is able to increase its market share more than the less expensive company Geico by offering discounts to new or existing customers. On the other hand, Geico looses less of its market share when it increases the prices for either new or existing customers than Allstate.

As mentioned in Section 3, the U.S. auto insurance industry is highly regulated. One of the regulations is that insurance companies have to offer premia to a customer based on the customer's belonging to a risk group. This also gives the intuition why discounts to new customers are not profitable: They have to be offered to all new customers which drives up the costs of these discounts to a very high level and is higher than the additional profits coming from winning over additional new customers. On the other hand, increasing insurance premia to existing customers is profitable to insurance companies. The profit decrease due to loosing customers is smaller than the profit increase due to higher premia for the remaining existing customers.

Table 30 showed the effects of the change in premia only. In reality, changes in prices are often accompanied by advertising campaigns or change the content of existing advertising campaigns. Discounts or price increases might also change the distributions of prices in the market and customer satisfaction with prices offered by their insurance provider. While capturing the joint effect of advertising and a change in prices is difficult within the framework of our model, we are currently working on incorporating the effect of a change in premia on the distribution of insurance prices on consumer search and purchase behavior in this counterfactual in the next version of the paper.

10 Robustness Checks

10.1 Pricing Regressions

We check whether the personal characteristics we have can explain the variation in insurance policy prices. To do so, we run two hedonic pricing regressions where we regress prices on observed characteristics. Similar to the model estimation, we only use prices charged by the previous insurer in the first pricing regression and all quoted prices in the second pricing regression. Results are shown in Table 20. The two middle columns show the estimates for the PI Pricing Regression, the two columns at the right the estimates for All Quotes Pricing Regression. Note that all coefficients have expected signs and reasonable magnitudes. The coefficient estimates across these two pricing regressions are also quite similar in magnitude and lie within two standard errors of each other suggesting some internal consistency. The R^2 for the PI and All Quotes Pricing Regressions are .69 and .57, respectively, and support our view that we have information on most of the consumer characteristics that determine insurance premiums. To show that these "low" R^2 are due to the fact we combine prices across all companies, we run company-specific pricing regressions. Due to data limitations, we are only able to do so for the largest four auto insurance companies (Allstate, Geico, Progressive, and State Farm). The R^2 of these separate pricing regressions (using all quoted prices) are .82, .80, .83, and .87, respectively. These regressions show that although the characteristics do explain a substantial portion of variation in prices, there are nevertheless factors left out that can justify the assumption of price search as noted previously.

10.2 Assumptions During Data Reconstruction Process

During the process of recontructing the competitive price quotes[33], we needed to make some data assumptions (Table 2) in order to be able to recover the premia. A valid question is how these assumptions influence our results. While it is prohibitively expensive to check the consequences of our data assumptions in the price recovery process by reconstructing competitive prices using a different set of assumptions, we try to mitigate those concerns by estimating our model using consumers who purchased auto insurance for one driver and one car only and did not have home insurance. The reasoning behind using this subset of consumers for our robustness check is that assumptions on additional drivers and cars as well as details on other insurance products should have the strongest influence on premia. By focusing on the subset of who purchased auto insurance for one driver and one car only and did not have home insurance, we avoid the effects of those assumptions of our estimates. The results are shown in Table 21. The estimates are very similar to those for Model 1 for the whole data set (see Table 14). The largest differences can be found in the search and price parameter estimates. Consumers seem to be more sensitive to prices and have a larger cost of search in utils. Search and the value on consumer inertia in dollars for this subset of consumers are \$52.46 (per search) and \$245.02, respectively, compared to \$48.29 and \$305.50, respectively, for the whole data set. We believe that the higher price sensitivity and lower inertia are due to the subsample consisting of younger, mostly single people or couples without children who make less money (e.g. average age 38.75 years in the subsample versus 45.23 years in the whole data set, 1.84 versus 2.33 people in the household). Despite the small differences in estimates, our results seem to be robust to the data reconstruction assumptions.

[33]For the details of this process, please see Section 4.

10.3 Decreasing Cost of Search

In our model, we assume that consumers incur the same cost when collecting a price quote independently of whether it is the first quote or the tenth quote the consumer is collecting. An alternative model specification is to assume that consumers incur decreasing cost of search, i.e. collecting the first quotes is more costly to consumers than later quotes. We estimate this alternative model by dividing all consumers into four groups (those who search zero to two times, three or four times, five or six times, and seven or more times) and estimating search costs for each of these four groups separately. This specification also allows for decreasing cost of search in a "non-parametric" way. The results are shown in Table 25. We find the cost of one search for consumers who search zero to two times to be $65.31, for consumers who search three or four times $51.87, for consumers who search five or six times $41.62, and $34.66 for consumers who search seven or more times. The loglikelihood for this model is -3,090.73. We conduct a likelihood ratio test and compare the likelihood of the model allowing for decreasing cost of search with the likelihood of Model 1 in Table 14. We find the model with decreasing cost of search to significantly better fit the data. We intend to make the model with decreasing search costs as our focal model in future versions of the paper.

10.4 Consumer's Knowledge of the Price Distributions

One of the assumptions that we make in our model is that consumers know the distributions of prices present in the market with the parameters of this distribution coming from the observed price distributions in the data. To examine the sensitivity of our results to this assumption, in particular to assess the consequences of consumers having less information, we increase the estimated variance of the price distributions two- and four-fold to reflect consumers knowing less about the prices in the market. The results are shown in Table 24. We find the cost of one search to be $105.04 when the variance is doubled and $213.55 when the variance is quadrupled. Recall that we found cost of one search of $48.29 for the same model specification under the assumption that consumers know the distributions of prices present in the market. This means that search costs are increasing when consumers are less knowledgeable about the distributions of prices. The loglikelihoods of the model specifications with doubled and quadrupled variances are -3,314.60 and -3,344.24, respectively. The loglikelihood of the same model where we assume consumers know the distribution of prices present in the market and estimate its variance from data on premia (Model 1, Table 14) is -3,295.35. Thus our assumption appears to be more consistent with the data as compared to the alternative assumption of less information.

10.5 Advertising

In our model, we control for a consumer's exposure to an insurance company's advertising by including an interaction effect between an insurance company's monthly advertising spending (during the month prior to the consumer's auto insurance purchase) and company-specific advertising recall measures reported by individual consumers. Here, we explore how sensitive our results are to different specifications of advertising in our model. We estimate three different models: In the first model specification, we only control for monthly advertising spending during the month prior to the consumer's auto insurance purchase in the utility function; in the second model specification, we only control for self-reported advertising recall by the consumer for each company; in the third model specification, we control for both monthly advertising spending and advertising recall; and in the fourth model specification, we control for both the main effect of advertising spending and advertising recall and the interaction effect between both variables. The estimates are shown in Table 26. In the first model specification, we find monthly advertising spending to be insignificant; in the second model specification, we find a coefficient for advertising recall of .77[34]; in the third model specification, we find coefficients for advertising recall and advertising spending of .50 and -.40, respectively, and in the fourth model specification, we find positive coefficients for both advertising spending (.06) and advertising recall (1.00) and a negative interaction effect between those two variables (-.13). Also note that our search cost estimates are robust across these three model specifications ($43.99, $44.56, $44.39, $44.96 compared to $48.29 from Model 1, Table 14). Similarly, the total effect of consumer inertia remains (stemming from both switching costs and customer satisfaction) constant across all three model specifications and similar to the one found in Model 1 (Table 14). The loglikelihoods of these three model specifications are -3,300.06, -3.295.60, -3,293.52, and -3.291.89, respectively. While the loglikelihood expectedly increases with the inclusion of more variables in the estimation, AIC and BIC paint a mixed picture. AIC supports the inclusion of more variables, while BIC which punishes the introduction of additional variables more severely does not support model specifications with more variables. Most importantly, we note that our results are robust to all alternative model specifications with regard to advertising.

A potential concern in the way that we currently incorporate advertising in our model is one of endogeneity. Insurance companies might target specific groups of consumers with their advertising from which they know that these consumers are likely to respond to their ads. We try to control for exposure to advertising by using consumer-specific advertising recall measures in our model. We acknowledge that this might not fully solve the advertising endogeneity problem, but point out that our goal to is mainly to control for advertising. We do not interpret the advertising coefficient structurally and are not conducting counterfactual analyses

[34]Recall that the interaction effect between advertising spending and advertising recall has a coefficient of .15 in Model 1, Table 14.

where we study how changes in advertising affect consumer behavior.

10.6 Sequential Search

As mentioned in the Section 5 on the model, the search method the consumer is using is not identified because in our data we observe the outcome of the search process, but not the sequence of searches. Either assumption on the search method, i.e. simultaneous or sequential search, can be justified by factors outside our model specification. In our model, we assumed consumers search simultaneously to form their consideration sets. In this section, we want to test for the robustness of our results under a sequential search assumption. We currently do not have estimation results for this model, but explain our estimation approach in detail in Technical Appendix C.

10.7 Adverse Selection

Previous literature has found mixed evidence on the existence of adverse selection and informational asymmetries in general in the auto insurance industry[35]. We tested for adverse selection in our data using the approaches suggested by Puelz and Snow (1994) and Dionne et al. (2001) and have no found evidence of adverse selection.

11 Limitations & Future Research

There are several limitations to our research. First, our model describes the consumer search process and the purchase decision given a choice of a policy with certain characteristics (coverage choices, payment terms), i.e. we do not jointly model policy choice and purchase. Our model assumes that consumers first make the decision what policy to buy and then start the shopping process. It is left for future research to develop a model where consumers choose policy characteristics and search at the same time evaluating different policies. Second, we assume that consumers have rational expectations about prices for all companies in the market. This points to two assumptions that we are making: Consumers are aware of all the companies operating in their markets and consumers have rational expectations about prices. Clearly, a model that uses information on all three stages of the purchase process (awareness, consideration, choice) would provide further details into the first two stages. A model has information on consumer price expectations or is able to recover them would enable researchers to test the hypothesis of rational price expectations.

[35]Among papers that found evidence for informational asymmetries in the auto insurace industry are Puelz and Snow (1994) and Israel (2007). Among papers that did not find evidence for informational asymmetries are Dionne et al. (2001), Cummins and Tennyson (1996), Abbring et al. (2003), and Chiappori and Salanié (2000).

Third, our data does not allow us to account for unobserved heterogeneity. While using information on quoting channels, demographics and psychographic factors allows us to account for observed heterogeneity in preferences, search and switching costs, it is possible that there are residual sources of heterogeneity that we have not accounted for. Fourth, some data limitations and institutional characteristics of the insurance industry make a direct comparison with models and conclusions from previous literature difficult. Not having price information for all companies in a consumer's market or knowing consumer price expectations together with the large number of companies that operate in each market make a meaningful comparison with previous literature that has estimated consideration sets impossible[36]. Another stream of literature on search (Hong and Shum (2006), Moraga-Gonzalez and Wildenbeest (2008)) assumes consumers search companies randomly, i.e. every company in the market has the same chance of being picked for a price quote by the consumer. A direct comparison of their results with ours is again not possible as we use a utility-based approach that allows for differences in brand preferences, advertising, and switching costs, while - in addition to random search - previous literature did not incorporate these characteristics and assumed consumers only care about prices. A comparison of our model and results with previous literature on search and consideration is left for future research in an industry with different characteristics.

Fifth, another area that future research could focus on is disentangling the part of search cost that is due to time cost from the part that is due to psychological cost. The results would help researchers understand why search costs have not disappeared with the introduction of the internet. Sixth, we found company type and regional differences in preferences similarly to Bronnenberg et al. (2009). Social interaction, the regulatory environment favoring one company over another, regional pricing, agent density, and historical brand preferences are all possible explanations. The contribution of each factor to the observed preference pattern is left for future research. And finally, we observed insurance companies to follow four distinctly different market entrance strategies: being present in (nearly) all markets, in the biggest markets, in geographically connected markets or concentrating on one market. Gaining insights into the reasoning behind these strategies may be an interesting area to explore.

12 Conclusion

The U.S. auto insurance industry presents ample potential for research in marketing. We contribute to this literature by providing an explanation for the high observed retention rate and the high savings advertised by insurance providers when switching insurers in this industry. We show that the levels of search and switching costs that we find in our data can explain this retention rate. From a methodological perspective,

[36]There are 11 - 17 companies operating in each market in our data. Previous literature that has estimated consumer consideration sets used markets with a much smaller number of brands, e.g. Mehta et al. (2003) use a market with four brands.

our paper contributes to the marketing literature by jointly modeling search, consideration, and switching. We provide a search cost estimation approach that is feasible in markets with a large number of alternatives where one observes the consideration set and the choice decision.

We find average switching costs of about \$85, cost of one price search over the internet of about \$45 and cost of a price search through local agents, mail, and calling centers of \$83, \$110, and \$70, respectively. When studying how search and switching costs, customer satisfaction, and consumer preferences other than those associated with inertia contribute to the level of retention found in the auto insurance industry, we find consumer search costs and customer satisfaction to be the two main drivers. These search and switching costs together with customer satisfaction provide an explanation for the very high retention rate observed in the auto insurance industry and appear to support the high potential savings advertised by the industry. To support managers in their decision making about investments in measures to increase customer satisfaction, we calculate additional profits resulting from a maximization of customer satisfaction. We find maximizing customer satisfaction to yield additional profits in the magnitude of several hundred million dollars. And finally, we study whether insurance companies should increase or decrease insurance premia to new and existing customers. We find insurance companies to lose money by offering discounts to new customers and increase their profits by increasing premia to existing customers. Our results are consistent with the "invest-harvest" theme found in the switching cost literature.

References

[1] Abbring J. H., P.-A. Chiappori, J. Pinquet (2003), "Moral Hazard and Dynamic Insurance Data", Journal of the European Economic Association, 1(4), 767 - 820.

[2] Allenby, G. M., and J. L. Ginter (1995), "The Effects of In-Store Displays and Feature Advertising on Consideration Sets", International Journal of Research in Marketing, 12(1), 67 - 80.

[3] Baye, M.R., J. Morgan, and P. Scholten (2007), "Information, Search, and Price Dispersion", Handbook of Economics and Information Systems, Vol. I.

[4] Ben-Akiva, M.E., and S.R. Lerman (1985), "Discrete Choice Analysis: Theory and Application to Travel Demand", MIT Press.

[5] Berndt, E. R., R. S. Pindyck, and P. Azoulay (2003), "Consumption Externalities and Diffusion in Pharmaceutical Markets: Antiulcer Drugs.", Journal of Industrial Economics, 51(2), 243 - 270.

[6] Berger, L. A., P. R. Kleindorfer, H. Kunreuther (1989), "A Dynamic Model of the Transmission of Price Information in Auto Insurance Markets", The Journal of Risk and Insurance, 56, 17 - 33.

[7] Bronnenberg, B.J., S.K. Dhar, and J.-P. Dubé (2009), "Brand History, Geography, and the Persistence of Brand Shares", Journal of Political Economy, 117(1), 87 - 115.

[8] Brynjolfsson, E., A. Dick, and M. Smith (2004), "Search and Product Differentiation at an Internet Shopbot", Working Paper.

[9] Carlson, J.A., and R.P. McAfee (1983), "Discrete Equilibrium Price Dispersion", Journal of Political Economy, 91(3), 480 - 493.

[10] Chade, H., and L. Smith (2006), "Simultaneous Search", Econometrica, 74(5), 1293 - 1307.

[11] Chade, H., and L. Smith (2005), "Simultaneous Search", Working Paper.

[12] Chiang, J., S. Chib, and C. Narasimhan (1999), "Markov Chain Monte Carlo and Models of Consideration Set and Parameter Heterogeneity", Journal of Econometrics, 89, 223 - 248.

[13] Chiappori, P.-A., and B. Salanié (2000), "Testing for Asymmetric Information in Insurance Markets", Journal of Political Economy, 108(1), 56 - 78.

[14] Chiappori, P.-A., B. Jullien, B. Salanié, and F. Salanié (2006), "Asymmetric Information in Insurance: General Testable Implications", The Rand Journal of Economics, 37(4), 783 - 798.

[15] Cummins, J. D., D. M. McGill, H. E. Winklevoss, and H. A. Zelten (1983), “Consumer Attitudes Toward Auto and Homeowner Insurance”, Philadelphia, Department of Insurance, Wharton School, University of Pennsylvania.

[16] Cummins, J. D., and S. Tennyson (1996), “Moral Hazard in Insurance Claiming: Evidence from Automobile Insurance”, Journal of Risk and Uncertainty, 12, 29 - 50.

[17] Dahlby, B., and D.S. West (1986), “Price Dispersion in an Automobile Insurance Market”, Journal of Political Economy, 94(2), 418 - 438.

[18] De los Santos, B., A. Hortacsu, and M. R. Wildenbeest (2010), “Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior”, Working Paper.

[19] Dionne, G., Ch. Gouriéroux, and Ch. Vanasse (2001), “Testing for Evidence of Adverse Selection in the Automobile Insurance Market: A Comment”, Journal of Political Economy, 109(2), 444 - 453.

[20] Draganska, M., and D. Klapper (2009), “Choice Set Heterogeneity and the Role of Advertising: An Analysis with Micro and Macro Data”, Working Paper.

[21] Dubé, J.-P., G. Hitsch, and P. E. Rossi (2009), “Do Switching Costs Make Markets less Competitive?”, Journal of Marketing Research, 46(4), 435 - 445.

[22] Dubé, J.-P., G. Hitsch, and P. E. Rossi (2010), “State Dependence and Alternative Explanations for Consumer Inertia”, RAND Journal of Economics, 41(3), 417 - 445.

[23] McFadden, D. (1986), “The Choice Theory Approach to Market Research”, Marketing Science, 5(4), 275 - 297.

[24] Gauri, D., K. Sudhir, and D. Talukdar (2008), “The Temporal and Spatial Dimensions of Price Search: Insights from Matching Household Survey and Purchase Data”, Journal of Marketing Research, 45(2), 226 - 240.

[25] Gumbel, E. J. (1961), “Bivariate Logistic Distributions”, Journal of the American Statistical Association, 56(294), 335 - 349.

[26] Handel, B. (2010), “Adverse Selection and Switching Costs in Health Insurance Markets: When Nudging Hurts”, Working Paper.

[27] Hong, H., and M. Shum (2006), “Using Price Distributions to Estimate Search Cost”, RAND Journal of Economics, 37(2), 257 - 275.

[28] Israel, M. (2005a), "Tenure Dependence in Consumer Firm Relationships: An Empirical Analysis of Consumer Departures from Automobile Insurance Firms", The Rand Journal of Economics, 36(1), 165 - 192.

[29] Israel, M. (2005b), "Services as Experience Goods: An Empirical Examination of Consumer Learning in Automobile Insurance", The American Economic Review, 95(5), 1444 - 1463.

[30] Israel, M. (2006a), "Where is All the Hidden Information Hiding? Evidence from Automobile Insurance Panel Data", Working Paper.

[31] Israel, M. (2006b), "Who Can See the Future? Information and Consumer Reactions to Future Price Discounts", Working Paper.

[32] Israel, M. (2007), "Do We Drive More Safely When Accidents are More Expensive? Identifying Moral Hazard from Experience Rating Schemes", Working Paper.

[33] Kim, J., B. Bronnenberg, and P. Albuquerque (2009), "Consumer Search and Online Demand for Durable Goods", Working Paper.

[34] Klemperer, P. (1995), "Competition when Consumers have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade", The Review of Economic Studies, 62, 515 - 539.

[35] Mehta, N., R. Surendra, and K. Srinivasan (2003), "Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation", Marketing Science, 22(1), 58 - 84.

[36] Moraga-Gonzalez, J.L., and M.R. Wildenbeest (2008), "Maximum Likelihood Estimation of Search Cost", European Economics Review, 52(5), 820 - 848.

[37] Morgan, P. and R. Manning (1985), "Optimal Search", Econometrica, 53(4), 923 - 944.

[38] Puelz, R. and A. Snow (1994), "Evidence on Adverse Selection: Equilibrium Signaling and Cross-Subsidization in the Insurance Market", Journal of Political Economy, 102(2), 236 - 257.

[39] Ratchford, B., M. Lee, and D. Talukdar (2003), "The Impact of the Internet on Information Search for Automobiles", Journal of Marketing Research, 40(2), 193 - 209.

[40] Ratchford, B., and N. Srinivasan (1993), "An Empirical Investigation of Returns to Search", Marketing Science, 12(1), 73 - 87.

[41] Schlesinger, H., and J.-M. Graf von der Schulenburg (1991), "Search Costs, Switching Costs and Product Heterogeneity in an Insurance Market", Journal of Risk & Insurance, 58(1), 109 - 119.

[42] Siddarth, S., R.E. Bucklin, and D.G. Morrison (1995), "Making the Cut: Modeling and Analyzing Choice Set Restriction in Scanner Panel Data", Journal of Marketing Research, 32(8), 255 - 266.

[43] Puelz, R., and A. Snow (1994), "Evidence on Adverse Selection: Equilibrium Signaling and Cross-Subsidization in the Insurance Market", Journal of Political Economy, 102(2), 236 - 257.

[44] Terui, N., M. Ban, and G.M. Allenby (2009), "The Effect of Media Advertising on Brand Consideration and Choice", Working Paper.

[45] Van Nierop, E., B. Bronnenberg, R. Paap, M. Wedel, and P. H. Franses (2010),"Retrieving Unobserved Consideration Sets from Household Panel Data", Journal of Marketing Research, 47(1), 63 - 74.

[46] Vishwanath, T. (1992), "Parallel Search for the Best Alternative", Economic Theory, 2, 495 - 507.

[47] Weitzman, M. L. (1979), "Optimal Search for the Best Alternative", Econometrica, 47(3), 641 - 654.

[48] Wilson, Ch. (2009), "Market Frictions: A Unified Model of Search and Switching Costs", Working Paper.

TECHNICAL APPENDIX A:

Derivation of Distribution of Maximum Utility

Distribution of prices p_j:

$p_j \sim EV\left(\eta_j, \mu\right)$ $\qquad E\left[p_j\right] = \eta_j + \frac{e_c}{\mu}$

Distribution of $\beta_p p_j$:

$\beta_p p_j \sim EV\left(\beta_p \eta_j, \frac{\mu}{\beta_p}\right)$ $\qquad E\left[\beta_p p_j\right] = \beta_p \eta_j + \frac{e_c \beta_p}{\mu}$

Distribution of utilities u_j:

$u_{ij} \sim EV\left(\alpha_j + X_{ij}\beta + \beta_{SC}PI_{ij} + \epsilon_{ij} + \beta_p \eta_j, \frac{\mu}{\beta_p}\right)$ $\qquad E\left[u_{ij}\right] = \alpha_j + X_{ij}\beta + \beta_{SC}PI_{ij} + \epsilon_{ij} + \beta_p \eta_j + \frac{e_c \beta_p}{\mu}$

Let us define $a_{ij} = \alpha_j + X_{ij}\beta + \beta_{SC}PI_{ij} + \epsilon_{ij} + \beta_p \eta_j$ and $b = \frac{\mu}{\beta_p}$. Then the distribution of the maximum utility is given by

$$\max_{j \in S} u_{ij} \ \sim \ EV\left(\frac{1}{b} \ln \sum_{j \in S} \exp\left(b a_{ij}\right), b\right)$$

TECHNICAL APPENDIX B: Estimation of Consideration Set and Purchase Probabilities

First, we define the optimality conditions for observed consumer search behavior:

1. Conditions for at least one search being optimal (i.e. PI+1):

   (a) $\min_{j \in S} (E[u_{ij}]) \geq \max_{j' \notin S} (E[u_{ij'}])$

   (b) $\Gamma_{i,k+1} \geq \max(\Gamma_{i,k'+1}) \quad \forall k \neq k'$

2. Condition for no search being optimal (i.e. PI only):

   (a) $\Gamma_{i,1} \geq \max(\Gamma_{i,k'+1}) \quad \forall k' \neq 0$

Then we take a large number of draws of ϵ_j (for each consumer/ company combination). For each set of draws of ϵ_j, we calculate the relevant differences. For example, for the case 1a, we calculate

$x_1 = \min_{j \in S} (E[u_{ij}]) - \max_{j' \notin S} (E[u_{ij'}])$

and $x_2 = \Gamma_{i,k+1} - \max(\Gamma_{i,k'+1})$

and then smooth the probabilities using a scaled multivariate logistic CDF (Gumbel (1961)). For the case 1a, the bivariate logistic CDF takes the form

$$F(x_1, x_2; s_1, s_2) = \frac{1}{1 + \exp(-s_1 x_1) + \exp(-s_2 x_2)} \tag{16}$$

where s_1 and s_2 are tuning parameters.

Second, we define the optimality condition for observed consumer purchase behavior:

$u_{ij} \geq u_{i,j'} \qquad \forall j \neq j' \quad j, j' \in C_i$

Using the same approach as for the consideration set probability, we calculate

$x_q = u_{ij} - u_{i,j'} \qquad \forall j' \in C_i$

and use the same scaled multivariate logistic CDF as for the consideration set probabilities where the number of dimensions of the multivariate logistic CDF is one less than the size of the consideration set.

We test the performance of our estimation approach by simulating data on search set decisions for 3,000 consumers. Consumers can choose among four brands with the first search being free. We simulate both the consideration and purchase decisions. We make 300 draws from the distribution of the error term and replicate the simulation 25 times. The true parameters and our estimation results are shown in Table 13 . The true values are recovered quite well.

TECHNICAL APPENDIX C: Sequential Search Model

Utility is the utility a consumer receives from a product. *Reservation* utility is the utility that makes a consumer indifferent between searching and not searching. Weitzman (1979) developed three rules which guide a consumer's optimal behavior under sequential search: (1) Sequence rule: The consumer ranks all alternatives according to their reservation utilities and searches first the top ranked one, then the second-ranked one, etc. (2) Stopping rule: The consumer stops searching when the maximum utility among the searched companies is larger than the maximum reservation utility among the non-searched companies. (3) Choice rule: The consumer picks the company which gives him the largest utility among the ones he has searched. In the sequential search model, after searching a company and receiving a price quote, the consumer makes the decision whether to continue searching or to stop searching and purchase from the set of searched companies. In contrast to the simultaneous search model, there are thus no separate consideration and purchase stages.

We will now describe how a sequential search model can be estimated in settings where the outcome of the search process is observed using a specific example. Suppose there are six brands in a market A, B, C, D, E, F. Utilities for these brands are denoted by U_A, U_B, U_C, U_D, U_E, U_F which contains a systematic part of utility and a linearly added EV Type I distributed error term. Reservation utilities are denoted by R_B, R_C, R_D, R_E, R_F. Kim et al. (2009) have shown that they also can be written as an observed (approximated) systematic part and a linearly added error term which we assume follows an EV Type I distribution. Suppose A is the previous insurer whose price quote the consumer knows (free first search). It is also observed in the data that the consumer searched B, C, and D and purchased B. Note that the alphabetical naming of the different brands does NOT imply and ranking of utilities or reservation utilities.

Using the Weitzman (1979) results, we know that

- $\min(R_B, R_C, R_D) > \max(R_E, R_F)$ because the consumer also searched B, C, and D, but not E and F.
- $\max(U_A, U_B, U_C, U_D) > \max(R_E, R_F)$ because the consumer stopped searching after the 4th search.
- $U_B > U_A, U_C, U_D$ because the consumer purchased B.

Additionally, because we observe the consumer searching four times (and not stopping the search earlier), the following must hold:

- $U_A < \max(R_B, R_C, R_D)$
- If $R_B > R_C, R_D$, then $U_B < R_C, R_D$
- If $R_C > R_B > R_D$, then $\max(U_A, U_C) < R_B$ and $U_B < R_D$

- If $R_D > R_B > R_C$, then $\max(U_A, U_D) < R_B$ and $U_B < R_C$
- If $R_C > R_D > R_B$, then $\max(U_A, U_C) < R_D$ and $\max(U_A, U_C, U_D) < R_B$
- If $R_D > R_C > R_B$, then $\max(U_A, U_D) < R_C$ and $\max(U_A, U_C, U_D) < R_B$

These conditions only make a few orderings of utilities and reservation utilities consistent with observed consumer behavior:

1. $R_B > R_C, R_D > U_B > U_A, U_C, U_D, R_E, R_F$
2. $R_C > R_B > R_D > U_B > U_A, U_C, U_D, R_E, R_F$
3. $R_D > R_B > R_C > U_B > U_A, U_C, U_D, R_E, R_F$
4. $R_C, R_D, U_B > R_B > U_A, U_C, U_D, R_E, R_F$
5. $R_C, R_D > R_B > U_B > U_A, U_C, U_D, R_E, R_F$

Recall that both utilities and reservation utilities can be written as a systematic part plus an added error term which follows an EV Type I distribution.

APPENDIX

Observed Variables	
1.	Number of quotes
2.	Identity of companies the respondent got quotes from
3.	Current auto insurance company
4.	Previous auto insurance company
5.	Premium paid with current and previous auto insurance company
6.	Demographics: age, gender, marital status, income, education, number of people in the household, children under 18 in the household
7.	Location: Zip code, Community Type (e.g. rural or urban area, suburb)
8.	Number of drivers in the policy
9.	Driver under 25 years in the policy
10.	Number of vehicles in the policy
11.	Primary vehicle make
12.	Primary vehicle model
13.	Primary vehicle year
14.	Number of years with current insurance company
15.	Self-assessed credit history
16.	Period with no insurance during last 3 years
17.	Two or more tickets during last 3 years
18.	Two or more accidents during last 3 years
19.	Home owner insurance with current auto insurance company
20.	Home owner insurance with other insurance company
21.	Relocated during the past 12 months and, if yes, state previously lived in
22.	Number of hours spent online per week
23.	Quoting channel (e.g. agent, insurer website)
24.	Psychographic questions
25.	Advertising recall

Table 1: Observed Variables

Data Assumptions

1.	All quotes are collected without SSN or VIN.
2.	All parties in the household work full-time, children under 18 go to school,children between 18 - 22 go to college, children 23 and older work full-time. Children who go to school or college qualify for the Good Student Discount.
3.	If there are two or more drivers under the policy: If the respondent is married or in a domestic partnership, the second driver is spouse or partner. Male spouses/ partners are 3 years older than female respondents, female spouses/ partners are 3 years younger than male respondents. If the respondent is 40 years or older and the second driver is under 25 years old, assume it is a daughter. The third and fourth drivers are assumed to be children. Their age is picked in accordance with the drivers under 25 years variable.
4.	All other characteristics for the 2nd, 3rd, and 4th driver are assumed to be similar to the respondent.
5.	If the respondent is over 60 years old, it is assumed that he/ she is retired. The he/ she uses the car for pleasure and drives 10,000 miles per year.
6.	Since we assume that all drivers work full-time, we assume that they commute to work (5 days a week) 10 miles. The annual mileage is 15,000. If there are more cars than drivers under the policy, we assume that the additional cars are used for pleasure and driver 5,000 annually.
7.	The respondent is the driver of the primary vehicle, the 2nd driver is the primary driver of the 2nd vehicle, etc. The 2nd vehicle is a 200 Honda Civic DX, the 3rd car a 1999 Ford Escort.
8.	All respondents got their driver's licenses at 16. No driver's license suspensions or revocations, no major violations. All respondents have continuously had insurance while having a car.
9.	All vehicles are owned. The respondent is the original owner if the car is younger than 5 years.
10.	Same coverage levels are picked for all cars under the policy.
11.	All members in the household have health insurance.
12.	No existing damage on cars.
13.	Insurance policy premiums are paid in full.
14.	To determine whether a respondent owns a home or not the two variables "Home owner insurance with current auto insurer" and "Home owner insurance with other insurer" were used. If either was indicated to be the case, the respondent owns a home. If it was required to differentiate between owning a house and owning a condo, a condo was picked for people under 30 years anda home for people over 30 years. If no home owner insurance was indicated, it was assumed that the respondent rents.
15.	If it is a required to pick an occupation, one was chosen that matches education and income levels. Occupations that often qualify for special discounts at some companies (e.g. nurse, physician, firefighter, policeman, teacher, etc.) were avoided.

Table 2: Data Assumptions

Consumer 005353
Male, 55 years old, married, lives in Illinois (urban area), two cars, three drivers, one driver under 25 years old previously insured with Safeco, requests two additional price quotes Safeco: \$997.29; MetLife: \$1,203.84; Travelers: \$1,221.76
Consumer 006963
Female, 54 years old, married, lives in California (rural area), three cars, three drivers previously insured with Geico, requests three additional price quotes Geico: \$537.86; 21st Century: \$658.24; Farmers: \$783.60; Progressive: \$832.48
Consumer 012672
Male, 26 years old, single, lives in New Jersey (Suburb of a medium city), one car, one driver, two or more tickets (excl. parking tickets) previously insured with Liberty Mutual, requests two additional price quotes Liberty Mutual: \$1,387.68; 21st Century: \$1,405.57; AIG: \$1,171.31

Table 3: Examples of Reconstructed Premia

Proportion in % of Consumers	**Final Data**	**Original Data**
Female	36.32	34.73
Period with no Auto Insurance during last 3 years	3.03	2.78
Relocated During the Past 12 Months	19.46	20.72
Presence of Children	25.79	26.66
Driver under 25 Years	8.25	10.74
Excellent Credit History	65.30	63.82
Good Credit History	22.83	23.57
Fair Credit History	8.46	7.99
Poor Credit History	1.89	2.81
Married	56.07	57.82
Divorced/ Separated	13.23	12.33
Widowed	.87	1.74
Single	22.48	21.44
Domestic Partnership	6.19	5.52
Income: below $50,000	20.71	19.91
Income: $50,000 - $99,999	39.84	40.89
Income: $100,000 - $149,999	24.65	18.76
Income: $150,000 or more	14.80	8.75
Education: High School or Less	7.24	5.58
Education: Some College or College Degree	54.72	56.70
Education: Some Graduate Courses or Advanced Degree	38.04	37.17
Community Type: Rural	14.67	14.99
Community Type: Suburb of a Medium City	29.63	31.24
Community Type: Suburb of a Large City	34.38	33.47
Community Type: Urban Area	19.66	19.13
Average Values	**Final Data**	**Original Data**
Number of Quotes	2.96	3.34
Number of Information Sources	2.12	2.20
Number of Methods to Obtain a Quote	2.01	2.09
Length of Price Gathering Process	1.54	1.65
Length of Choosing Process	1.36	1.52
Number of Vehicles	1.58	1.62
Number of Drivers	1.64	1.69
Vehicle Year	2001.98	2001.88
Premium with CI	592.97	612.60
Number of Hours Online	15.21	15.29
Age	45.23	45.09
Number of People in the Household	2.33	2.39

Table 4: Original and Final Data Set Comparison
PI: Previous Insurer; CI: Current Insurer

Variable	N	Mean	Std. Dev.	Minimum	Maximum
Number of Quotes	945	2.96	1.38	1	10
Number of Information Sources	945	2.12	1.28	1	8
Number of Methods to Obtain a Quote	945	2.01	0.90	1	6
Number of Online Quote Sites Visited	419	1.14	0.40	0	4
Length of Price Quote Gathering Process	879	1.54	1.27	0	12
Length of Choosing Process	854	1.36	1.27	0	11
Premium for 6-months Policy with PI	270	756.41	365.10	105.00	2,700.00
Premium for 6-months Policy with CI - Same Sample as PI	270	554.49	265.90	89.00	2,750.00
Premium for 6-months Policy with CI	945	592.97	288.28	74.00	2,750.00
Number of Vehicles with CI	945	1.58	0.64	1	3
Number of Drivers with CI	945	1.64	0.59	1	4
Number of Years with CI	945	7.07	9.04	0	50
Number of People in the Household	927	2.33	1.12	1	8
Number of Hours Online per Week	918	15.21	12.26	0	80
Vehicle Year	945	2001.98	4.19	1960	2007
Respondent Age	945	45.23	12.94	20	84
From Reconstructed Price Quotes					
Matched Premium for 6-months Policy with PI	945	606.58	280.07	108.40	1,947.50
Matched Premium for 6-months Policy with PI - Same Sample as above	270	626.19	266.11	173.50	1,778.00
Matched Premium for 6-months Policy with CI	945	585.17	227.78	118.89	2,705.85

Table 5: Descriptive Statistics
PI: Previous Insurer; CI: Current Insurer

Variable	No search/ No switch	Search/ No switch	Search/ Switch
Number of Respondents	56	586	303
Number of Quotes	1	2.89	3.51
Number of Information Sources	2.27	2.00	2.33
Number of Methods to Obtain a Quote	1.20	2.00	2.2
Number of Online Quote Sites Visited	1.21	1.13	1.15
Length of Price Quote Gathering Process	1.42	1.44	1.75
Length of Choosing Process	1.07	1.26	1.62
Premium for 6-months Policy with PI			756.41
Premium for 6-month Policy with CI	660.13	606.36	551.44
Number of Vehicles with CI	1.61	1.60	1.54
Number of Drivers with CI	1.73	1.65	1.60
Number of Years with CI	7.83	10.14	
Number of People in the Household	2.55	2.26	2.43
Number of Hours Online per Week	16.23	14.88	15.73
Vehicle Year	2002.93	2001.86	2002.05
Respondent Age	42.17	46.90	42.35

Table 6: Averages Across Customer Types
PI: Previous Insurer; CI: Current Insurer

Company	Considered by % Respondents	Accounting for Limited Availability*
Geico	53.91	53.91
Progressive	46.98	46.98
Allstate	36.60	36.60
State Farm	34.26	34.26
AIG	20.81	20.81
The Hartford	13.98	13.98
Nationwide	11.94	12.21
Farmers	11.86	11.86
Liberty Mutual	10.38	10.38
21st Century	7.91	7.91
Travelers	7.85	7.93
American Family	6.61	23.22
MetLife	6.37	2.20
GMAC	5.65	5.68
Mercury	5.23	7.68
Erie	5.00	11.77
Safeco	4.82	4.87

Table 7: Consideration Set Composition
*Probability that respondent considers an insurer if this insurer is offering policies in the respondent's state

Company	Market Shares in %	Retention Rates in %	Conversion Rates in %
21st Century	2.81	81.72	35.51
AIG	3.96	63.29	18.74
Allstate	14.16	69.84	38.70
American Family	2.53	91.03	38.25
Erie	2.83	96.25	56.66
Farmers	4.92	79.62	41.46
Geico	18.41	64.10	34.16
GMAC	1.60	67.12	28.34
The Hartford	4.24	87.07	30.35
Liberty Mutual	4.73	70.44	45.53
Mercury	2.22	81.07	42.46
MetLife	2.29	64.51	37.02
Nationwide	3.35	74.69	28.06
Progressive	12.42	62.67	26.56
Safeco	1.68	52.52	33.41
State Farm	13.18	72.40	38.48
Travelers	4.51	72.63	57.47

Table 8: Market Shares and Conversion Rate
Conversion Rate: Given that a company is being considered, what is the probability that it will be chosen?

	January	February	March	April	May	June	July	August	September	October	November	December	Mean
21st Century	2.93	2.09	2.36	1.66	4.63	5.68	2.52	2.07	4.03	2.58	1.08	0.25	2.66
AIG	4.03	4.34	3.70	4.99	5.33	4.04	4.66	4.94	3.85	4.28	2.19	2.65	4.08
Allstate	24.15	60.53	28.31	25.22	29.81	20.51	19.79	18.26	21.69	35.44	28.12	12.57	27.03
Amfam	4.83	3.07	2.09	1.83	1.75	1.13	1.63	1.37	3.58	3.77	3.30	2.01	2.53
Erie	0.01	0.00	0.00	0.00	0.01	0.01	0.02	0.01	0.00	0.01	0.00	0.00	0.01
Farmers	7.49	3.38	5.13	6.25	2.40	7.72	4.64	7.05	1.80	3.57	2.27	4.94	4.72
Geico	50.14	41.99	43.56	35.26	38.56	51.06	46.06	55.73	55.16	54.14	45.34	33.67	45.89
GMAC	0.00	0.10	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
Hartford	0.98	0.39	1.17	0.21	1.17	0.36	1.21	0.62	1.34	0.39	0.87	0.02	0.73
Liberty	1.38	0.91	0.76	0.78	0.06	13.13	9.56	8.06	14.41	9.43	8.17	2.03	5.72
Mercury	5.55	3.33	3.83	3.78	3.15	3.28	4.15	4.08	3.98	4.79	4.12	6.32	4.20
MetLife	0.43	0.43	0.07	0.00	0.54	1.57	0.95	0.72	1.40	1.23	0.60	0.25	0.68
Nationwide	3.86	2.58	5.76	4.95	8.60	5.30	5.35	6.42	7.57	6.11	9.73	8.82	6.25
Progressive	30.53	27.61	29.20	31.73	30.97	30.62	29.17	30.98	28.74	32.35	30.05	22.72	29.56
Safeco	0.00	0.00	0.00	0.01	0.01	0.01	0.06	0.11	0.01	0.00	0.07	0.11	0.03
State Farm	20.12	14.75	30.99	23.43	21.79	22.55	13.25	16.85	16.70	26.58	12.87	7.46	18.95
Travelers	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 9: Advertising Spending by Company in $1,000,000 in 2006

Company	% Respondents Recalling Advertising
21st Century	15.22
AIG	49.00
Allstate	78.66
American Family	20.16
Erie	3.45
Farmers	36.05
Geico	90.13
GMAC	20.66
Hartford	34.14
Liberty Mutual	38.11
Mercury	14.34
MetLife	44.36
Nationwide	55.53
Progressive	78.62
Safeco	13.41
State Farm	79.33
Travelers	27.48

Table 10: Advertising Recall by Company

	PI Prices		All Quotes Prices	
Location Parameter	Coefficient	Std. Error	Coefficient	Std. Error
Constant	4.6597a	(.6733)	4.4736a	(.3836)
Male	-.1631	(.1378)	-.1951c	(.0888)
Marital Status: Married	.2064	(.2701)	.0407	(.1627)
Marital Status: Divorced/ Separated	-.1003	(.2497)	-.2581	(.1589)
Marital Status: Widowed	.8726	(.6531)	1.1233b	(.3939)
Marital Status: Domestic Partnership	.3048	(.3155)	-.2168	(.2116)
Age	-.0088	(.0057)	-.0074c	(.0037)
Driver under 25	.5410d	(.3120)	.8347a	(.1926)
Two Vehicles	1.3434a	(.1731)	1.2804a	(.1086)
Three Vehicles	2.5610a	(.2762)	1.7428a	(.1799)
Two Drivers	-.0524	(.2304)	-.2170	(.1506)
Three Drivers	1.1211b	(.4065)	.7285b	(.2675)
Four Drivers	1.2561d	(.7335)	1.4547c	(.6805)
Location: Suburb of a Medium City	-.1907	(.2237)	-.3892b	(.1409)
Location: Suburb of a Large City	.1370	(.2290)	.1065	(.1404)
Location: Urban Area	.1398	(.2455)	.1307	(.1537)
Home Owner Insurance with CI	-.2615d	(.1437)	-.2158c	(.0963)
Other Insurance with CI	-.4528b	(.1543)	-.3180a	(.0985)
Two or More Accidents	1.9080a	(.4800)	1.9669a	(.3253)
Two or More Tickets	1.4560a	(.3356)	1.3206a	(.2242)
Model Age	-.0739a	(.0185)	-.0329b	(.0111)
Chosen coverage Dummies	yes		yes	
State	yes		yes	
Make*Class	yes		yes	
Scale Parameter				
Constant	1.5681a	(.0915)	1.8091a	(.1032)
Loglikelihood	-1,855.98		-6,263.03	

Table 11: Pricing Regressions
a: <.001, b: <.01, c < .05, d: <.10
Comment: Prices are measured in $100

	PI Price Residuals		All Quotes Price Residuals	
	Mean	Std. Dev.	Mean	Std. Dev.
21st Century	.0151	2.5034	-.0804	2.8971
AIG	.5510	1.8503	1.0306	3.6156
Allstate	.4517	2.0701	.8408	2.8539
American Family	-.9181	1.8783	-.7703	2.5609
Erie	-.6752	2.2029	-.6610	2.3801
Farmers	-.0896	1.983	.2885	3.6831
Geico	-.7473	1.766	-1.0244	2.2044
GMAC	1.4234	2.0993	2.7861	5.2494
The Hartford	-.3811	1.8675	-.0942	2.4076
Liberty Mutual	.4063	1.9079	.6257	2.8183
Mercury	.4727	2.7348	.0929	2.9579
MetLife	.5818	2.2262	.8196	2.8249
Nationwide	-.1635	1.8803	.3167	3.1631
Progressive	.0911	2.2659	.2204	3.0434
Safeco	.0355	1.3778	-.1110	1.9692
State Farm	.0709	1.8341	.3214	2.4835
Travelers	.5714	2.1509	.4330	2.4217

Table 12: Price Residual Means and Variances by Company

	True Values	Estimates	Std. Error
Brand Intercept 1	1	.8739	(.2226)
Brand Intercept 2	-.5	-.6272	(.2422)
Brand Intercept 3	1.5	1.2942	(.2004)
Brand Intercept 4	-.75	-.7662	(.2324)
X1	1.5	1.3906	(.1832)
X2	-1.5	-1.5618	(.1293)
Search Cost	.4	.4317	(.0533)
Loglikelihood		-9,023.75	

Table 13: Simulation Results

	Model 0		Model 1	
	Estimate	Std. Error	Estimate	Std. Error
	Brand Preferences		Brand Preferences	
21st Century	-2.3104a	(.1641)	-2.3147a	(.2005)
AIG	-1.8281a	(.1682)	-1.8283a	(.1732)
Allstate	-2.1001a	(.1682)	-2.0973a	(.1647)
American Family	-1.7098a	(.2495)	-1.7102a	(.2110)
Erie	-2.1258a	(.2544)	-2.1246a	(.2556)
Farmers	-2.8391a	(.1812)	-2.8371a	(.1897)
Geico	-2.9831a	(.1820)	-2.9735a	(.1909)
GMAC	-2.4646a	(.1947)	-2.4671a	(.2372)
Hartford	-1.7641a	(.1591)	-1.7679a	(.1683)
Liberty Mutual	-1.6877a	(.1556)	-1.6887a	(.1687)
Mercury	-2.3894a	(.2580)	-2.3933a	(.2328)
MetLife	-2.5922a	(.2454)	-2.5863a	(.2345)
Nationwide	-2.5151a	(.1824)	-2.5159a	(.1942)
Progressive	-2.1005a	(.1711)	-2.1016a	(.1749)
Safeco	-2.8041a	(.2213)	-2.8000a	(.2230)
State Farm	-2.3980a	(.1573)	-2.3956a	(.1666)
Travelers	-2.1498a	(.1890)	-2.1532a	(.2133)
	Other Parameters		Other Parameters	
Price in $100	-.4623a	(.0234)	-.4641a	(.0258)
Recall*Advertising in $10,000,000	.1579a	(.0242)	.1561a	(.0254)
PI			.5076a	(.0391)
Search Cost	-1.5008a	(.1578)	-1.4956a	(.1619)
Loglikelihood	-3,299.37		-3,295.35	
AIC	6,638.75		6,632.71	
BIC	6,749.64		6,749.14	

Table 14: Models 0 - 1
a: <.001, b: <.01, c < .05, d: <.10

	Model 2		Model 3	
	Estimate	Std. Error	Estimate	Std. Error
	Brand Preferences		Brand Preferences	
21st Century	-.1838	(.3900)	-.3277	()
AIG	.2818	(.3833)	.1324	()
Allstate	.0243	(.3784)	.2712	()
American Family	.3821	(.4105)	-.0965	()
Erie	-.0300	(.4411)	-.2640	()
Farmers	-.7548d	(.3931)	-.4126	()
Geico	-.9122c	(.3832)	-.7491	()
GMAC	-.2875	(.3888)	-.4189	()
Hartford	.3336	(.3733)	.3261	()
Liberty Mutual	.4279	(.3732)	.0764	()
Mercury	-.2383	(.4152)	-.3988	()
MetLife	-.4605	(.3991)	-.6943	()
Nationwide	-.4023	(.4059)	-.4480	()
Progressive	.0719	(.3764)	.1664	()
Safeco	-.6956d	(.3905)	-.8187	()
State Farm	-.3671	(.3743)	-.1694	()
Travelers	-.0241	(.3995)	.0480	()
	Other Parameters		Other Parameters	
Price in $100	-.4613a	(.0203)	-.5352	()
Recall*Advertising in $10,000,000	.1641a	(.0212)	.1427	()
Loglikelihood	-3,241.39		-2,856.08	
AIC	6,592.78		5,874.16	
BIC	6,897.71		6,323.25	

Table 15: Models 2 - 3
a: <.001, b: <.01, c < .05, d: <.10
Note: Std. errors for Model 3 are currently being calculated.

	Model 2		Model 3	
	Estimate	Std. Error	Estimate	Std. Error
	Search Costs		Search Costs	
Search Cost Constant	-1.5386a	(.1492)	-.8063	()
Mail			.2751	()
Insurer Website			-.5724	()
Online Quoting Service			-.7057	()
Insurer Calling Center			-.2042	()
	Switching Costs		Switching Costs	
PI	.4606a	(.0365)	-.7655	()
PI*Age			-.0133	()
PI*Male			-.3276	()
PI*Income $25,000 - $49,999			.0302	()
PI*Income $50,000 - $74,999			.1498	()
PI*Income $75,000 - $99,999			.2937	()
PI*Income $100,000 - $149,000			.3605	()
PI*Income $150,000 or more			.3043	()
PI*Education: High School Graduate			.6561	()
PI*Education: College Graduate			.6973	()
PI*Education: Some Graduate Courses			.6704	()
PI*Education: Advanced Degree			.7832	()
PI*Attitude Towards Auto Insurance Shopping & Switching			.0920	()
PI*New Technology Adoption			-.0497	()
PI*Technology Usage			.1215	()
PI*Loyalty			.1819	()
PI*Interest in Finance			-.4419	()
PI*Satisfaction with Billings and Payment Process			-.3317	()
PI*Satisfaction with Price			.3506	()
PI*Satisfaction with Interacting with Insurer			-.0452	()
PI*Satisfaction with Policy Offerings			-.3424	()
PI*Satisfaction with Claims Experience			.0714	()
PI*Overall Satisfaction			.5219	()

Table 16: Models 2 - 3 (continued)
a: $<.001$, b: $<.01$, c $< .05$, d: $<.10$
Note: Std. errors for Model 3 are currently being calculated.

	Model 2		Model 3	
	Estimate	Std. Error	Estimate	Std. Error
	Demographics		Demographics	
Credit History: Excellent	-2.6107b	(.9223)	-2.3987	()
Credit History: Good	-2.7904b	(.9407)	-2.5881	()
Credit History: Fair	-1.8697	(1.1973)	-1.6812	()
Age	-.0079	(.0131)	-.0103	()
Male	-.3762	(.3596)	-.3254	()
Marital Status: Married	.3016	(.4029)	.2998	()
Marital Status: Divorced/ Separated	.9668	(.6356)	.9656	()
Marital Status: Widowed	.0747	(1.5096)	.1117	()
Marital Status: Domestic Partnership	.1969	(.7618)	.2526	()
Community Type: Suburb of a Medium City	-.0597	(.5874)	-.0106	()
Community Type: Suburb of a Large City	-.8328	(.5458)	-.8238	()
Community Type: Urban Area	-.9390	(.5841)	-.9014	()
	Psychographics		Psychographics	
Attitude Towards Auto Insurance Shopping & Switching	-.4207c	(.1712)	-.4117	()
New Technology Adoption	-.3463d	(.1790)	-.3479	()
Technology Usage	-.1951	(.1832)	-.1646	()
Loyalty	-.0437	(.1794)	-.0802	()
Interest in Finance	.2818	(.2003)	.2881	()
	Quoting Channels		Quoting Channels	
Mail	1.5326b	(.5322)	1.6628	()
Insurer Website	1.9803a	(.4371)	1.5959	()
Online Quoting Service	1.8588a	(.3840)	1.3991	()
Insurer Calling Center	.4195	(.3264)	.2929	()
	Regions		Regions	
Northeast	-1.0647d	(.5551)	-1.0328	()
Mid-Atlantic	-.1650	(.5304)	-.1099	()
Midwest	-.4229	(.4702)	-.4080	()
Florida	-.1680	(.6033)	-.1883	()
South	1.0965	(1.0040)	.9813	()
Texas	.0323	(.5726)	.0088	()
Mountain Area	-.6070	(.7895)	-.5465	()
	Customer Satisfaction		Customer Satisfaction	
Satisfaction with Billings and Payment Process	.0687	(.1098)	.0695	()
Satisfaction with Price	-.1408	(.1048)	-.1284	()
Satisfaction with Interacting with Insurer	.1017	(.0841)	.0924	()
Satisfaction with Policy Offerings	.1054	(.1130)	.1123	()
Satisfaction with Claims Experience	.0458	(.0420)	.0434	()
Overall Satisfaction	-.1478	(.1650)	-.1485	()

Table 17: Models 2 - 3 (continued)
a: <.001, b: <.01, c < .05, d: <.10
Note: Std. errors for Model 3 are currently being calculated.

Factor 1: Attitude Towards Auto Insurance Shopping & Switching
Switching to another insurer is not worth the risk. Shopping for a new auto insurer is too difficult or time consuming. I have invested too much into building a relationship with my current agent or insurer to switch to a new auto insurer.
Factor 2: New Technology Adoption
I am always one of the first of my friends to try new products and services. I enjoy reading about new technology products. I am among the first of my friends and colleagues to try new technology products.
Factor 3: Technology Usage
Computers are too confusing to be of much use to me. Technology has litte impact on my daily life.
Factor 4: Loyalty
If a product or service is offered by a company I trust, I will buy it even if it is slightly more expensive. Compared to other people, I am more likely to be loyal to a brand. When I find a brand I like, I stick with it.
Factor 5: Interest in Finance
I find the ups and downs of the financial markets exciting. I regularly read financial news or financial publications.

Table 18: Psychographic Factors

	Own-price Elasticities	**Std. Errors**
21st Century	-2.11	(.13)
AIG	-2.05	(.18)
Allstate	-2.27	(.12)
American Family	-1.01	(.20)
Erie	-1.60	(.22)
Farmers	-1.91	(.15)
Geico	-1.69	(.11)
GMAC	-2.55	(.23)
Hartford	-2.11	(.16)
Liberty Mutual	-2.80	(.15)
Mercury	-2.77	(.19)
MetLife	-2.36	(.16)
Nationwide	-2.18	(.17)
Progressive	-2.31	(.12)
Safeco	-2.29	(.16)
State Farm	-2.62	(.13)
Travelers	-1.36	(.18)
Mean	-2.10	(.14)

Table 19: Implied Own-Price Elasticities

	PI Pricing Regression		All Quotes Pricing Regression	
Variable	Estimate	Std. Error	Estimate	Std. Error
Intercept	1.9773c	(.7424)	2.4690a	(.5830)
Male	-.1664	(.1668)	-.2064	(.1316)
Marital Status: Married	-.9297b	(.2785)	-1.0845a	(.2273)
Marital Status: Divorced/ Separated	-.1970	(.2690)	-.4406c	(.2182)
Marital Status: Widowed	.2683	(.7729)	-.0982	(.6670)
Marital Status: Domestic Partnership	-.0809	(.3417)	-.5018d	(.2719)
Age	-.0189c	(.0065)	-.0286a	(.0052)
Driver under 25	1.017c	(.4339)	.5822d	(.3429)
Male Driver under 25	.6802	(.5433)	1.1676c	(.4232)
Two Vehicles	2.0510a	(.2004)	2.3324a	(.1617)
Three Vehicles	4.1209a	(.3050)	4.6517a	(.2434)
Two Drivers	.3482	(.2749)	.4501c	(.2224)
Three Drivers	2.0190a	(.5067)	2.7491a	(.3916)
Four Drivers	2.5340c	(.9241)	3.3351c	(1.3224)
Location: Suburb of a Medium City	-.0498	(.2565)	-.1441	(.1988)
Location: Suburb of a Large City	.6078c	(.2550)	.4633c	(.1971)
Location: Urban Area	.8257c	(.2827)	1.0085a	(.2243)
Home Owner Insurance with CI	-.0709	(.1631)	-.1450	(.1300)
Other Insurance with CI	-.1989	(.1713)	-.5212a	(.1347)
Two or More Accidents	2.5943a	(.4766)	3.4573a	(.3630)
Two or More Tickets	1.4285a	(.3753)	1.9747a	(.2969)
Model Age	-.0560c	(.0185)	-.0642a	(.0156)
Chosen coverage Dummies	yes		yes	
State	yes		yes	
Make*Class	yes		yes	
R2	.69		.57	

Table 20: Pricing Regressions
a: <.001, b: <.01, c < .05, d: <.10

	Model 1	
	Estimate	Std. Error
	Brand Preferences	
21st Century	-1.4360a	(.3442)
AIG	-1.1929b	(.4126)
Allstate	-1.1370b	(.3860)
American Family	-1.7327a	(.4694)
Erie	-2.2909b	(.8228)
Farmers	-1.6834a	(.4066)
Geico	-2.1122a	(.5344)
GMAC	-2.5468a	(.6022)
Hartford	-2.0497a	(.4350)
Liberty Mutual	-1.6408a	(.4113)
Mercury	-1.6518a	(.3613)
MetLife	-2.1917a	(.3988)
Nationwide	-2.3200a	(.3887)
Progressive	-1.2496b	(.4115)
Safeco	-2.4102a	(.7307)
State Farm	-1.5612a	(.3947)
Travelers	-1.4703a	(.3976)
	Other Parameters	
Price in $100	-.6049a	(.1070)
Recall*Advertising in $10,000,000	.1660d	(.0848)
PI	.3935a	(.0939)
Search Cost	-1.1298a	(.1768)
Loglikelihood	-695.79	

Table 21: One Driver, One Car, No Home Insurance

Mark-up in %	**2007**
AIG	20.80
Allstate	33.00
American Family	19.50
Erie	28.80
Farmers	28.20
Geico	13.60
GMAC	28.70
The Hartford	28.90
Liberty Mutual	25.90
Mercury	30.60
MetLife	30.20
Nationwide	32.90
Travelers	31.10
State Farm	19.60

Table 22: Industry Mark-ups

	Consideration Set Composition		Purchase Market Shares		Retention Rates	
Company	**Data**	**Predicted**	**Data**	**Predicted**	**Data**	**Predicted**
21st Century	2.73	3.90	2.81	3.48	81.72	66.76
AIG	7.17	5.99	3.96	4.99	63.29	65.93
Allstate	12.61	10.82	14.16	12.38	69.84	67.56
American Family	2.28	2.32	2.53	3.03	91.03	88.67
Erie	1.72	2.66	2.83	3.55	96.25	86.48
Farmers	4.09	4.75	4.92	5.22	79.62	73.31
Geico	18.58	16.52	18.41	18.32	64.10	80.70
GMAC	1.95	2.55	1.60	1.38	67.12	50.25
The Hartford	4.82	6.03	4.24	5.14	87.07	75.63
Liberty Mutual	3.58	4.53	4.73	5.01	70.44	68.57
Mercury	1.80	2.22	2.22	2.47	81.07	67.50
MetLife	2.20	2.23	2.29	1.77	64.51	49.39
Nationwide	4.11	5.73	3.35	3.83	74.69	72.90
Progressive	16.19	13.69	12.42	13.10	62.67	76.68
Safeco	1.66	2.47	1.68	2.18	52.52	60.37
State Farm	11.81	9.44	13.18	10.58	72.40	73.75
Travelers	2.71	4.14	4.51	3.56	72.63	58.91

Table 23: Data and Predictions

	Doubled Variance		Quadrupled Variance	
	Estimate	Std. Error	Estimate	Std. Error
	Brand Preferences		Brand Preferences	
21st Century	-2.5803a	(.2281)	-2.4646a	(.1924)
AIG	-2.1652a	(.2126)	-1.9996a	(.1794)
Allstate	-2.2492a	(.2562)	-2.1517a	(.1398)
American Family	-2.2540a	(.3247)	-1.9284a	(.2223)
Erie	-2.4076a	(.2417)	-2.1031a	(.2290)
Farmers	-2.4886a	(.2116)	-2.3661a	(.1947)
Geico	-3.0952a	(.4234)	-2.6529a	(.3872)
GMAC	-2.6253a	(.2006)	-2.6518a	(.1976)
Hartford	-2.3029a	(.1977)	-2.0961a	(.1953)
Liberty Mutual	-2.3721a	(.2098)	-2.4106a	(.1887)
Mercury	-2.5992a	(.2332)	-2.5515a	(.2278)
MetLife	-2.5961a	(.2287)	-2.5356a	(.2065)
Nationwide	-2.7218a	(.2287)	-2.5479a	(.1845)
Progressive	-2.3484a	(.3280)	-2.1975a	(.3251)
Safeco	-3.0469a	(.2273)	-3.0069a	(.2406)
State Farm	-2.5267a	(.2492)	-2.4255a	(.1645)
Travelers	-2.3731a	(.1961)	-2.3464a	(.2141)
	Other Parameters		Other Parameters	
Price in $100	-.3857a	(.0319)	-.2069a	(.0220)
Recall*Advertising in $10,000,000	.1536a	(.0682)	.1435a	(.0745)
PI	.3371a	(.0554)	.3931a	(.0464)
Search Cost	-.9035a	(.1324)	-.8166a	(.1362)
Loglikelihood	-3,014.60		-3,044.24	
AIC	6,071.20		6,130.48	
BIC	6,187.63		6,246.91	

Table 24: Larger Variances
a: <.001, b: <.01, c < .05, d: <.10

	Model 1	
	Estimate	Std. Error
	Brand Preferences	
21st Century	-2.2225a	(.2138)
AIG	-1.7537a	(.1751)
Allstate	-1.8299a	(.1788)
American Family	-1.9842a	(.2293)
Erie	-1.9943a	(.2433)
Farmers	-2.1962a	(.1956)
Geico	-2.7032a	(.2323)
GMAC	-2.3138a	(.2243)
Hartford	-1.9688a	(.1724)
Liberty Mutual	-2.0099a	(.1846)
Mercury	-2.2992a	(.1955)
MetLife	-2.1320a	(.2362)
Nationwide	-2.3278a	(.1991)
Progressive	-1.9564a	(.1980)
Safeco	-2.7341a	(.2117)
State Farm	-2.0731a	(.2106)
Travelers	-1.9727a	(.2129)
	Other Parameters	
Price in $100	-.4773a	(.0351)
Recall*Advertising in $10,000,000	.1365a	(.0375)
PI	.3941a	(.0443)
Cost 0 - 2 Searches	-1.1655a	(.1954)
Cost 3 - 4 Searches	-1.3959a	(.1543)
Cost 5 - 6 Searches	-1.6161a	(.1561)
Cost 7 - 9 Searches	-1.7990a	(.1703)
Loglikelihood	-3,090.73	

Table 25: Decreasing Cost of Search
a: $<.001$, b: $<.01$, c $< .05$, d: $<.10$

	Advertising Only		**Recall Only**		**Main Effects Only**		**Main and Interaction Effects**	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
	Brand Preferences		Brand Preferences		Brand Preferences		Brand Preferences	
21st Century	-2.5261a	(.1745)	-2.8024a	(.2096)	-2.6377a	(.2288)	-2.9261a	(.2020)
AIG	-2.0622a	(.1532)	-2.5233a	(.1766)	-2.3232a	(.1696)	-2.6387a	(.1903)
Allstate	-1.8357a	(.1556)	-2.4672a	(.1894)	-2.0475a	(.2787)	-2.5369a	(.1743)
American Family	-2.2708a	(.1527)	-2.7997a	(.2054)	-2.5370a	(.2015)	-2.9945a	(.2303)
Erie	-2.4024a	(.2275)	-2.5817a	(.2188)	-2.5099a	(.2076)	-2.6158a	(.2473)
Farmers	-2.4209a	(.1867)	-2.7504a	(.1910)	-2.5459a	(.1780)	-2.9236a	(.1822)
Geico	-2.4184a	(.1531)	-3.1380a	(.1953)	-2.7750a	(.5032)	-3.0733a	(.1928)
GMAC	-2.5250a	(.1643)	-2.6121a	(.1892)	-2.7068a	(.1757)	-3.0312a	(.2257)
Hartford	-2.2489a	(.1659)	-2.6613a	(.2039)	-2.4989a	(.1692)	-2.7300a	(.2233)
Liberty Mutual	-2.2875a	(.1459)	-2.6873a	(.2284)	-2.4336a	(.1924)	-2.8174a	(.2035)
Mercury	-2.5578a	(.1562)	-2.8416a	(.2111)	-2.6273a	(.2125)	-2.9189a	(.2455)
MetLife	-2.4853a	(.2261)	-2.8223a	(.2134)	-2.7150a	(.2018)	-3.017a	(.2299)
Nationwide	-2.6029a	(.1214)	-3.1373a	(.1960)	-2.8647a	(.2287)	-3.2818a	(.1937)
Progressive	-1.9168a	(.1423)	-2.5503a	(.1769)	-2.1820a	(.3173)	-2.6168a	(.1835)
Safeco	-2.9817a	(.1522)	-3.1735a	(.1974)	-3.0147a	(.2195)	-3.2522a	(.2003)
State Farm	-2.2200a	(.1583)	-2.8000a	(.1842)	-2.4790a	(.2309)	-2.9163a	(.0920)
Travelers	-2.2730a	(.1764)	-2.3428a	(.1850)	-2.4945a	(.1794)	-2.6647a	(.2258)
	Other Parameters		Other Parameters		Other Parameters		Other Parameters	
Price in $100	-.4721a	(.0246)	-.5056a	(.0313)	-.4954a	(.0352)	-.5236a	(.0329)
Advertising Recall			.7692a	(.0785)	.5027a	(.0849)	.9967a	(.1097)
Advertising Spending in $10,000,000	-.0050	(.0251)			-.4033a	(.1041)	.0635	(.0323)
Recall*Advertising in $10,000,000							-.1291a	(.0466)
PI	.3229a	(.0467)	.3791a	(.0461)	.3464a	(.0514)	.4075a	(.0595)
Search Cost	-1.5718a	(.1674)	-1.4903a	(.1605)	-1.5147a	(.1714)	-1.4464a	(.1583)
Loglikelihood	-3,300.06		-3,295.60		-3,293.52		-3,291.89	
AIC	6,642.12		6,633.20		6,631.04		6,629.78	
BIC	6,758.55		6,749.63		6,753.02		6,757.30	

Table 26: Advertising
a: $<.001$, b: $<.01$, c $< .05$, d: $<.10$

	Predicted Retention Rate in %
Predicted under current consumer behavior	70.60
Search costs eliminated	43.41
Additionally customer satisfaction eliminated	14.42
Additionally switching costs eliminated	7.16

Table 27: Decomposition of Consumer Retention

Average Premium	Company Size	Company		At Reported Satisfaction Levels	At Maximized Satisfaction Levels
High	Large	**Allstate**	Consideration Set Share in %	10.82	11.19
			Purchase Market Share in %	12.38	13.54
			Retention Rate in %	69.84	82.95
Low	Large	**Geico**	Consideration Set Share in %	16.52	17.08
			Purchase Market Share in %	18.32	22.41
			Retention Rate in %	64.10	76.14
High	Small	**Travelers**	Consideration Set Share in %	4.14	4.24
			Purchase Market Share in %	3.56	3.78
			Retention Rate in %	72.63	81.21
Low	Small	**Safeco**	Consideration Set Share in %	2.47	2.73
			Purchase Market Share in %	2.18	2.66
			Retention Rate in %	52.52	67.18

Table 28: Value of Customer Satisfaction

	Profit Increase under	
	15% Profit Margin	25% Profit Margin
Allstate	$330 million	$521 million
Geico	$658 million	$1,108 million
Travelers	$124 million	$194 million
Safeco	$135 million	$219 million

Table 29: Increase in Profits from Maximizing Customer Satisfaction

	Allstate		Geico	
	0% Discount	10% Discount	0% Discount	10% Discount
New Customers	12.37	12.62	18.66	18.69
Existing Customers	12.37	13.58	18.66	18.94
	Allstate		**Geico**	
	0% Increase	**10% Increase**	**0% Increase**	**10% Increase**
New Customers	12.37	12.23	18.66	18.64
Existing Customers	12.37	12.18	18.66	18.50

Table 30: Effects of Changes in Premia

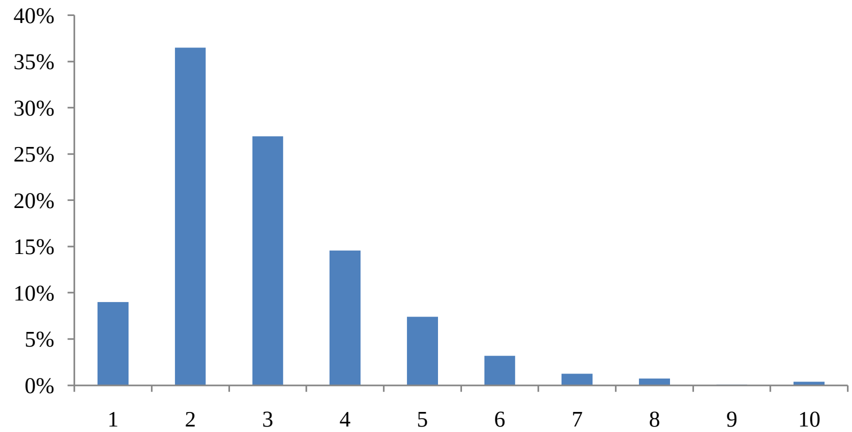

Figure 1: Distribution of Consideration Set Sizes

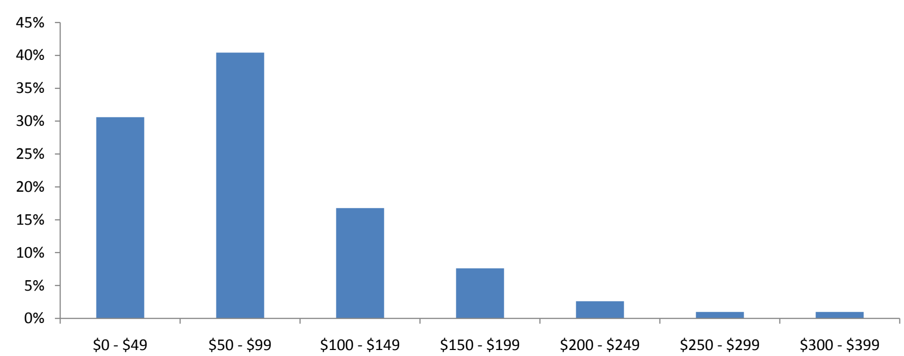

Figure 2: Switching Cost Distribution