Dynamic Price Competition in Auto Insurance Brokerage

Luis H. B. Braido[†] Bruno C. A. Ledo[‡]

June 6, 2018

Abstract

We analyze Brazilian data on auto insurance and document that: (a) about 20 percent of policies are sold without brokerage commission; (b) over 40 percent are sold at the highest fee allowed; and (c) the remaining contracts are associated with a spread-out distribution of fees. Static models cannot rationalize these findings. We develop a dynamic model of price competition with search and switching costs that reproduces them. We use the equilibrium structure to estimate the model parameters and infer the brokers' expected earnings, the frequency that insurees switch brokers, and the counterfactual effects of a price ceiling policy.

Keywords: Brokerage, commission, fee, price competition, structural estimation, auto insurance. JEL Classification: D4; L1.

[*]We are thankful for comments from Rodrigo de Losso Bueno, Marcelo Medeiros, Humberto Moreira, Mauricio Moura, Heleno Pioner, Bernard Salanié, Marcelo Sant'Anna, and seminar participants at the EUI, FGV-SP, FUCAPE, IPEA-RJ, PUC-Rio, UFRJ, USP-RP, USP-SP, and at the following meetings: Econometric Society (Malaga 2012 and Shanghai 2010), European Economic Association (Glasgow 2010), and SBE (Salvador 2010). We are also grateful to Antero Alves Neto for computing support.

[†]Getulio Vargas Foundation, FGV/EPGE, Rio de Janeiro, Brazil; luis.braido@fgv.br.

[‡]University of Sao Paulo, USP/FEA-RP, Ribeirao Preto, Brazil; brunoaurichio@gmail.com.

1 Introduction

This article analyzes the brokerage activity in the Brazilian auto insurance market. The standard auto insurance policy lasts for one year. Automatic extensions are not available, and new contracts need to be signed after expiration. Contracts are sold through certified brokers.[1] There are more than fifteen auto insurance companies and thousands of brokers. The typical broker runs a small local business, but the brokerage service is also offered by bank branches and internet firms. Insurance companies provide software programs for brokers to compute the baseline insurance price for each given list of observables — i.e., a list of characteristics of the insuree, the insured vehicle, and the insurance contract. Brokers add a fee to the baseline price. Brokerage fees must not exceed a ceiling level defined by the insurance company. Ceilings vary from 15 to 50 percent of the policy final price, depending on observed and unobserved characteristics of the market and the contract.

Our data document the coexistence of policies being sold with zero and positive brokerage fees. About 20 percent of the contracts are sold without any brokerage commission. Over 40 percent are sold at the highest fee allowed (the ceiling level). The remaining policies are associated with commission values that are smoothly distributed within an interval that ranges from zero to over 845 BRL (equivalent to about 422.50 PPP-adjusted USD).

It is challenging to rationalize the economics behind this evidence. On the one hand, the mass of zero fees suggests that the brokerage market is very competitive. On the other hand, the mass of fees at ceiling levels suggests exactly the opposite. Moreover, our regression analysis shows that the variability in brokerage fees is poorly related to a large set of observable characteristics, suggesting that randomness is an important element in brokerage pricing.

We recall that mixed strategy is sometimes used to model equilibrium price dispersion that is not related to observables — see Varian [1980], Stahl [1989], Sharkey and Sibley [1993], and Braido [2009]. We notice however that Nash equilibrium re-

[1] We use the term broker to refer to any person licensed to trade auto insurance contracts. We ignore the technical distinction between brokers and agents. In the insurance market, brokers represent the insuree and cannot have contractual agreements with insurance companies, while agents make direct selling on behalf of the carriers. They both interact with the insuree and charge a commission. There are national and regional associations of insurance brokers working to prevent carriers from discriminating against brokers.

quires that prices in the support of each player's mixed strategy generate the same expected payoff, given the other players' strategies. Therefore, the usual static games do not generate zero and positive commissions being simultaneously played in equilibrium.

We introduce switching costs into a model of price competition with informed and uninformed consumers.[2] Costs to switch brokers generate an intertemporal value for brokerage. The game has a symmetric stationary Nash equilibrium. Depending on the parameters of the model, this equilibrium is either in pure or in mixed strategies. Loosely speaking, when the switching cost is very high, the pricing strategy is degenerated, and the insurees do not change brokers over time. Otherwise, the brokers use mixed strategies over a large set of fees.

The equilibrium mixed strategies depend on the matching status of each insuree. When dealing with a new client, brokers are tempted to set low fees to increase the probability of making a sale and becoming matched to the consumer for the following period. This type of loss leading generates a mass of contracts without brokerage commission. When dealing with a regular customer, brokers take advantage of the switching cost and only offer contracts with positive fees. They face a tension between extracting consumer surplus and risking to lose the customer to a competitor. Their mixed strategy will have a mass point at the highest price allowed for that contract.

We take this symmetric stationary Nash equilibrium to data. We derive a theory-consistent data generating distribution and use it to perform a maximum likelihood estimation of the model parameters. Among them, we found that the brokers' discount factor equals .95 and that the insurees' switching cost is 213.36 BRL (equivalent to 106.68 PPP-adjusted USD). We then use the estimated parameters for three different analyses.

First, we compute that the brokers' lifetime expected profit is 790.62 BRL (equivalent to 395.31 PPP-adjusted USD) when dealing with a regular customer, and it is 577.26 BRL (equivalent to 288.63 PPP-adjusted USD) when facing a new insuree. These are important concepts in industrial organization. They are useful for analyzing mergers and acquisitions as they allow us to compute the value of a portfolio with regular customers and a stable flow of new potential clients. They are com-

[2]Similar to Varian [1980], this last feature accounts for search frictions on the consumer side.

puted here through our structural procedure, which essentially relies on price data. Second, we predict that each insuree switches brokers with a probability of 17.17 percent. This number is consistent with reports from the Brazilian Association of Insurance Brokers (FENACOR).[3] Finally, we analyze the counterfactual effects of a price ceiling policy and find that any ceiling fee lower than 63.86 BRL (equivalent to 31.93 PPP-adjusted USD) could implement a Pareto efficient allocation with much lower commission values and without costly switching.

The remaining of this article is organized as follows. We discuss related papers in Section 2 and describe the data in Section 3. We use Section 4 to formally describe our dynamic game of price competition and to derive a symmetric stationary Nash equilibrium. In Section 5, we take the equilibrium distribution and the individual decision rule to derive the theoretical data generating distribution. This distribution is used in Section 6 to identify the parameters of the model through a maximum likelihood estimation. In Section 7, we express structural parameters as functions of observable characteristics. This introduces flexibility into the model by allowing for variation in the structural parameters across consumers with different observables. However, likelihood ratio tests show that this modification did not improve the model performance. Counterfactual exercises and additional implications of our structural analysis are presented in Section 8. Concluding remarks appear in Section 9.

2 Related Literature

We document a case in which a reasonably homogeneous service is sold at very different prices. This is not consistent with frictionless models of competitive markets in which all units should be sold at the same price. However, this can be rationalized by transaction costs related to the fact that consumers waste resources to search for price quotations and to switch suppliers. We start this section by reviewing the main theories on each of these frictions. We also review some recent empirical papers analyzing markets for differentiated goods that are subject to search or switching frictions.

[3] For instance, FENACOR (2013) reports that brokers retain 70-80 percent of their customers.

2.1 Theory Review

In a seminal article, Stigler [1961] pointed out the relation between price dispersion and the fact that consumers spend resources to gather information about prices. Since then, many authors in labor economics and industrial organization have contributed to this topic. In an important work, Varian [1980] developed a model of sales in which identical oligopolistic firms set prices for a homogeneous good produced at a constant marginal cost. There are two types of consumers: informed and uninformed. Informed consumers buy at the lowest available price, whereas uninformed consumers buy from a randomly selected firm whenever its price is below the consumer reservation level. Firms face a tradeoff between lowering prices to increase the probability of making a sale to informed consumers and increasing prices to extract rent from randomly selected uninformed consumers. In equilibrium, they use mixed strategies to balance these forces. This generates random sales and ex post price dispersion.

A natural concern about Varian's model is that uninformed consumers should be able to search for price information. To address this issue, Stahl [1989] introduced a search model in which a fraction of (informed) consumers faces no search cost, whereas another fraction faces a positive search cost. As in Varian [1980], firms use mixed strategy to balance the tradeoff between selling to informed customers and extracting rent from consumers with positive search cost.

We model the search friction in a way that resembles these papers. This feature is important to generate equilibrium in mixed strategies. However, given its history independence, search cost does not rationalize the coexistence of zero and positive fees in equilibrium. Switching cost is the key element behind this equilibrium outcome in our approach, as it introduces an intertemporal value for selling today without a brokerage commission to become matched to the insuree for the next period.

Switching Cost Klemperer [1987] introduced a model in which two firms set prices over two periods of time, and consumers bear a fixed cost for switching from one firm to another in the second period. Prices are restricted to be the same for regular and new consumers. The authors focus on computing a stationary equilibrium in which the second-period pricing strategy depends on the firm's stock of regular customers. The firms charge the same price and capture the same share of the market at each given period. Padilla [1995] and Anderson et al. [2004] extended this analysis to an environment with overlapping generations and infinitely many periods.[4] Like in Klemperer [1987], the two firms cannot price discriminate between regular and new consumers, and the equilibrium pricing depends on each firm's market share.

In a second front, there are papers analyzing environments with price discrimination. Chen [1997] and Taylor [2003] presented models in which two firms set different prices for regular and new consumers over a finite number of periods.[5] Consumers face different switching costs in each period, according to *i.i.d.* draws from an exogenous distribution function. In equilibrium, the two firms offer the same price for their own regular customers and the same discount price for their rival's consumers. The exogenous distribution of switching costs determines the fraction of consumers who switch suppliers.[6]

There are several differences between our work and those in the switching cost literature.[7] The most important is the type of price differentiation allowed in our model. Insurance brokers usually deal to one consumer at a time. It is then natural in our setting to allow for price differentiation between regular and new consumers as well as across insurees within each of these groups. Our equilibrium mixed strategies depend on whether or not the insuree is matched to the broker. They are unrelated to the broker's market share. Because brokers set prices to each consumer at a time, identical consumers end up facing different prices.

2.2 Empirical Papers

There is a large empirical literature on the importance of search costs and product differentiation to explain price dispersion. Models of product differentiation

[4] See Farrell and Shapiro [1988] and Beggs and Klemperer [1992] for alternative infinite-horizon frameworks.

[5] Fudenberg and Tirole [2000] and Villas-Boas [2006] also studied dynamic models of price discrimination based on consumer's purchasing history.

[6] In a relatively different context, Chen and Rosenthal [1996] derived a mixed-strategy equilibrium for a dynamic duopoly in which consumers are loyal to firms, but loyalty varies over time according to firms realized prices and an exogenous stickiness rule.

[7] We refer the reader to Farrell and Klemperer [2007] for an extensive survey on this topic.

typically consider one product with different brands and consumers with heterogeneous preferences for the existing brands. We refer the reader to following works in this literature: Dahlby and West [1986] for an analysis of auto insurance prices in Alberta, Canada; Sorensen [2000] for data on prescription drugs in New York drugstores; Brown and Goolsbee [2002] for life-insurance prices in the USA; Lach [2002] for prices of some retail items (such as refrigerator, chicken, coffee, and flour) selected from the Israeli Consumer Price Index; Hortaçsu and Syverson [2004] for fees charged by hedge funds in the USA; Hong and Shum [2006] for online prices of books in the USA; Wildenbeest and Moraga-González [2008] for online prices of computer memory chips; Wildenbeest [2011] for grocery prices in the UK; De Los Santos et al. [2012] for very rich data on consumers' searching behavior and prices from online bookstores; Allen et al. [2014a] for interest rates in mortgage contracts in Canada; and De Los Santos et al. [2017] for data on web browsing and purchasing behavior of MP3 players.

We are aware of a few empirical works analyzing the importance of switching costs.[8] Dubé et al. [2009] analyzed dynamic price competition in an environment with differentiated goods (i.e., one good with different brands). In each period, consumers' utilities depend on: (a) the price of the good for that period; (b) a random coefficient associated with each brand; (c) a fixed switching-cost parameter that interacts with a categorical variable indicating the brand selected by the consumer in the previous period; and (d) a random variable with a type I extreme value distribution. Consumers are myopic and make choices that maximize their current utility (static problem). This leads to a demand system represented by a logit model with random coefficients for brands and a fixed switching-cost parameter. Firms make dynamic choices regarding prices. The authors conjectured that a pure-strategy Markov equilibrium exists and simulate the equilibrium conditions in the estimation of the model parameters. Honka [2014] took the basic structure from Dubé et al. [2009] and introduced a search model to endogeneize the set of brands (in her case, the set of insurance companies) available for each consumer.

[8] We focus this review on structural papers, but we must mention an early empirical work by Israel [2005]. The author performed a reduced-form analysis using a panel data on auto insurance policies contracted in Georgia, USA, from 1991 to 1998. He showed that the length of time that an insuree is covered by a given firm reduces the probability that this insuree switches to another firm.

Allen et al. [2014b] used data from the Canadian mortgage market to analyze three frictions acting together: search, switching, and bargaining. Consumers face a static problem. They must choose one bank to finance their home. They do not hold preferences for banks (brands), but they are born attached to one of them (where they hold a currency account, for instance). They face no randomness in their utility functions and only differ from each other in the cost of making the deal with any bank different from the one where they already hold an account. Banks, on the other hand, face random costs that depend on a common factor associated with each borrower and on a privately observed idiosyncratic component. Each consumer observes a price offer from their home bank and decides whether to take it or to costly search for other banks. In the second case, they select a set of banks to visit and start a multilateral negotiation with them and with the home bank. Search costs are linearly proportional to the number of banks visited. The multilateral negotiation is modeled through an English auction. The equilibrium is simple to compute, and prices depend on the random components of the bank's cost structure. In other words, while search costs affect the endogenous number of banks bidding in the auction, price dispersion is strongly related to exogenous variations in costs.

Our paper differs from all these empirical papers in one important regard. We model homogeneous brokers selling to homogeneous consumers. As a result, our price dispersion is fully endogenous and is neither related to exogenous heterogeneity on consumer's preferences for brands nor to firm's costs. We model search cost in a very simple fashion, based on Varian [1980]. It is not our goal to evaluate different ways to model the search friction. When compared with the three structural papers with switching costs, we stress that: (a) different from Dubé et al. [2009] and Honka [2014], consumers in our paper are not myopic, and we do not rely on simulated methods as we explicitly derive the Markov equilibrium strategy; and (b) different from Allen et al. [2014a], we do not account for multilateral negotiation, but our firms solve a dynamic optimization problem.

3 Data

We use data recorded by the Brazilian Superintendence of Private Insurance (SUSEP), a federal institution responsible for regulating the insurance market in the country. Auto insurance companies are commanded to provide detailed information on all active policies. The SUSEP database conveys cross-section information about the population of auto insurance policies that have been active (for at least one day) during each semester. We use data relative to one-year policies that were active in the first semester of 2003. We focus our analysis on comprehensive policies contracted by individuals.[9] For each policy, we access information about the characteristics of the insuree (such as gender and age), the insured vehicle (such as model and year), and the insurance contract (such as the insurance company, type of coverage, and deductible values). The insurance premium and the brokerage fee embedded in that price are also available.[10] There is no information about the broker. Table 1 defines all variables used in the article.

[Table 1]

The insurance policies are associated with a wide range of brokerage fees. This feature is very robust, and similar patterns are observed in different subsamples. Figure 1 presents histograms for nominal brokerage fees expressed in BRL, and Figure 2 displays the brokerage fees expressed as percentages of each policy final price. We report results for the following six different groups.

Subsample A contains the fees paid to brokers across all insurance contracts in the country. Subsample B focuses on contracts in the metropolitan area of Sao Paulo (the largest in the country). Subsample C presents fees on policies covering eight comparable compact cars with identical power train in the metropolitan area of Sao Paulo.[11] Subsample D refines subsample C to include only contracts with positive bonus discount. Insurees with positive bonus discount are particularly interesting because they must have bought an insurance policy before and, then, they most likely conform our model assumption of being initially matched to a broker. (The insurees carry over the bonus status after changing the insurance company or the broker.) Subsamples E refines subsample D to include only new compact cars. Subsample F

[9]We do not analyze: (i) policy endorsements and cancelations; (ii) collective policies covering multiple vehicles; (iii) policies contracted by firms; and (iv) noncomprehensive contracts.

[10]We also access files containing information on reported accidents, other claims, and disbursements. However, these files are not useful for the purposes of this article.

[11]The vehicle models are: Celta 1.0L, Clio 1.0L, Corsa 1.0L, Fiesta 1.0L, Gol 1.0L, Ka 1.0L, Palio 1.0L, and Uno 1.0L. We did not include sportive, turbo, or 16V versions of these models.

further refines subsample E and includes only new Volkswagen Gol 1.0L and Fiat Palio 1.0L, the two most popular vehicle models.

On average, about 20 percent of the insurance policies are sold with zero fees. Furthermore, there is large dispersion in both nominal and percentage fees. The shapes of the histograms are surprisingly similar across the different subsamples.

[Figures 1 and 2]

The class width of the histograms in Figure 2 is 0.5 percentage point, and there are contract fees in virtually all classes between zero and 42 percent. We notice a higher frequency of fees equal to 1, 5, 10, 15, 20, 25, 30, and 35 percent of the contract's final price. In addition to the fact that brokers round some numbers, the insurance companies impose percentage caps to brokerage fees. These percentage caps vary across companies and among contracts of a given company. The most frequent caps are 15, 20, 25, 30, and 35 percent. We stress however that policies associated with nonround percentage fees — such as 12.9 and 23.39 percent — are not rare. (They account, for instance, for 14 percent of subsample E.)

What Brokers Say We conducted structured interviews with 20 registered brokers to improve our understanding about institutional details of this market. The computer programs used by brokers record the contract final price, the brokerage nominal fee, and the percentage fee. Brokers argued they use these three pieces of information to make their pricing decisions. In many times, they round the percentage fee, the final price, or the monthly installment (when the sale is financed). They also confirmed that insurance companies impose ceilings on brokerage fees expressed in percentage terms.

Selected Subsample We use subsample E in our quantitative analyses throughout the article. The choice of analyzing a more homogeneous subsample is justified by the fact that we attempt to understand the economics behind the price dispersion that is not related to consumer heterogeneity. We recall that subsample E contains policies signed by insurees with positive bonus discount (and therefore not new in the market), covering new compact cars with identical power train, in the metropolitan area of Sao Paulo. This is the largest metropolitan area in the country and, according to SUSEP's categories, it includes 29 connected cities. Figure 3 lists these 29 cities with their respective zip codes. We stress that the city of Sao Paulo itself is indicated with number 1. It is subdivided in 99 neighborhoods in Figure 4.

[Figures 3 and 4]

We perform reduced-form OLS regressions (using subsample E) for four different dependent variables: the insurance premium; the broker percentage fee; the brokerage commission; and a dummy variable indicating whether the policy was sold without brokerage commission. Our econometric models use independent variables describing details about the policy coverage, the insuree's gender, age (linear, quadratic, and categorical controls), and bonus category. We also control for the vehicle model, the insuree's bonus category, and the insurance company. Moreover, we divide the metropolitan area in 135 regions according to the first 3 digits of the zip code and introduce categorical controls for these regions.

The results are presented in Table 2. Column zero lists the independent variables used in the estimation. The regression in column 1 analyzes the insurance premium. The R-squared in this regression shows that 49 percent of the variability in the insurance premium is explained by the linear regression model. This is a good fit considering that we use a simple linear model and that the insurance companies access additional information that is not required to be reported to SUSEP.

We then pursue a similar analysis with respect to the variables describing different aspects of the brokerage commission. The regression presented in column 2 shows that our control variables account for only 5 percent of the variation in the percentage fees. This strongly suggests there are economic forces driving broker's decisions that are not related to the characteristics of the insurees and the contracts.

Next, we use column 3 to show that our control variables explain 13 percent of the variation in brokerage fees. We recall that these fees are defined by the product of the insurance premium and the broker percentage fee (i.e., the variables in the first two regressions). Finally, these same controls explain 15 percent of the variation in the dummy variable describing policies sold with and without brokerage commission (see column 4).

We conclude that differences in the observable variables are more relevant to explain the dispersion in the insurance premium than the variability in the three variables measuring different aspects of the brokerage commission. This supports our view that an important part of the dispersion in brokerage fees is exogenously random.

[Table 2]

4 A Recursive Model of Price Competition

There are a few distinguishing features in the auto insurance market. Policies last for one year and must be sold through an insurance broker. There is no automatic extension, and long-term contracts are not enforceable. Insurance companies define a positive baseline price b for each policy. Brokers add a fee q to that price. Insurance companies require that the brokerage fees do not exceed a percentage cap τ of the policy final price $p := b + q$. The condition $q \leqslant \tau p$ is equivalent to $q \leqslant r$, where $r := \frac{\tau}{1-\tau} b$ is a nominal ceiling for brokerage fees.

Remark 1. The baseline price b and the percentage cap τ are random variables, in the sense that they take different values for each observation in the data. We assume brokers are well-informed and take as given the values of b and τ associated with each insuree.

Game Structure Brokers are allowed to set a different fee for each quotation. This is as if they competed for each insuree at a time. We therefore model an environment with a single insuree, a continuum of identical brokers, and an infinite sequence of periods $t \in \mathbb{T} := \{0, 1, ...\}$. The insuree demands one insurance policy every period and seeks to minimize the expected lifetime discounted costs. He starts each period matched to a broker and faces a fixed cost $c > 0$ for switching to another broker.[12]

In each period, the insuree receives a price quotation from his matched broker. With probability $\mu \in (0, 1)$, this is the only quotation he receives. However, with probability $(1 - \mu)$, he accesses a second price quotation from an unmatched broker,

[12] In period $t = 0$, the matched broker is randomly selected. After that, the insuree stays matched to the last broker he purchased from.

randomly selected from the continuum set of brokers. The second quotation event is *i.i.d.* across periods and privately observed by the insuree.

Brokers select a price offer for each period they are called to play. They cannot commit to prices in the future. They maximize expected lifetime discounted profits and share the same intertemporal discount factor $\delta \in (0,1)$. The brokerage commission cannot exceed the ceiling $r > 0$. In principle, fees are not allowed to be negative. However, when dealing with a new insuree, the broker can offer assistance services that reduces the consumer switching cost. To represent this possibility, we allow the unmatched broker to set negative fees down to $-c$. Hence, brokerage fees q must lie in the interval $[0, r]$ when the broker deals with a regular (matched) insuree and in $[-c, r]$ when she deals with a new (unmatched) insuree.

The timing of the game is as follows. Nature moves first. It selects an unmatched broker out of the continuum of brokers and determines whether the insuree meets this broker or not. The latter event (not meeting an unmatched broker) occurs with probability $\mu \in (0,1)$. It is private to the insuree and cannot be credibly revealed to the matched broker. Second, the two selected brokers choose their pricing strategies. Then, the insuree purchases one policy. The broker who makes the sale becomes matched to the insuree, and the game ends for the other broker. An identical game starts next period.

Strategies Brokers' payoffs depend on their matched/unmatched statuses. Conditional on this status, the payoffs do not directly depend on the history of the game or on the broker's name. Although strategies could still depend on this information, we will focus on equilibria in which all players select the same stationary strategy (σ_m, σ_u), where σ_m is a probability distribution over $[0, r]$ to be used when playing under the matched status, and σ_u is a distribution over $[-c, r]$ to be used when unmatched.

Brokers' Payoff Functions For a given strategy σ_u followed by the unmatched broker, we denote by $\pi_m(q, \sigma_u)$ the probability that the matched broker sells the policy when she poses the fee q. The expected lifetime discounted payoff of a broker facing a matched insuree is implicitly defined by the following equation:

$$V_m(\sigma_m, \sigma_u) = \int_0^r \pi_m(q, \sigma_u)\,(q + \delta V_m(\sigma_m, \sigma_u))\,d\sigma_m(q),$$

where $\delta V_m(\sigma_m, \sigma_u)$ is the equilibrium continuation value of becoming matched to the consumer next period after selling a policy today.[13]

For a given strategy σ_m, we let $\pi_u(q, \sigma_m)$ represent the probability that the unmatched broker — after being selected to play — sells the contract when her fee is q. The expected lifetime discounted payoff of a broker selected to play under the unmatched status is

$$V_u(\sigma_m, \sigma_u) := \int_{-c}^r \pi_u(q, \sigma_m)\,(q + \delta V_m(\sigma_m, \sigma_u))\,d\sigma_u(q).$$

We impose a tie-breaking rule that favors the matched broker in case of a tie. When the distribution σ_u is atomless, we obtain

$$\pi_m(q, \sigma_u) = \mu + (1-\mu)(1 - \sigma_u(q - c)) \tag{1}$$

and

$$\pi_u(q, \sigma_m) = 1 - \sigma_m(q + c). \tag{2}$$

In equation (1), the fraction μ stands for the probability that the insuree has no other option but to take the offer from his matched broker. Moreover, the term $(1-\mu)(1-\sigma_u(q-c))$ is the probability that the insuree meets a second (unmatched) broker, but her fee does not beat offer q posed by the matched broker. In equation (2), the term $1 - \sigma_m(q + c)$ is the probability that the fee offered by the matched broker exceeds $q+c$ and turns out to be less attractive than the fee q posed by the unmatched broker.

Remark 2. The unmatched broker infers that she only plays after nature has promoted a meeting with the insuree. For this reason, the probability $(1 - \mu)$ does not affect the definition of π_u and V_u. The value of being called to play under the unmatched status before nature decides whether the insuree will or will not access

[13] Throughout the article, we use $\int_{\theta_0}^{\theta_1}$ to represent the Lebesgue integral over the set of real numbers larger than θ_0 and smaller than θ_1. In cases where $\theta_1 < \theta_0$, this set is empty and the integral is zero.

the unmatched offer is $(1-\mu)V_u(\sigma_m, \sigma_u)$.

Definition 1 (Equilibrium Concept)**.** A symmetric stationary Nash equilibrium is described by a strategy (σ_m, σ_u) for which there is no profitable one-period deviation — that is, there is no pair of probability distributions σ'_m over $[0, r]$ and σ'_u over $[-c, r]$ such that either

$$V_m(\sigma_m, \sigma_u) < \int_0^r \pi_m(q, \sigma_u)\,(q + \delta V_m(\sigma_m, \sigma_u))\, d\sigma'_m(q)$$

or

$$V_u(\sigma_m, \sigma_u) < \int_{-c}^r \pi_u(q, \sigma_m)\,(q + \delta V_m(\sigma_m, \sigma_u))\, d\sigma'_u(q).$$

Remark 3. We follow Duffie et al. [1994] and focus on independent one-period deviations to define recursive equilibrium. This is without loss of generality. In a stationary equilibrium, players' strategies do not relate their current actions to past information that is not summarized by their current matching status. In principle, deviations are allowed to depend on it, but no broker can profit from conditioning behavior on the history of this game given that no other player does so.

4.1 Pure-Strategy Equilibrium When $\frac{\mu r}{1-\delta\mu} \leqslant c$

Consider a deviation strategy in which the matched broker charges the ceiling fee r with probability 1. In the worst-case scenario, this broker only sells to the fraction μ of insurees who do not access another quotation. If this was repeated over time, then the broker's discounted payoff would be $\sum_{t=0}^{\infty}(\delta\mu)^t \mu r = \frac{\mu r}{1-\delta\mu}$.

Let us consider now an economy with $\frac{\mu r}{1-\delta\mu} \leqslant c$. In this case, there is a pure-strategy equilibrium in which σ_m degenerates to $\min\{(1-\delta)c, r\}$ and σ_u degenerates to $\min\{-\delta c, r-c\}$. The matched broker sells the policy every period and earns a lifetime payoff of $\min\left\{c, \frac{r}{1-\delta}\right\}$. All others earn zero from this insuree.[14]

Given the equilibrium strategies, the unmatched broker would make a loss if she undercut prices. Moreover, the matched broker does not profit by increasing prices. If she did so, she would sell only to the fraction μ of insurees who do not access another quotation. She would then earn at most $\frac{\mu r}{1-\delta\mu}$, which is smaller than her equilibrium payoff $\min\left\{c, \frac{r}{1-\delta}\right\}$ by construction.

[14]Recall that our tie-breaking rule favors the matched broker.

4.2 Mixed-Strategy Equilibrium When $\frac{\mu r}{1-\delta\mu} > c$

The pure strategy described before is not a Nash equilibrium when the matched broker can deviate and collect $\frac{\mu r}{1-\delta\mu} > c$. The following mixed-strategy equilibrium emerges in these environments.

Let σ_u be an atomless probability distribution with support given by $[-a, r-c)$, where $-c < -a < r-c$. Let σ_m be a probability distribution with support given by $[c-a, r]$, which is atomless in $[c-a, r)$ and has a mass point at r. The lifetime expected payoff associated with an action $q \in [0, r]$ taken by the matched broker is

$$v_m(q) := \left[\mu + (1-\mu)(1-\sigma_u(q-c))\right](q + \delta V_m(\sigma_m, \sigma_u)). \tag{3}$$

The lifetime expected payoff associated with a fee $q \in [-c, r]$ for the broker called to play under the unmatched status is

$$v_u(q) := (1-\sigma_m(q+c))(q + \delta V_m(\sigma_m, \sigma_u)). \tag{4}$$

Strategy When Unmatched We define

$$\bar{v}_m := \frac{\mu r}{1-\delta\mu} \tag{5}$$

and set $\sigma_u(\cdot)$ to be such that

$$v_m(q) = \bar{v}_m, \ \forall q \in [c-a, r]. \tag{6}$$

If we take $v_m(q)$ and $\bar{v}_m$ from equations (3) and (5) and solve condition (6) for σ_u, then we obtain the following expression for the strategy of brokers under the unmatched status:

$$\sigma_u(q) = 1 - \frac{\mu}{1-\mu}\left(\frac{r}{\delta\mu r + (1-\delta\mu)(q+c)} - 1\right), \ \forall q \in [-a, r-c). \tag{7}$$

By setting $\sigma_u(-a) = 0$, we obtain

$$a = c - \frac{(1-\delta)\mu r}{1-\delta\mu}. \tag{8}$$

It is simple to verify that $-c < -a < r - c$, $0 \leqslant \sigma_u(\cdot) \leqslant 1$, and $\lim_{q \to r-c} \sigma_u(q) = 1$.

Strategy When Matched Using analogous reasoning, we take $\bar{v}_u > 0$ and define σ_m to be such that

$$v_u(q) = \bar{v}_u, \ \forall q \in [-a, r-c). \tag{9}$$

We set $\sigma_m(c-a) = 0$ and use equations (4), (5), and (8) to obtain

$$\bar{v}_u = \frac{\mu r}{1 - \delta\mu} - c. \tag{10}$$

Because $\frac{\mu r}{1-\delta\mu} > c$, we have $\bar{v}_m > \bar{v}_u > 0$. From our previous derivation, we obtain

$$\sigma_m(q) = \begin{cases} 1 - \frac{\mu r - (1-\delta\mu)c}{\delta\mu r + (1-\delta\mu)(q-c)}, & \text{for } q \in [c-a, r); \\ \\ 1, & \text{for } q = r. \end{cases} \tag{11}$$

Notice that $\sigma_m(c-a) = 0$ as initially set. The distribution σ_m is continuous at any $q < r$ and has a mass at the point r. We also have $0 \leqslant \sigma_m(\cdot) \leqslant 1$. Moreover, thanks to $\frac{\mu r}{1-\delta\mu} > c$, we have

$$0 < 1 - \lim_{q \to r_-} \sigma_m(q) = \frac{\mu r - (1-\delta\mu)c}{r - (1-\delta\mu)c} < 1. \tag{12}$$

Proposition 1. *The strategy (σ_m, σ_u) defined by conditions (7), (8), and (11) is a symmetric stationary Nash equilibrium in economies with $\frac{\mu r}{1-\delta\mu} > c$.*

Proof. The matched broker is indifferent to any action in $\operatorname{supp} \sigma_m = [c-a, r]$. Moreover, it follows from conditions (3), (5), and (8) that fees smaller than $c - a$ generate a payoff smaller than $\bar{v}_m$. The unmatched brokers are also indifferent to any action in $\operatorname{supp} \sigma_u = [-a, r-c)$, and fees outside this interval cannot increase this broker's payoff. □

Remark 4. Notice from equation (8) that $c - a > 0$. Therefore, the fees posed by the matched broker are always positive. When $-a < 0$, there is a positive probability that the unmatched broker offers assistance services to alleviate the switching cost.

5 Data Generating Probability

In our recursive model, brokers pose prices for each insuree at a time. They share the same discount factor $\delta \in (0,1)$ and are assumed to know the values of r, c, and μ associated with each insuree. When taking the model to data, we assume that $\mu \in (0,1)$ and $c \in \mathbb{R}_{++}$ are constant parameters and let r vary across insurees. Moreover, we recall that

$$r := \frac{\tau}{1-\tau} b. \tag{13}$$

Case 1 Our data is generated from one of two possible scenarios. When $\frac{\mu r}{1-\delta\mu} \leqslant c$, our model predicts that the insurance policy is sold at $\min\{(1-\delta)c, r\}$. In these cases, the data generating probability depends only on the stochastics of the random variables b and τ.

Case 2 When $\frac{\mu r}{1-\delta\mu} > c$, we have a mixed-strategy equilibrium in which the insuree may take any of the quotations posed by the brokers. We do not observe the price quotations but only the contracted fees. Our theoretical data generating probability must then combine the equilibrium strategy (σ_m, σ_u) with the insuree's selection rule.

According to our model, each insuree observes a matched offer q_m drawn from σ_m. Moreover, with probability $(1-\mu)$, he also observes an unmatched offer q_u drawn from σ_u. When observing two offers, the insuree accepts the fee q_m if $q_m \leqslant q_u + c$, and he takes q_u otherwise.

Let $y \in \mathbb{R}_+$ denote the brokerage fee recorded in the data (i.e., the winning offer). The conditional probability $\Pr(y \mid b, \tau)$ represents the data generating distribution of y given (b, τ). Recall that negative fees represent the situation in which the unmatched broker charges zero and offers assistance services to alleviate the switching costs. Because assistance services are not recorded in the data, our observations of y are censored at zero. Moreover, nonpositive fees must be drawn from the distribution σ_u. Each insuree accesses an unmatched offer with probability $1-\mu$ and takes it whenever it exceeds the quotation posed by the matched broker plus the switching cost. This implies

$$\Pr(0 \mid b, \tau) = (1-\mu) \int_{-a}^{\min\{0, r-c\}} (1 - \sigma_m(q+c))\, d\sigma_u(q).^{15} \tag{14}$$

The distribution $\Pr(y \mid b, \tau)$ has no mass point over the open interval $(0, r)$. For $y \in (0, r)$, we have:

$$\Pr(y \mid b, \tau) = (1-\mu) \int_{-a}^{y} (1 - \sigma_m(q+c))\, d\sigma_u(q) \\ + (1-\mu) \int_{c-a}^{y} (1 - \sigma_u(q-c))\, d\sigma_m(q) + \mu\sigma_m(y). \tag{15}$$

The first two terms in this equation are associated with the scenario in which the insuree accesses two offers. This occurs with probability $1-\mu$. In this case, with probability $1-\sigma_m(q+c)$, the insuree takes the offer posed by the unmatched broker. Alternatively, with probability $1-\sigma_u(q-c)$, the insuree takes the offer drawn from the matched broker strategy σ_m. The last term in equation (15) accounts for the fact that, with probability μ, the insuree accesses only one offer $q < y$ drawn from the matched broker strategy σ_m. We also recall that:

$$\sigma_m(q) = d\sigma_m(q) = 0,\ \forall q \leqslant c - a; \tag{16}$$

and

$$d\sigma_u(q) = 0,\ \forall q \geqslant r - c. \tag{17}$$

This means that fees in $(0, c-a)$ are never generated from the distribution σ_m which support is $[c-a, r]$, whereas fees in $(r-c, r)$ are never generated from the distribution σ_u which support is $[-a, r-c)$.

We conclude by pointing out that the distribution $\Pr(\cdot \mid b, \tau)$ has a second mass point at the ceiling fee $r := \frac{\tau}{1-\tau} b$. Because the support of σ_u is $[-a, r-c)$, fees equal to r can only be taken by insurees who do not meet an unmatched broker. This is

[15]Notice from equation (8) that $-a < r - c$. Moreover, recall from footnote 13 than this integral is zero when $-a > 0$.

to say that

$$1 - \lim_{y \to r} \Pr(y \mid b, \tau) = \mu \left(\frac{\mu r - (1 - \delta\mu)c}{r - (1 - \delta\mu)c} \right).^{16} \tag{18}$$

6 Maximum Likelihood Estimation

The insurance baseline price b takes a different value for each observation, and these realizations are available from the data. This is unfortunately not the case for the percentage ceiling τ. We accessed documents from the insurance companies and reports from the broker's association mentioning ceilings on brokerage fees at five different percentage levels: 15, 20, 25, 30, and 35 percent of the policy final price. For about 99.86 percent of observations in subsample E, the observed percentage fees do not exceed the 35 percent cap. There are, however, 13 observations with percentage fees in $(.35, .42]$. To account for that, we also include the value 42 percent as a ceiling.

The realizations of τ are known by the brokers, but they are not recorded in our data. We then treat τ as a latent random variable occurring with probability $\varphi(\tau)$. More precisely, we assume φ is represented by

$$\varphi(\tau|\xi) := \frac{\exp(\xi_\tau)}{\sum_\tau \exp(\xi_\tau)}, \tag{19}$$

where $\tau \in \{.15, .20, .25, .30, .35, .42\}$, $\xi := (\xi_{.15}, ..., \xi_{.35}) \in \mathbb{R}^5$, and $\xi_{.42}$ is normalized to zero to make the system determined.

The logistic functional form is assumed here without loss of generality. It is also used later to make the model conditional on observables, in which case it imposes restrictions on the conditional effects.

6.1 Likelihood Function

We adapt the maximum likelihood procedure for censored data described in Meeker and Escobar [1994] to estimate the parameters. Each observation j is associated with a realization of the baseline price b_j. Moreover, each observa-

[16] We notice that $\int_{-a}^{r-c} (1 - \sigma_m(q + c))\, d\sigma_u(q) + \int_{c-a}^{r} (1 - \sigma_u(q - c))\, d\sigma_m(q) = 1$. This is because, for any independent realizations of q_u and q_m, respectively draw from σ_u and σ_m, we have that $q_m > q_u + c$ if and only if $q_u < q_m - c$. This implies $\int_{-a}^{r-c} (1 - \sigma_m(q + c))\, d\sigma_u(q) = \int_{c-a}^{r} \sigma_u(q - c) d\sigma_m(q)$.

tion j is generated from one of six possible games, indexed by τ. For each observation j and a game τ, we set $r_{j,\tau} := \frac{\tau}{1-\tau} b_j$ and define an auxiliary likelihood function $L(\delta, \mu, c \mid y_j, b_j, \tau)$ as follows.

Case 1 When $\frac{\mu r_{j,\tau}}{1-\delta\mu} \leqslant c$, we define

$$L(\delta, \mu, c \mid y_j, b_j, \tau) := \begin{cases} 1, & \text{if } y_j = \min\{(1-\delta)c, r_{j,\tau}\}; \\ 0, & \text{otherwise.} \end{cases} \tag{20}$$

Case 2 When $\frac{\mu r_{j,\tau}}{1-\delta\mu} > c$, we have three possible scenarios. For policies sold without brokerage commission (censored observation), we take equation (14) and define

$$L(\delta, \mu, c \mid y_j, b_j, \tau) := \Pr(0 \mid b_j, \tau), \quad \text{if } y_j = 0. \tag{21}$$

For observations associated with a brokerage fee at the percentage ceiling τ, we take equation (18) and write

$$L(\delta, \mu, c \mid y_j, b_j, \tau) := \mu \left(\frac{\mu r_{j,\tau} - (1-\delta\mu)c}{r_{j,\tau} - (1-\delta\mu)c} \right), \quad \text{if } y_j = r_{j,\tau}. \tag{22}$$

For observations associated with a brokerage fee $y_j \in (0, r_{j,\tau})$, we derive equation (15) to find a density probability and write

$$\begin{aligned} L(\delta, \mu, c \mid y_j, b_j, \tau) := {} & (1-\mu)(1-\sigma_m(y_j + c))\, d\sigma_u(y_j) \\ & + [(1-\mu)(1-\sigma_u(y_j - c)) + \mu]\, d\sigma_m(y_j). \end{aligned} \tag{23}$$

We stress that equations (8) and (16) imply

$$\sigma_m(y_j) = d\sigma_m(y_j) = 0, \quad \text{for } y_j \leqslant \frac{(1-\delta)\mu r_{j,\tau}}{1-\delta\mu}. \tag{24}$$

Moreover, equation (17) implies

$$d\sigma_u(y_j) = 0, \quad \text{for } y_j \geqslant r_{j,\tau} - c. \tag{25}$$

We recall that each τ is independently withdrawn with probability $\varphi(\tau|\xi)$. Hence, the likelihood function associated with each observation j is

$$\mathbf{L}(\delta, \mu, c, \xi \mid y_j, b_j) := \sum_{\tau} L\left(\delta, \mu, c \mid y_j, b_j, \tau\right) \varphi(\tau|\xi). \tag{26}$$

The final likelihood function for a given data $D := \{(y_j, b_j), \forall j\}$ is

$$\mathcal{L}(\delta, \mu, c, \xi \mid D) := \prod_{j} \mathbf{L}(\delta, \mu, c, \xi \mid y_j, b_j). \tag{27}$$

The associated log-likelihood function is

$$ln\mathcal{L}(\delta, \mu, c, \xi \mid D) = \sum_{j} ln\left[\mathbf{L}(\delta, \mu, c, \xi \mid y_j, b_j)\right]. \tag{28}$$

6.2 Numerical Estimation

Our model has 3 structural parameters (δ, μ, c) and 5 logit parameters ξ_τ determining the probabilities in equation (19). This leaves us with a total of 8 parameters to estimate. We use the software *Wolfram Mathematica* to obtain a closed-form solution for the integral in equation (14). This is used to compute the function (21). The optimization code is written in MATLAB and uses the patternsearch algorithm. This algorithm performs numerical optimization by means of different direct search algorithms such as the generalized pattern search (GPS), the generating set search (GSS), and the mesh adaptive search (MADS). These are derivative-free methods that are convenient here because our likelihood function is not continuous.

We use a grid search and a random algorithm to select the initial seeds for the optimization. In the fist method, we compute the log-likelihood function for a grid of points. We draw 1 million *i.i.d.* points for the parameters (δ, μ, c) and find values for ξ_τ that minimize the distance between the theoretical and the empirical fraction of observations at the mass points 0, 15, 20, 25, 30, 35, and 42 percent. We use the value of the log-likelihood function to sort the results from this grid and use the first 2,000 points as seeds for our numerical optimization on the set of admissible parameters (δ, μ, c, ξ). In the alternative approach, we also perform 2,000 optimizations using random seeds on our 8 parameters. The estimated values for the two methods coincide.

Our results are presented in Table 3. The estimated brokers' discount factor is .95. The estimation for μ is .80. This means that about 20 percent of the insurees see a price quotation from a new (unmatched) broker in addition to the quotation from the regular (matched) broker. The estimated value for the switching cost c is 213.36 BRL (i.e., approximately 106.68 PPP-adjusted USD). Table 3 also presents the estimates for the logit parameters ξ and 95-percent confidence intervals. These intervals were computed by reestimating the model for 1,000 subsamples with one-third the size of the main sample. The observations are randomly withdrawn with replacement.

[Table 3]

6.3 Discontinuities

We have mentioned that our likelihood function is not continuous. Let us briefly comment here on each of these discontinuity points.

The first source of discontinuity derives from the fact that the equilibrium strategy depends on whether or not $\frac{\mu r_j}{1-\delta\mu} \leqslant c$ (pure- versus mixed-strategy equilibrium). When $\tau = .42$, the estimation selects the mixed-strategy equilibrium for all observations. In the other extreme, when $\tau = .15$, the ML estimates selects the pure-strategy equilibrium for 277 observations and the mixed-strategy equilibrium for all others. For every observation j and all possible values of τ, the estimated parameters are never at the discontinuity point in which $\frac{\mu r_j}{1-\delta\mu} = c$.

The second point of discontinuity is when $y_j = (1-\delta)c$ at a pure-strategy equilibrium. This follows from equation (20).[17] Again, in our estimation, this equality is never active for (j, τ) associated with a pure-strategy equilibrium.

The third and fourth points of discontinuities are associated with equations (24) and (25) in the mixed-strategy equilibrium. Once again, the estimated parameters are such that the equalities $y_j = \frac{(1-\delta)\mu r_{j,\tau}}{1-\delta\mu}$ and $y_j = r_{j,\tau} - c$ do not hold, for any (j, τ) associated with a mixed-strategy equilibrium.

[17] Notice that $r_{j,\tau}$ is not a parameter to be estimated.

6.4 Model Fit

We analyze the empirical performance of our model. We compare the continuous part of the theoretical and empirical distributions in Figure 5. We use bars to represent the empirical histogram and dots to simulate our likelihood function. The mass points are analyzed in Table 4, which presents the empirical and estimated frequency of policies sold at each different percentage ceilings. The overall results are very good. The empirical distribution of positive nominal fees is nicely replicated by the model. The same is true for all mass points of the model.

[Figure 5]
[Table 4]

7 Conditioning on Observable Characteristics

Our model was estimated assuming that brokers and insurees are identical within each group—that is, the broker's discount factor (δ) is constant, and the other parameters are equal across all insurees. The model can explain a great deal of the dispersion in the brokerage fees without relying on any type of heterogeneity of sellers or buyers.

We do observe, however, some characteristics of the insuree. We can then allow their parameters to depend on this information and check whether this improves the model fit. Let x be a k-dimensional random vector representing the insuree's characteristics and write the switching cost as a function:

$$c(x) = \exp(\alpha_0 + x \cdot \alpha),$$

where $\alpha_0 \in \mathbb{R}$ and $\alpha := (\alpha_1, ..., \alpha_k) \in \mathbb{R}^k$ represent parameters. Moreover, let us use the logistic model to represent the probability that a insuree with characteristics x does not meet an unmatched broker, that is

$$\mu(x) = \frac{\exp(m_0 + x \cdot m)}{1 + \exp(m_0 + x \cdot m)}$$

where $m_0 \in \mathbb{R}$ are $m = (m_1, ..., m_k) \in \mathbb{R}^k$ are parameters.

We are also informed about which insurance company has issued each policy. Let z be a k'-dimensional the random variable representing this information. We use a mixed logit model to represent the conditional probability of each percentage cap $\tau \in \{.15, .20, .25, .30, .35, .42\}$, that is

$$\varphi(\tau \mid x, z) = \frac{\exp\left(\xi_\tau + x \cdot \beta + z \cdot \phi\right)}{\sum_\tau \exp\left(\xi_\tau + x \cdot \beta + z \cdot \phi\right)},$$

where $\beta = (\beta_1, ..., \beta_k) \in \mathbb{R}^k$ and $\phi = (\phi_1, ..., \phi_{k'}) \in \mathbb{R}^{k'}$ are conditional parameter vectors, and ξ_τ is a multinomial parameter associated with equation τ. Like before, we normalize $\xi_{.42} = 0$ to make the system determined.

After trying a few alternatives, we selected age and gender as the individual characteristics in x, and we let z be given by dummy variables for the four largest insurance companies (which account for about 88 percent of the policies in our sample). We estimate this model and perform a likelihood ratio test. We do not reject the null hypothesis that α, m, β, and ϕ are simultaneously zero at the usual 5-percent level of significance.

It is worth emphasizing that our conditional model does not allow for random effects. This would require us to specify a probability distribution for the random coefficients and then use simulation methods to integrate the likelihood function over that distribution. This procedure may smooth out the likelihood function at the cost of increasing the computational time required to evaluate it. The results are likely to depend on the distributional assumptions.

Bayesian methods provide alternatives for modeling randomness of the structural parameters. We refer the reader to Chernozhukov and Hong [2004] for a methodological discussion on this topic. Random effects and Bayesian methods are potentially interesting extensions to be considered in the future.

8 Structural Predictions and Counterfactual

We start our structural analysis by computing the model prediction for the average frequency of switching. According to our model, a consumer switches the broker only if the equilibrium is in mixed strategies. Thus, when $\frac{\mu r_{j,\tau}}{1-\delta\mu} > c$, switching occurs with probability

$$\gamma\left(b_j,\tau,\delta,\mu,c\right)=(1-\mu)\int_{-a}^{r-c}\left(1-\sigma_m(q+c)\right)d\sigma_u(q).$$

Moreover, we have $\gamma\left(b_j,\tau,\delta,\mu,c\right)=0$ when $\frac{\mu r_{j,\tau}}{1-\delta\mu}\leqslant c$ (pure-strategy equilibrium).

We average this probability over the estimated frequency of τ and the sample values for b_j to define

$$\bar{\gamma}=\sum_{\tau}\sum_{j}\frac{1}{N}\varphi(\tau\mid x_j)\gamma\left(b_j,\tau,\delta,\mu,c\right),$$

where N is the sample size. Given our estimated parameters, insurees should switch brokers with probability $\bar{\gamma}=17.17\%$. This prediction is close to the number of policies sold without brokerage fees which, according to our model, must have been issued by unmatched brokers. According to a recent survey conducted by the Brazilian association of insurance brokers, FENACOR [2013], about 82 percent of the brokers in Brazil report that they typically retain over 80 percent of their current customers and only 5 percent of them report a retention rate smaller than 70 percent. This anecdotal evidence suggests that the magnitude of our model prediction is reasonable.

8.1 Lifetime Brokerage Value

The model also allows us to compute the expected present value of dealing with a consumer under different circumstances. This is useful for analyzing mergers and acquisitions, as it allows one to access the value of a portfolio with regular (matched) customers and a stable flow of potential new clients.

Recal that, when $\frac{\mu r_{j,\tau}}{1-\delta\mu}\leqslant c$ (pure-strategy equilibrium), the lifetime value of dealing with a matched insuree is c. Moreover, when $\frac{\mu r_{j,\tau}}{1-\delta\mu}>c$ (mixed-strategy equilibrium), this lifetime value becomes $\bar{v}_{m,j}(\tau):=\frac{\mu r_{j,\tau}}{1-\delta\mu}$, for each given insurees with observables (x_j,b_j) and environment with percentage ceiling τ. When we average these values over all insurees j in the population and across the different percentage caps τ, we obtain an average lifetime value of 790.67 BRL (i.e., approximately 395.34 PPP-adjusted USD).

We can also compute the value of a client asking for a price quotation. This is zero in the pure-strategy equilibrium and $E(\bar{v}_u) = E(\bar{v}_m) - c$ when $\frac{\mu r_{j,\tau}}{1-\delta\mu} > c$. We find that the average lifetime value of dealing with a new customer is 577.32 BRL (i.e., approximately 288.66 PPP-adjusted USD).

8.2 Efficient Price Ceiling

Consumer surplus is maximized in our model when price goes to zero. The total surplus, however, only depends on the costs of unnecessary switching. As a consequence, Pareto efficiency is attained when no insuree switches brokers. This can be obtained by setting ceilings to be at most $\frac{(1-\delta\mu)c}{\mu}$ and implementing a pure-strategy equilibrium.

In this case, the matched broker offers an equilibrium price of $(1-\delta)c$ and always sell the policy. According to our estimated parameters, this amounts to having a fee ceiling of 63.86 BRL (equivalent to 31.93 PPP-adjusted USD) and an equilibrium price of 51.01 BRL (about 25.51 PPP-adjusted USD). For matched brokers, the (per insuree) lifetime earnings would drop from 783.90 BRL to 213.36 BRL. Unmatched brokers would make zero.

9 Conclusion

We documented a few stylized facts from the Brazilian market for auto insurance brokerage. We found policies being sold with zero and positive brokerage commissions. The distribution of positive fees is smooth and exhibits a wide dispersion. We also registered that insurance companies impose ceilings on brokerage fees which are expressed as fractions of the policy final price. The data exhibit mass of policies being sold at each of these percentage caps.

We rationalize this evidence through a dynamic model of price competition with search frictions, in the spirit of Varian [1980], and switching costs. In our model, brokers face the following conflict: on the one hand, they are tempted to set high fees and extract rents; on the other hand, they are also interested in setting low fees to either attract or retain the customers who are looking for a better price. In this context, a mixed-strategy equilibrium emerges as the natural way to balance these opposite forces. Our equilibrium mixed strategy generates: (a) loss leading that we claim to be behind the recorded zero fees; (b) a spread-out distribution of nominal fees; and (c) mass points at the maximum fees allowed by the insurance company.

Our model is very parsimonious and has only three structural parameters. Using data on prices, we estimated these parameters, predicted the equilibrium fees with a pretty good accuracy, and computed the distribution of switching costs and the brokerage lifetime expected profits.

References

Jason Allen, Robert Clark, and Jean-François Houde. Price dispersion in mortgage markets. *Journal of Industrial Economics*, 62(3):377–416, 2014a.

Jason Allen, Robert Clark, and Jean-François Houde. Search frictions and market power in negotiated price markets. *NBER Working Paper 19883*, 2014b.

Eric T. Anderson, Nanda Kumar, and Surendra Rajiv. A comment on: "Revisiting dynamic duopoly with consumer switching costs". *Journal of Economic Theory*, 116(1):177–186, 2004.

Alan Beggs and Paul Klemperer. Multi-product competition with switching costs. *Econometrica*, 60(3):651–666, 1992.

Luis H.B. Braido. Multiproduct price competition with heterogeneous consumers and nonconvex costs. *Journal of Mathematical Economics*, 45(9-10):526–534, 2009.

Jeffrey Brown and Austan Goolsbee. Does the internet make markets more competitive? evidence from the life insurance industry. *Journal of Political Economy*, 110(3):481–507, 2002.

Yongmin Chen. Paying customers to switch. *Journal of Economics and Management Strategy*, 6(4):877–897, 1997.

Yongmin Chen and Robert W. Rosenthal. Dynamic duopoly with slowly changing customer loyalties. *International Journal of Industrial Organization*, 14(3):269–296, 1996.

Victor Chernozhukov and Han Hong. Likelihood estimation and inference in a class of nonregular econometric models. *Econometrica*, 72(5):1445–1480, 2004.

Bev Dahlby and Douglas S. West. Price dispersion in an automobile insurance market. *Journal of Political Economy*, 94(2):418–438, 1986.

Babur De Los Santos, Ali Hortaçsu, and Matthijs R. Wildenbeest. Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review*, 102(6):2955–2980, 2012.

Babur De Los Santos, Ali Hortaçsu, and Matthijs R. Wildenbeest. Search with learning for differentiated products: evidence from e-commerce. *Journal of Business & Economic Statistics*, 35(4):626–641, 2017.

Jean-Pierre Dubé, Gunter J. Hitsh, and Peter E. Rossi. Do switching costs make markets less competitive? *Journal of Marketing Research*, 46(4):435–445, 2009.

Darrell Duffie, John D. Geanakoplos, Andreu Mas-Colell, and Andrew M. McLennan. Stationary markov equilibria. *Econometrica*, 62(4):745–781, 1994.

Joseph Farrell and Paul Klemperer. Coordination and lock-in: Competition with switching costs and network effects. In M. Armstrong and R. Porter, editors, *Handbook of Industrial Organization*, volume 3. Amsterdam: North-Holland, 2007.

Joseph Farrell and Carl Shapiro. Dynamic competition with switching costs. *RAND Journal of Economics*, 19(1):123–137, 1988.

FENACOR. Estudo socioeconômico das empresas corretoras de seguros, 2013.

Drew Fudenberg and Jean Tirole. Customer poaching and brand switching. *RAND Journal of Economics*, 31(4):634–657, 2000.

Han Hong and Matthew Shum. Using price distributions to estimate search costs. *RAND Journal of Economics*, 37(2):257–275, 2006.

Elisabeth Honka. Quantifying search and switching costs in the us auto insurance industry. *RAND Journal of Economics*, 45(4):847–884, 2014.

Ali Hortaçsu and Chad Syverson. Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds. *Quarterly Journal of Economics*, 119(2):403–456, 2004.

Mark Israel. Tenure dependence in consumer-firm relationships: An empirical analysis of consumer departures from automobile insurance firms. *RAND Journal of Economics*, 36(1):165–192, 2005.

Paul Klemperer. Markets with consumer switching costs. *Quarterly Journal of Economics*, 102(2):375–394, 1987.

Saul Lach. Existence and persistence of price dispersion: An empirical analysis. *Review of Economics and Statistics*, 84(3):433–444, 2002.

William Q. Meeker and Luis A. Escobar. Maximum likelihood methods for fitting parametric statistical models to censored and truncated data. In J. Stanford and S. Vardeman, editors, *Probabilistic and Statistical Methods in the Physical Sciences*. New York: Academic Press, 1994.

A. Jorge Padilla. Revisiting dynamic duopoly with consumer switching costs. *Journal of Economic Theory*, 67(2):520–530, 1995.

Willian W. Sharkey and David S. Sibley. A bertrand model of pricing and entry. *Economic Letters*, 41(2):199–206, 1993.

Alan T. Sorensen. Equilibrium pice dispersion in retail markets for prescription drugs. *Journal of Political Economy*, 108(4):833–850, 2000.

Dale O. Stahl. Oligopolistic pricing with sequential consumer search. *American Economic Review*, 79(4):700–712, 1989.

George J. Stigler. The economics of information. *Journal of Political Economy*, 69 (2):13–25, 1961.

Curtis R. Taylor. Supplier surfing: Competition and consumer behavior in subscription markets. *RAND Journal of Economics*, 34(2):223–246, 2003.

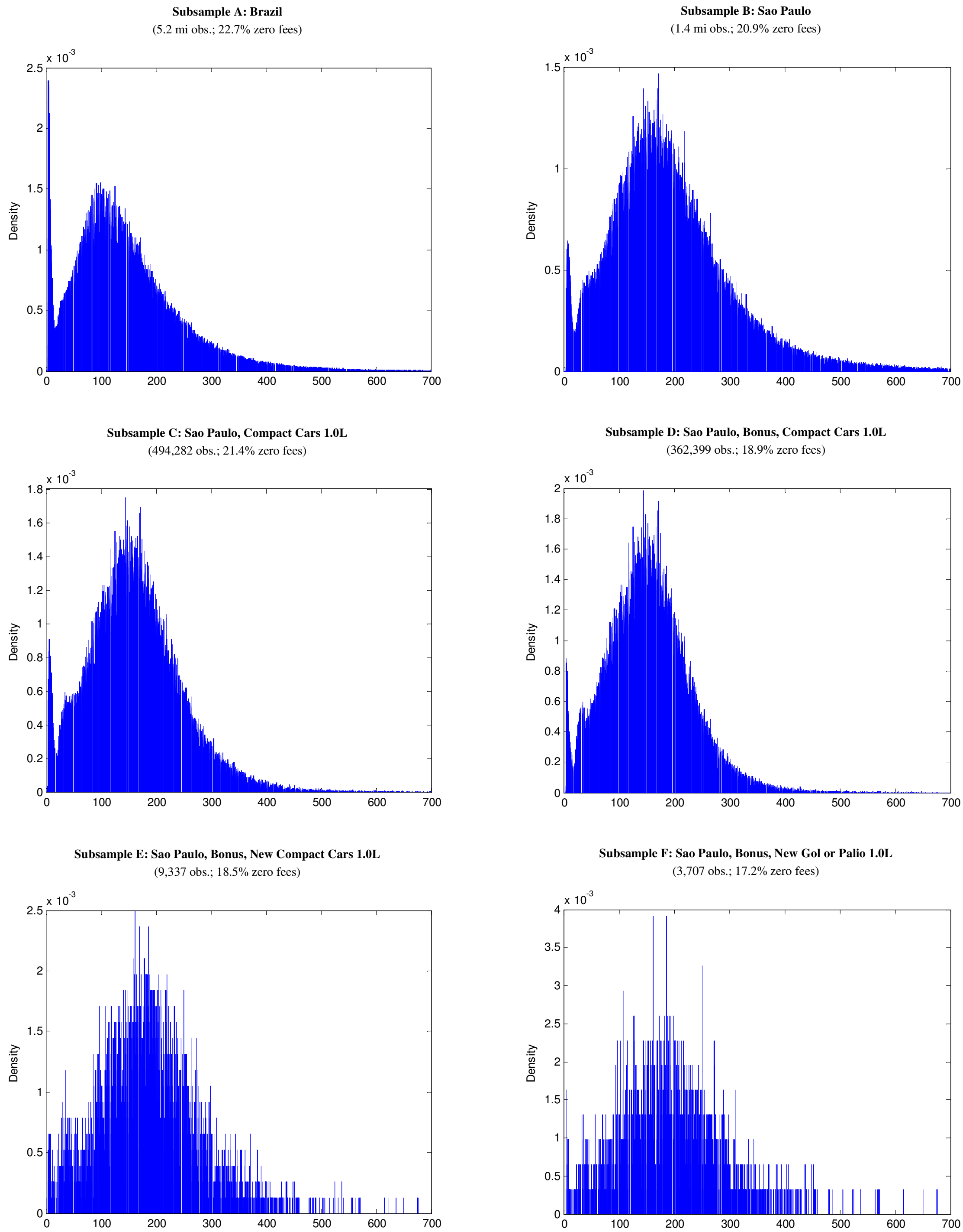
Hal R. Varian. A model of sales. *American Economic Review*, 70(4):651–659, 1980.

J. Miguel Villas-Boas. Dynamic competition with experience goods. *Journal of Economics and Management Strategy*, 15(1):37–66, 2006.

Matthijs R. Wildenbeest. An empirical model of search with vertically differentiated products. *RAND Journal of Economics*, 42(4):729–757, 2011.

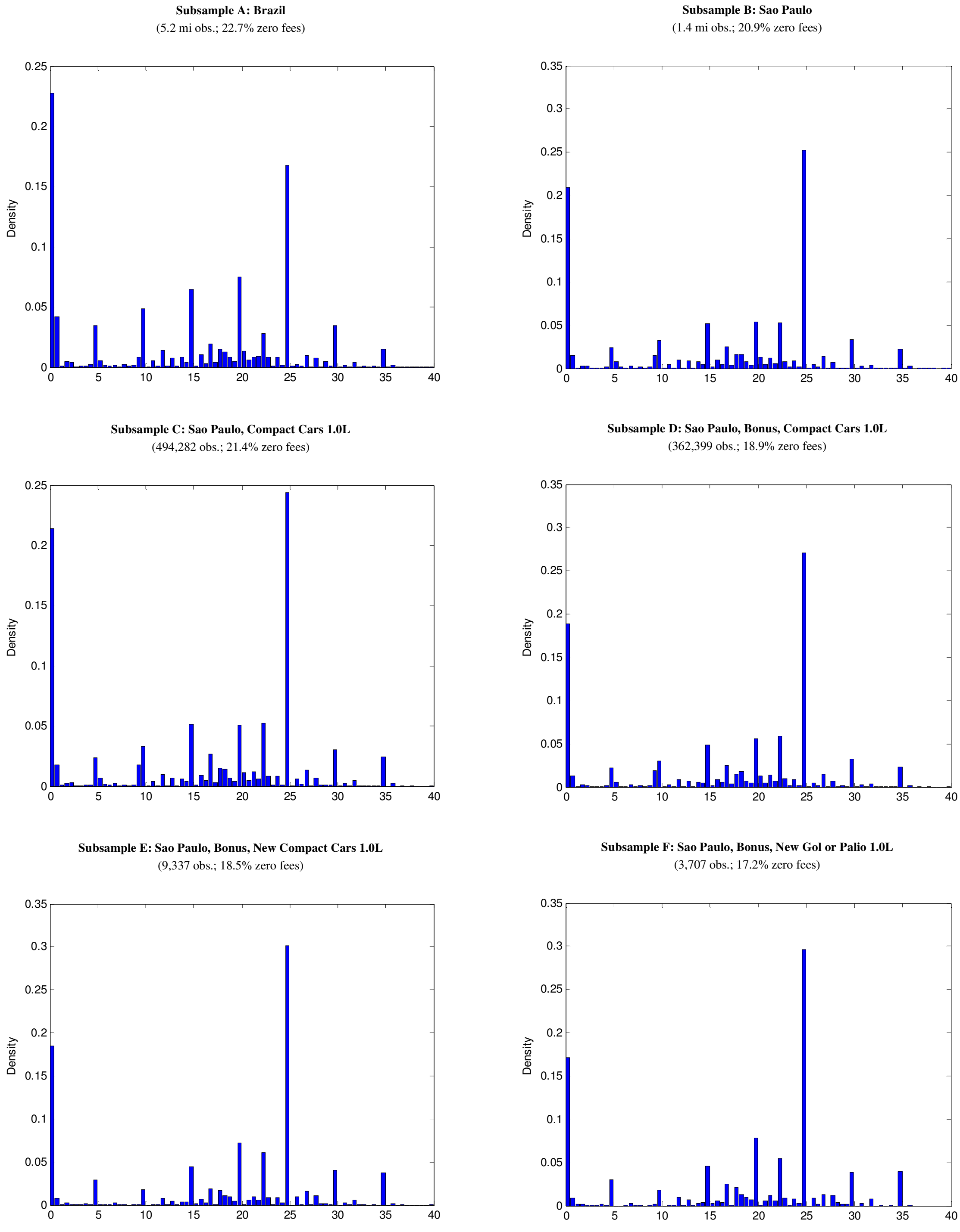
Matthijs R. Wildenbeest and José L. Moraga-González. Maximum likelihood estimation of search costs. *European Economic Review*, 52(5):820–848, 2008.

Figure 1. Nonzero Brokerage Fees (in BRL)



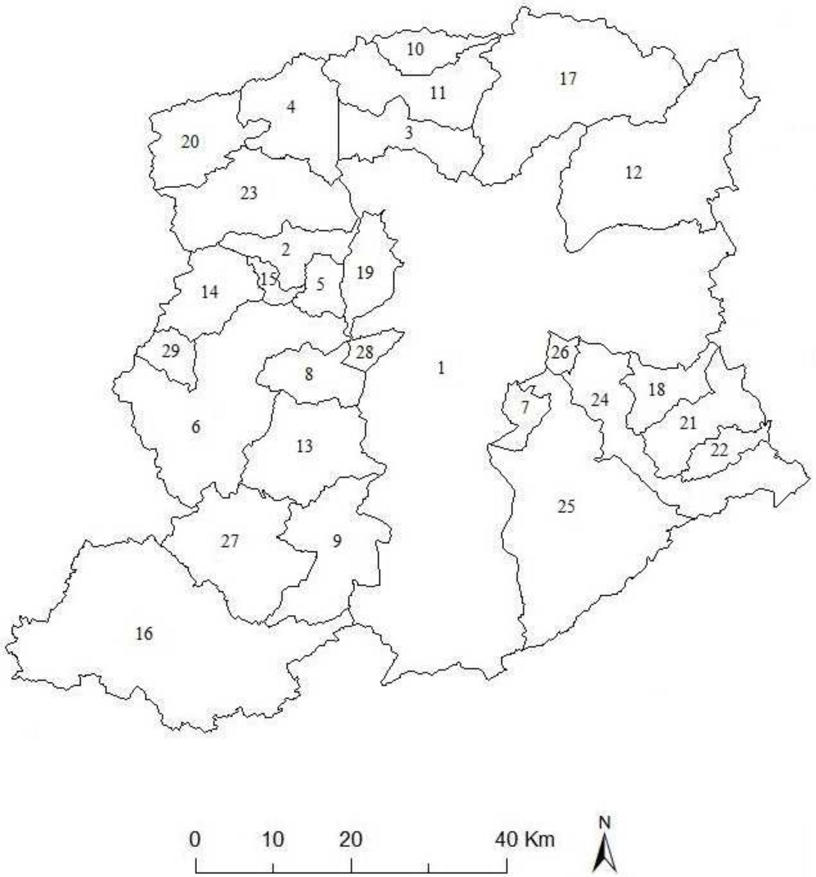
Histogram Width: 0.25 BRL. To better adjust the graph scales, we do not include policies without commission (very frequent, as shown in Table 2) and those with fees above 700 BRL (very infrequent).

Figure 2. Broker Percentage Fees



Histogram Width: 0.5 percentage point. All observations included.

Figure 3. Metropolitan Area of Sao Paulo



	City	ZIP Codes
1	São Paulo	01000 to 05899
2	Barueri	06400 to 06499
3	Caieiras	07700 to 07749
4	Cajamar	07750 to 07759
5	Carapicuíba	06300 to 06399
6	Cotia	06700 to 06729
7	Diadema	09900 to 09999
8	Embu	06800 to 06849
9	Embu-Guaçu	06900 to 06949
10	Francisco Morato	07900 to 07999
11	Franco da Rocha	07800 to 07870
12	Guarulhos	07000 to 07399
13	Itapecerica da Serra	06850 to 06889
14	Itapevi	06650 to 06699
15	Jandira	06600 to 06649
16	Juquitiba	06950 to 06999
17	Mairiporã	07600 to 07699
18	Mauá	09300 to 09399
19	Osasco	06000 to 06299
20	Pirapora do Bom Jesus	06550 to 06599
21	Ribeirão Pires	09400 to 09449
22	Rio Grande da Serra	09450 to 09499
23	Santana de Parnaíba	06500 to 06549
24	Santo André	09000 to 09299
25	São Bernardo do Campo	09600 to 09899
26	São Caetano do Sul	09500 to 09599
27	São Lourenço da Serra	06890 to 06899
28	Taboão da Serra	06750 to 06799
29	Vargem Grande Paulista	06730 to 06749

Figure 4. City of Sao Paulo

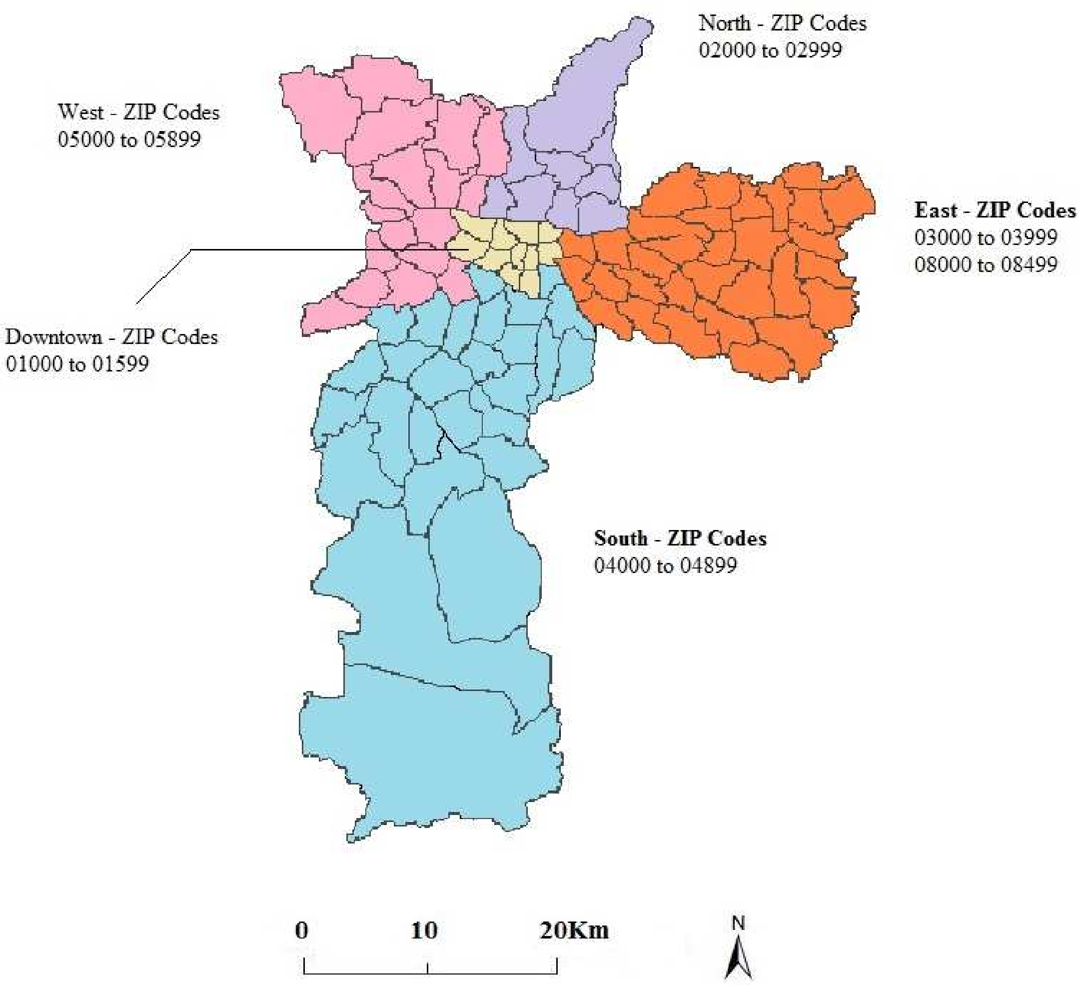
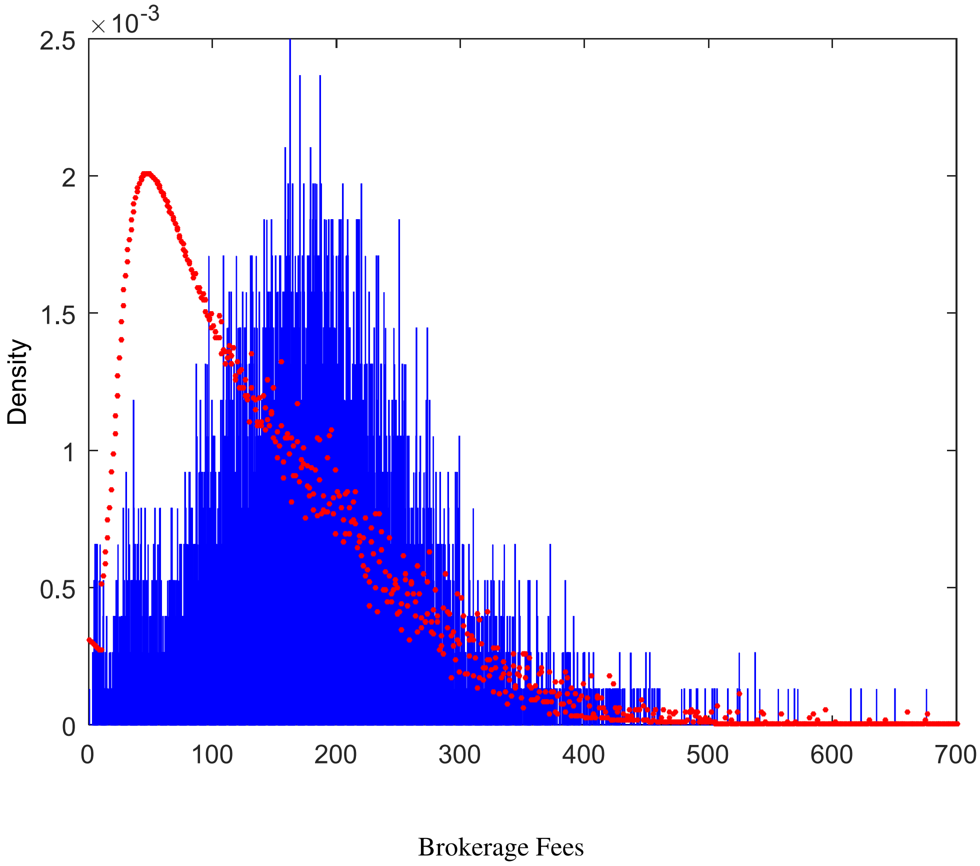


Figure 5. Empirical and Estimated Density Functions of Brokerage Fees



Dots for the estimated density function; bars for the empirical histogram (histogram width: 0.25 BRL). To better adjust the graph scales, we do not include policies without commission (about 18.46%) and those with fees above 700 BRL (very infrequent). The estimated fraction of zero fees is 14.86%.

Table 1. SUSEP/AUTOSEG Auto Insurance Data

Variable	Description
Total Insurance Premium	Total premium paid by the insuree to the insurance company (sum of the variables pre_casco, pre_rcdm, pre_app, pre_rcdp, and pre_outros).
Broker Percentage Fee	Fraction of the total premium that is transferred from the insurance company to the broker (perc_corr).
Brokerage Fee	Product of the Broker Percentage Fee and the Total Insurance Premium.
Baseline Insurance Price	Value set and retained by the insurance company. Total premium minus brokerage fee.
Vehicle Type	Alpha-numerical variable describing the make, model, and power train of the insured vehicle (cod_modelo). For the main estimation, we restrict attention to nine models of compact cars with 1.0L engines.
Vehicle Model Year	Year of the vehicle model (ano_modelo). When performing the estimations, we restrict attention to new vehicles (model year 2003).
Geographic Region	Numerical variable ranging from 1 to 41 that describes the geographic region of the insured policy (regiao). For our estimations, we restrict the sample to the metropolitan region of Sao Paulo (code 11).
Insurance Company	Numerical code for the insurance company (cod_seg).
Coverage Type	Qualitative variable describing the type of insurance coverage for the insured vehicle (cobertura). Code 1 is used for comprehensive coverage; 2 for coverage against fire and theft; 3 for fire only; 4 for write-off, collision, and theft; 9 for others. We restrict attention to contracts with comprehensive coverage (code 1).
Deductible Type	Qualitative variable describing the level of deductible for physical damages to the insured vehicle (tipo_franq). Codes 1, 2 and 3 indicate low, regular, and high deductible levels, respectively. Code 9 indicates no deductible.
Physical Damage Coverage	Coverage limits for damages in the insured vehicle (is_casco) and in third-party vehicles (is_rcdm). Values expressed in BRL.
Personal Injury Coverage	Coverage limits for personal injuries caused to passengers of the insured vehicle (is_app) and to passengers of third-party vehicles and pedestrians (is_rcdp). Values expressed in BRL. This includes medical expenses and death/disability benefits in extension of the mandatory coverage.
Bonus Category	Bonus categories summarizing the insuree's record of claims (perc_bonus). Available for all companies.
Gender	Qualitative variable describing the gender of the vehicle's primary driver, as declared in the contract (sexo). The code "M" is used for male; "F" for female; and "J" for corporate vehicles. We exclude policies with code "J".
Date of Birth	Primary driver's date of birth as stated in the contract (data_nasc). The value "00000000" is used for corporate vehicles.
Zip Code	Zip code of the insuree address as registered in the contract (cep).

Table 2. Reduced-Form OLS Regressions

	Premium	Broker Percentage Fee	Brokerage Fee	Zero-Fee Dummy
Physical Damage Coverage Limits				
Insured Vehicle (is_casco)	0.019***	0.000	0.004***	-0.000
Third Parties (is_rcdm)	0.002***	0.000	0.001***	-0.000
Personal Injury Coverage Limits				
Insured Vehicle (is_app)	0.001***	0.000	0.001***	0.000
Third Parties (is_rcdp)	0.000	-0.000	-0.000	0.000
Low Deductible Dummy	-80.089***	0.676*	-9.104***	-0.020
Insuree Gender (male = 1)	40.516***	-0.310	6.167***	0.006
Insuree Age	-9.746***	-0.007	-2.816***	0.003
Insuree Age^2	0.055**	0.000	0.016	-0.000
Age Groups	6 categories	6 categories	6 categories	6 categories
Insuree Bonus Discount	8 categories	8 categories	8 categories	8 categories
Vehicle Model	8 categories	8 categories	8 categories	8 categories
Insurance Company	15 categories	15 categories	15 categories	15 categories
3-Digit Zip Code	135 categories	135 categories	135 categories	135 categories
Sample Size	9,337	9,337	9,337	9,337
R^2	0.49	0.05	0.13	0.15

* significant at 10%; ** significant at 5%; *** significant at 1% (with robust standard errors).

Table 3. Structural Maximum Likelihood Estimation

	Estimate	95% Confidence Interval	
		Lower Bound	Upper Bound
Structural Parameters			
Discount Factor (δ)	0.9527	0.9046	0.9924
Probability of Accessing Two Brokers (1-μ)	0.2013	0.1851	0.2144
Switching Cost (c)	213.36	184.56	255.52
Multinomial Logit Parameters			
$\xi_{0.15}$	2.4823	2.1659	3.9276
$\xi_{0.20}$	3.1273	2.7659	4.5247
$\xi_{0.25}$	4.6801	4.2901	6.0720
$\xi_{0.30}$	2.9586	2.5813	4.3837
$\xi_{0.35}$	2.9765	2.6397	4.3381
Sample Size	9,337	From 1,000 subsamples with 3,112 observations randomly drawn with replacement	

Table 4. Mass Points

τ	**Empirical Frequency**	**Model Prediction**
0%	18.46%	14.86%
15%	4.40%	3.05%
20%	7.22%	6.88%
25%	30.05%	34.37%
30%	4.07%	6.32%
35%	3.76%	6.54%
42%	0.01%	0.34%