Competing under Information Heterogeneity: Evidence from Auto Insurance*

Marco Cosconati, IVASS
Yi Xin, Caltech
Fan Wu, Caltech
Yizhou Jin, Toronto

November 18, 2024

Abstract

This paper studies competition under information heterogeneity in selection markets and examines the impact of public information regulations aimed at reducing information asymmetries between competing firms. We develop a novel model and introduce new empirical strategies to analyze imperfect competition in markets where firms have *heterogeneous information* about consumers, vary in cost structures, and offer differentiated products. Using data from the Italian auto insurance market, we find substantial differences in the precision of risk ratings across insurers, and those with less accurate risk-rating algorithms tend to have more efficient cost structures. We assess the equilibrium effects of giving firms equal access to aggregated risk information from a centralized bureau. This policy significantly reduces prices by increasing competition, leading to a 13% boost in consumer surplus, nearly reaching the efficiency benchmark where firms have full knowledge of consumers' true risk. Aggregating information through the bureau favors low-risk consumers and reduces average costs by 18 euros per contract through more efficient insurer-insuree matching.

*Corresponding author: Yi Xin (email: yixin@caltech.edu). Cosconati and Xin are co-first authors, listed alphabetically. Cosconati: IVASS and Bank of Italy, via del Quirinale, 21, 00187, Roma. Xin: Division of the Humanities and Social Sciences, California Institute of Technology, 1200 East California Blvd, MC 228-77, Pasadena, CA 91125. Wu: Division of the Humanities and Social Sciences, California Institute of Technology, 1200 East California Blvd, MC 228-77, Pasadena, CA 91125. Jin: Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario M5S 3E6.

We appreciate valuable discussions with José Ignacio Cuesta, Liran Einav, Hanming Fang, Jeremy Fox, Jean-François Houde, Ginger Jin, Elena Krasnokutskaya, Fei Li, Yao Luo, Matthew Shum, Paulo Somaini, Andrew Sweeting, Xun Tang, Pietro Tebaldi, Tiffany Tsai, and Mo Xiao. We thank seminar and conference participants at Caltech, CUHK, JHU, HKU, HKUST, NTU, UC Davis, University of Indiana at Bloomington, University of Maryland, UNC-Chapel Hill, University of Pennsylvania, SITE (2024), SHUFE IO Conference (2024), SoCal Structural Econometrics Conference (2024). Financial support from the Ronald and Maxine Linde Institute of Economic and Management Sciences are gratefully acknowledged. The opinions and errors in this paper are solely those of the authors and do not reflect the views of IVASS and the Bank of Italy.

Keywords: Heterogeneous Information, Imperfect Competition, Information Regulation, Selection Markets, Auto Insurance.
JEL Codes: D82, D43, L13, L51, G22.

1 Introduction

As data analytics and technology advance, the gap between firms in both information access and data analysis capabilities continues to widen. Recent evidence suggests that banks and credit card lenders use different screening technologies to assess default and pre-payment risks (Grodzicki, 2023; Matcham, 2023; Becker et al., 2024; Blickle et al., 2024); auto insurance companies assess driver risk using different types of information, ranging from basic factors like age to advanced telematic data on real-time driving behavior.[1] In addition to the well-known asymmetry between buyers and sellers since the seminal work of Akerlof (1970) and Rothschild and Stiglitz (1976), this phenomenon introduces another layer of complexity: information asymmetries between competing firms. However, how these between-firm information asymmetries impact equilibrium pricing strategies, consumer choices, and market efficiency remains an open question in the literature.

On the policy side, improving information availability and reducing information gap between firms have become a focal point in recent discussions. For example, data-sharing policies like the UK's open banking initiative facilitate the flow of data between financial institutions. Similar debates are unfolding in Italy, where the insurance industry is advocating for more detailed consumer risk information to be publicly disclosed.[2] However, reducing information asymmetries between competing firms goes beyond simply sharing data. It requires ensuring that all participants have the same analytical capabilities to interpret and effectively use the information—a process that can be both costly and challenging to implement. An alternative approach is to establish a centralized bureau that collects and aggregates analyzed data (such as insurers' risk estimates) and makes this information equally accessible to all.

[1] Information asymmetries between firms have been documented in other markets as well. For example, Boomhower et al. (2024) show that certain homeowner insurance companies have access to finer measures of wildfire risk. In oil and gas lease auctions, "neighbour firms" are better informed about lease values than "non-neighbour firms" (Hendricks et al., 1987; Hendricks and Porter, 1988; Hendricks et al., 1994; Porter, 1995). E-commerce platform owners has an information advantage over third-party sellers (Chen and Tsai, 2023). Schools that conduct personal interviews gain deeper insights into students' unobserved talents (Friedrich et al., 2023).

[2] A recent update on Open Banking initiatives can be found on the Competition and Markets Authority website. Also see discussions from the the National Association of Insurance Companies (ANIA) at the IVASS workshop (web link).

Our paper seeks to address two critical questions: first, the fundamental question of how information asymmetries between firms shape market equilibrium, and second, from a policy perspective, whether establishing a centralized bureau to equalize information access can effectively promote competition, enhance consumer welfare, and improve overall market efficiency.

To address these questions, we develop and estimate a novel empirical model of imperfect competition in selection markets, where firms have *heterogeneous information* about consumers, differ in cost structures, and offer differentiated products. Our analysis is based on a unique market-level dataset from the Italian auto insurance industry. We find substantial differences in the precision of risk-rating across insurers. Insurers with more accurate risk-rating algorithms can cream-skim lower-risk consumers, but they often have less efficient cost structures. Equalizing information access through a centralized bureau significantly lowers market prices by increasing competition. This policy boosts consumer surplus by 13%, nearly reaching the efficiency benchmark where firms have perfect knowledge of consumers' true risk, and reduces average costs by 18 euros per contract through more efficient insurer-insueree matching.

Our study focuses on the Italian auto insurance industry for several key reasons. First, we use a representative sample of auto insurance contracts from *all* insurers in Italy between 2013 and 2021. The data include individual-level demographic information, vehicle characteristics, contract details, transaction prices, and claim records, with consumers tracked even if they switch insurers. Second, liability insurance in Italy is mandatory for all drivers, and rejections are prohibited by law, allowing us to focus on how consumers are sorted into different insurers within the market. Third, while insurers share similar contract features, they employ notably different pricing strategies. Survey evidence suggests that insurers use heterogeneous information in their actuarial models, and our regression analysis further reveals significant variation in the extent to which prices are based on commonly observed risk factors across insurers.

We begin our analysis by investigating whether insurers differ in the precision of their risk assessments. Specifically, we analyze the correlation between each firm's premiums and consumer risk, which is *ex-ante* unknown to insurers, but *ex-post* estimable based on realized claim data.[3] Our analysis demonstrates that certain firms' premiums are more responsive to, or more accurately reflect, true risk, indicating that these firms are potentially

[3]Using a long panel of claim records, we estimate an individual-specific risk measure. Similarly, Porter (1995) estimates the ex-post value of oil leases (which is ex-ante unknown to bidders) based on amounts of extracted oil, industry costs, and the price of oil, etc. Biglaiser et al. (2020) use the ex-post likelihood of resale as a measure of the quality of used cars and they find that dealers have better information about car quality.

more informed and better at assessing consumer risk.[4]

Motivated by these empirical findings, we develop a static model of price competition among insurers. Each firm draws a private signal about the consumer's risk type and infers the expected cost to insure the consumer based on the signal. The dispersion of the signal distribution is determined by the firm's information precision. Firms' pricing strategies are functions of their risk evaluations. In equilibrium, firms optimally choose pricing coefficients to maximize their profits, taking into account their opponents' strategies and demand-side responses. Our model allows for rich and flexible supply-side heterogeneity: insurers can differ in their information precision, benefits of contracting with a customer, efficiency at processing claims, and unobserved product attributes (e.g., quality and brand loyalty). The demand side of our model follows the standard differentiated product framework as in Berry (1994) and Berry et al. (1995).

Our paper introduces novel econometric techniques to identify and estimate demand parameters when only transaction prices are available. We leverage the relationship between observed transaction prices and the likelihood of consumers selecting those prices, as governed by the demand model, to infer the underlying distribution of offered prices. Importantly, we impose no parametric restrictions on the offered price distributions and allow them to vary fully across firms, which are essential for analyzing supply-side heterogeneity. Our estimation strategy also allows consumer preference parameters to vary with observable characteristics and risk type, making it applicable to a broad range of empirical settings.

The firm-specific distribution of offered prices, recovered from the demand side, plays an important role in identifying the signal distribution. The key insight is that the offered price monotonically increases with the firm's signal, similar to auction models in which bids monotonically increase in bidders' valuations. As a result, the dispersion in price distributions *within firms* helps identify signal variance. Finally, we estimate firms' cost parameters using the first-order conditions derived from their profit maximization problem.

We apply the model and the estimation strategy to the Italian auto insurance industry, focusing on a sample of new customers in Rome.[5] Our structural estimation suggests that there is a significant amount of firm heterogeneity along *all* dimensions. In particular, we find substantial differences in the precision of risk rating across insurers, and the ones with lower information precision tend to attract more high-risk customers.[6] We also find

[4]A similar intuition appeared in Panetta et al. (2009), who argue that banks with superior screening abilities should charge an interest rate that is more "sensitive" to the firm's default risk.

[5]Focusing on new customers is common in empirical analysis of asymmetric information to avoid complicated issues of dynamic pricing, and consumer/firm learning. See the discussions in Chiappori and Salanie (2000) and Crawford et al. (2018).

[6]Relatedly, Einav et al. (2013) document that high-performing dealers are better at assessing borrowers'

that insurers enjoy large and different net benefits from contracting with new customers (potentially due to inertia and dynamic pricing strategies) ranging from 355 to 1052 euros per contract, and their efficiency at processing claims also differs significantly. Another interesting finding to highlight is that insurers with less accurate risk-rating algorithms tend to have more efficient cost structures, leading to distinct types of comparative advantages.

Using our estimates, we simulate a counterfactual scenario where all firms have equal access to consumer information through a centralized risk bureau. This bureau collects signals from all firms (i.e., their analyzed data), aggregates them based on each firm's information precision, and then makes the combined information equally accessible to all. As an efficiency benchmark, we also consider a hypothetical scenario where all firms fully observe each consumer's true risk type, thereby eliminating information asymmetry completely.

We observe a substantial reduction in average premiums, ranging from 16.73% when a centralized risk bureau is introduced to 19.20% when firms can observe consumers' true risk levels. This reduction in premiums is driven by two key factors. First, eliminating heterogeneity in information precision weakens the market power of firms with superior information, forcing them to compete more aggressively on price. The second, and more intriguing, factor is the elimination of the winner's curse. When firms share the same evaluation of consumer risk, they can undercut each other's prices more effectively, further intensifying competition. Unsurprisingly, the reduction in premiums leads to an overall decline in average profits for firms, though the impact varies across firms. Those utilizing more advanced risk-rating technologies experience greater losses, which, in our case, may include smaller and more specialized firms.

Due to the reduction in prices, we find that establishing a centralized risk bureau boosts consumer surplus by 13.15%, nearly matching the 14.62% increase observed in the efficiency benchmark, where firms observe consumers' true risk. With better information provided by the bureau, firms can more accurately price discriminate based on risk type. As a result, this policy primarily benefits low-risk consumers, whose surplus increases by 63.27%, while high-risk consumers experience a 27.52% reduction in their surplus.[7]

Eliminating information heterogeneity also significantly impacts consumer sorting patterns, which critically affects market efficiency. We find that when insurers have equal access to more accurate information about consumers' risk distributed by the centralized

risks in auto loans.

[7]In an alternative counterfactual motivated by privacy regulations, we assume that firms are restricted to basic consumer information and set their signal variances to the highest level in the market. This results in opposite distributional effects, where high-risk consumers see a 7.76% increase in surplus as they become harder to distinguish from low-risk consumers.

bureau, those that are more efficient at processing claims adjust their pricing strategies to target high-risk consumers. As a result, consumers' risk sorting patterns shift significantly from the baseline, now driven by specialization based on cost advantages rather than information asymmetries. The new sorting pattern leads to a more efficient match between insurers and insurees, reducing the average cost by 2.56% (approximately 18 euros per contract). The efficiency gains under the centralized risk bureau scenario closely approach the infeasible efficiency benchmark, highlighting the effectiveness of this information policy. This also suggests that better utilization and more efficient aggregation of existing market information could significantly improve market outcomes.

Related Literature Our paper contributes to the large empirical literature on selection markets (for an excellent review, see Einav et al. (2021)). In particular, our paper develops the first tractable empirical framework for analyzing imperfect competition when firms have heterogeneous information about consumers. The seminal works of Einav et al. (2010) and Einav and Finkelstein (2011) laid the foundation for analyzing perfectly competitive selection markets. Mahoney and Weyl (2017) present a theoretical model of symmetric oligopoly competition. Azevedo and Gottlieb (2017) develop a competitive model of adverse selection that allows for endogenously determined contract characteristics and multidimensional private information. Recent empirical studies focus on the interaction between asymmetric information and market power using data from credit and health insurance markets (e.g., Cabral et al., 2018; Crawford et al., 2018; Nelson, 2018; Decarolis et al., 2020; Jaffe and Shepard, 2020; Curto et al., 2021; Cuesta and Sepúlveda, 2021; Tebaldi, 2024), all under the assumption that information is symmetrically distributed among competing firms.[8] Importantly, our paper also incorporates multidimensional cost heterogeneity on the supply side (Salanié, 2017; Serna, 2023; Nelson, 2018) and studies how it interacts with heterogeneous information.

From a methodological point of view, our paper extends the classic demand estimation techniques by introducing a "selection loop" on top of the contraction mapping for mean utilities pioneered by Berry (1994) and Berry et al. (1995). We address the common issue of missing full price menus and impose no parametric restrictions or symmetry assumptions on price distributions across firms. More theoretical results and technical details of our approach can be found in Wu and Xin (2024). On the supply side, we identify and estimate

[8]The health insurance context differs significantly from the auto insurance industry we study, particularly in terms of regulatory frameworks. For example, in the US the Affordable Care Act prevents insurers from denying coverage or charging higher premiums based on preexisting conditions. It also establishes risk adjustment, reinsurance, and risk corridor programs to encourage competition among insurers to provide high-quality services at low cost and to mitigate risk selection. For further details and other related work, see the recent review of health insurance markets by Handel and Ho (2021).

information precision for competing firms leveraging the one-to-one mapping between the offered price and the signal, which borrows ideas from the empirical nonparametric auction literature (Guerre et al., 2000).

Lastly, our paper also contributes to the recent discussions on antitrust policies and consumer protection with the rise of big data (Lam and Liu, 2020; Jin and Wagman, 2021; Krämer, 2021; Alcobendas et al., 2023; Jeon et al., 2023). We provide empirical evidence showing that public policies that equalize information access can enhance competition and improve overall consumer welfare, though with complex distributional effects.

The rest of the paper is organized as follows. In Section 2 we introduce the institutional background of the Italian auto insurance market and provide empirical evidence on the heterogeneity in insurers' pricing strategies and risk-rating precision. We then present the model and estimation strategies in Sections 3 and 4, respectively. Section 5 reports the empirical estimation results and model fit. The details of our counterfactual experiments are provided in Section 6. Section 7 concludes.

2 Institutional Background

2.1 Italian Auto Insurance Market

Our paper focuses on the market for mandatory third-party liability insurance for motor vehicles (i.e., *Responsabilità Civile Auto*) in Italy. This insurance covers damage to third parties' health and property in accidents where one is "not at fault."[9] The policies last for one year and are exclusive. Consumers decide whether to renew the contract or change insurers at the end of the year. Insurance companies are legally prohibited from rejecting consumers. The Italian mandatory liability insurance market is large; for example, in 2018, about 31 million contracts were underwritten, and over 4 million claims were filed. There are more than 50 competing firms nationwide.

The data we use for this research come from a micro dataset, IPER (*Indagine sui Prezzi Effettivi RC Auto*), collected by IVASS, the Italian insurance supervising authority. The IPER dataset covers a nationally representative sample of matched insurer–insuree panel with rich information on observable risk factors, premiums, coverage, and contractual clauses. The data also include information on the frequency and severity of claims for each sampled consumer in each contract year. Importantly, policyholders are tracked through changes of insurers.

[9]The Italian system follows the common fault-based system. In the event of an accident, all individuals who experience physical damages, except for the liable driver, are compensated by the company covering the liable vehicle.

Contract Design Two important features of the contract design stand out in the Italian auto insurance market. First, only a negligible fraction of contracts feature a deductible, and therefore consumers do not face complex deductible choices. Second, the law establishes a mandatory minimum of liability coverage.[10] The liability limits are 1 million and 6 million euros for property and health damage, respectively. In practice, claim payouts almost never exceed the mandatory minimum liability limits, and therefore consumers essentially enjoy full coverage.

Premium The premiums of the contracts are determined by insurers' actuarial algorithms based on various risk customization factors such as age, residence, and driving record. In Italy (as in many other countries), a uniform experience rating system relates the history of accidents to a class of risk, known as bonus-malus (BM) class. There are 18 BM classes with class 1 being the best. Young drivers with no driving history are assigned to BM class 14; if no accident occurs, the BM class decreases by one each year. The experience rating system suffers from the well-known saturation problem. The BM class, although publicly available, is not very informative for purposes of risk evaluations: about 80% of policyholders are in class 1. This provides strong incentives for firms to collect additional information and develop pricing algorithms to more accurately price consumers' risks.

Premiums can also vary due to optional contractual clauses and the number of installments chosen by consumers. For example, the "Exclusive Driving" clause lowers the premium, but the vehicle can be used only by the driver indicated in the contract. The "Protected Bonus" clause, which increases the premium, prevents the future premium from going up as much if an at-fault accident occurs. Not paying the entire amount upfront (i.e., in one installment) also implies a surcharge.

In this paper, we abstract away from consumers' choices of additional contract features and payment method; instead, we treat them as consumer characteristics that can directly affect premiums and risks. We include these variables as controls when we predict consumer expected claim payouts (i.e., risk types), as detailed in the next section. We model additional contractual clauses in a reduced-form manner rather than through a comprehensive structural analysis, mainly because our paper focuses on price competition among insurers with heterogeneous risk-rating technologies and cost structures. Incorporating an extra layer of endogenous contract design and consumer sorting across contract types would be even more challenging, but it is an interesting avenue for future research.[11]

[10] That is, if the damage to third parties exceeds this limit, the policyholder is responsible for any amount exceeding the limit.

[11] Empirical work endogenizing contract designs in selection markets is scarce. Einav et al. (2012) analyze the optimal design of minimum down payment and interest rate in auto loans. Decarolis and Guglielmo

2.2 Descriptive Analysis

In our empirical analysis, we focus on a sample of new customers in Rome's metropolitan area.[12] Our sample includes 124,592 liability insurance contracts sold between 2013–2021. This market has around 50 insurance companies. To reduce computational burden, we focus on the top 10 largest firms at the national level and the group of remaining fringe firms (which we denote as Firm 11).

Table 1 reports the summary statistics for contract premiums, claim frequency and severity, and consumer and vehicle characteristics. In our sample, the average annual premium paid by consumers is 478 euros. Accidents are rare events: on average, the number of claims filed in each contract year is around 0.08. The average number of accidents over the previous five years is 0.81. Approximately 56% of the sampled consumers are male. The average driver age and BM class are 48 and 2, respectively.

Table 1: Summary statistics

Variables	Mean	Std. Dev.	Min	Max	N
Premium (€)	478.06	209.03	134.26	1,334.70	124,592
Claim size (€)	264.90	10,213.62	0	2,521,014	124,592
No. of claims (within contract year)	0.08	0.30	0	6	124,592
No. of accidents in last 5 years	0.81	1.22	0	3	124,592
BM class	2.06	2.51	1	15	124,592
Age	48.24	14.11	18	99	124,592
Man	0.56	0.50	0	1	124,592
Median city	0.10	0.30	0	1	124,592
Big city	0.62	0.49	0	1	124,592
Car age	8.30	5.27	0	19	124,592
Horsepower	66.88	26.84	0	493	124,592
Petrol vehicle	0.52	0.50	0	1	124,592
One installment	0.67	0.47	0	1	124,592

Price Variations and Consumer Sorting We plot the average premium and claim payouts for 11 firms separately in Figure 1, where the size of the circle represents the market share of each firm. We observe a substantial cross-firm price dispersion in this market, even

(2017) document that insurers adjust plan generosity in response to potential changes in selection. In our context, there are more than 10 optional contractual clauses offered by the insurers. Allowing endogenous choices in this extremely large space of contract features is exceptionally challenging. Several recent papers incorporating endogenous product attributes in conventional markets include Sweeting (2013), Crawford et al. (2019), Fan and Yang (2020), and Barwick et al. (2023).

[12]We define customer tenure as the number of years since the customer initiated her contract with the insurer. Our sample includes all customers with tenure equal to 0 in a given contract year for a given insurer (i.e., the customers are new to the insurer).

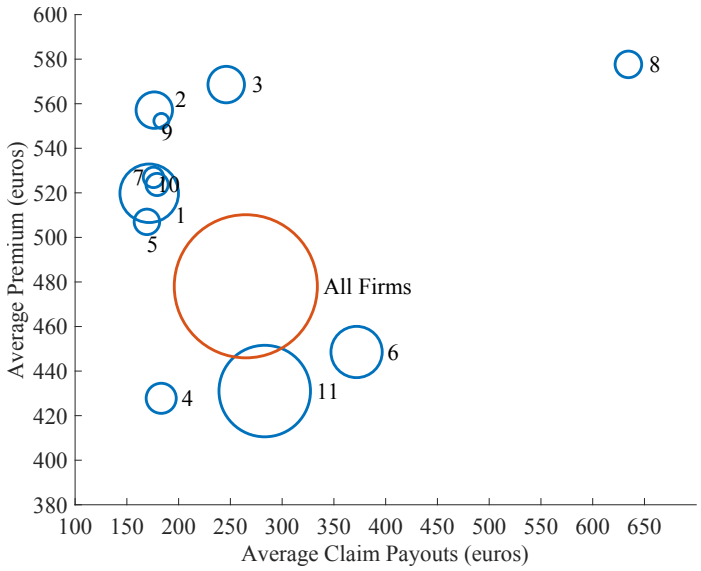


Figure 1: Heterogeneity across firms: Average premium and claim payouts (in euros). The size of each circle represents the market share of the firm, with firm IDs displayed next to each circle.

if we take into account price adjustments based on consumer risk profiles. For two insurers with similar market sizes and average claim payouts, such as firms 4 and 5, the difference in their average premiums can be as much as 80 euros. Moreover, firms differ significantly in their average claim payouts. For example, Firm 8's average claim payout is more than double the market average. Compared horizontally to other firms with similar market sizes, such as firms 7 and 10, Firm 8 attracts consumers with significantly higher average risk. These patterns suggest that consumers of different risk types are being sorted into different insurers. What drives the price variation and risk sorting observed in this market? Next, we provide empirical evidence suggesting that firms adopt significantly different pricing strategies and display considerable heterogeneity in the precision of their risk assessments.

Heterogeneity in Pricing Strategies In the data, we observe a set of commonly used factors for motor insurance pricing, including consumer characteristics (e.g., age, BM class, number of accidents over the previous five years), vehicle features (e.g., power, years since registry), province of residence, and city-level geographic characteristics. To assess how well these observable factors explain the price variation in the data, we run a series of firm-specific regressions of premiums on these variables. The estimated coefficients for a subset of covariates, along with the R^2 values of the regressions, are reported in Table 2.

We highlight several important observations from Table 2. First, the observed risk factors account for approximately 50% of the price variations in the data, suggesting that

Table 2: Regressing premiums on observable characteristics: Estimated coefficients and R^2

Firm ID	Age	BM	1 Acc.	Vehicle Age	Power	Big City	R^2
1	-1.04	19.94	124.33	-1.44	1.34	34.23	0.39
2	-0.72	24.95	148.86	-0.49	1.09	45.30	0.39
3	-1.48	26.53	105.82	2.14	1.57	57.00	0.47
4	-0.25	21.75	76.67	2.23	1.00	54.96	0.44
5	-0.40	31.11	112.47	2.61	1.23	39.44	0.48
6	-0.79	21.87	112.03	5.31	1.04	17.01	0.59
7	-1.40	23.58	125.34	2.47	1.06	49.51	0.45
8	-1.67	13.64	119.67	-2.61	1.32	54.89	0.39
9	-0.93	31.52	145.87	7.50	1.85	83.89	0.51
10	-1.02	11.91	182.24	-0.63	1.22	52.42	0.44
11	-0.52	23.10	120.83	0.93	1.07	11.66	0.52
All	-0.76	22.02	121.77	0.78	1.20	28.69	0.48

Note: full estimation results with standard errors are reported in Table 8 in Appendix A. In all regressions, we control for time dummies, optional contract clauses, and a rich set of consumer and vehicle characteristics we have access to.

premiums also depend on other information not observable to researchers. These unobservable pricing variables may include consumer or vehicle characteristics that are directly used in firms' actuarial algorithms but not available in the dataset, as well as "soft information" (e.g., a sales agent's perception of a consumer's risk type).

Second, the extent to which these commonly observed factors explain price variation differs significantly across firms, with R^2 values ranging from 0.39 to 0.59. This indicates that firms vary in how much they rely on these observable factors when setting premiums. Additionally, the estimated coefficients for key risk factors, such as age, BM class, and driving record, vary notably across firms. For instance, being one year older reduces premiums by 0.25 to 1.48 euros, a higher BM class increases premiums by 12 to 32 euros, and having one accident in the previous five years raises premiums by 77 to 182 euros across firms.

To further confirm that firms use additional and heterogeneous information for pricing, we conducted a survey of five major insurance companies in Italy and list a subset of pricing variables they use in Table 7 of the appendix.[13] Importantly, Table 7 shows that the set of variables used for pricing varies significantly across firms. For instance, some firms consider the presence of safety devices, finer-level zip codes, occupation, and socioeconomic factors, while others do not. Together, these findings provide concrete evidence of significant heterogeneity in the variables and algorithms used for pricing across firms.

[13]The variables in Table 7 are not included in our data. We report a subset of pricing variables to protect the confidentiality of the company and to ensure anonymity.

Heterogeneity in Information Precision While the pricing strategies vary across firms, do insurers also differ in the precision of their risk assessments? To explore this question, we examine the extent to which each firm's premium is correlated with consumer risk. Although a consumer's true risk type is *ex-ante* unknown to insurers, it is eventually revealed after observing the consumer over multiple periods. In other words, the consumer's risk type becomes *ex-post* estimable based on realized claim data. Intuitively, if a firm's premiums are more responsive to (or more accurately reflect) the true risk, this suggests that the firm is better at evaluating consumer risk, and vice versa.

We estimate an individual-specific riskiness measure using a panel dataset of claim records, which includes the number of accidents and the amount paid to the consumers following an accident in each contract year. The important advantage of our dataset is that we are able to track individuals across different firms over a long period of time.[14] Specifically, we estimate claim frequency and severity separately and multiply the two to represent consumers' risk level. This two-part approach is widely adopted in the auto insurance industry. Actuaries rely heavily on the generalized linear models (GLM) and they typically model claim frequency and severity separately (see more details and discussions in Goldburd et al., 2016).

Let i be the index of each customer and t be the index of a contract year. For each consumer in the dataset, we observe a vector of characteristics (such as age, vehicle features, geographic locations, and contractual clauses), which we denote by $\boldsymbol{X}_{it}$. We first estimate a Poisson regression of claim counts on observables with an individual fixed effect. Specifically,

$$E[ClaimCount_{it}|\boldsymbol{X}_{it}, \lambda_i] = \exp(\boldsymbol{X}_{it}\delta_c + \omega_i). \tag{2.1}$$

We then estimate a log-normal regression of claim size conditional on the consumer being involved in an accident:[15]

$$\log(ClaimSize_{it}) = \delta_0 + \boldsymbol{X}_{it}\delta_s + \eta_{it}. \tag{2.2}$$

With the estimated regression coefficients δ and the individual fixed effects ω from Equations (2.1) and (2.2), we predict the expected number of accidents and the expected indem-

[14]Existing literature on auto insurance markets often has access to contract and claim data from a single insurer (e.g., Cohen, 2005; Cohen and Einav, 2007; Jeziorski et al., 2017). Estimating consumers' risks using data from one firm is susceptible to selective censoring issues. This is because in the auto insurance markets, high-risk drivers are more likely to switch, especially after being involved in an accident (Cohen, 2005; Cosconati, 2023).

[15]We assume that claim size is independent of the individual fixed effect following Jeziorski et al. (2017). This assumption is motivated by the arguments in the actuarial literature that accident severity is more random and less related to the individual's driving ability.

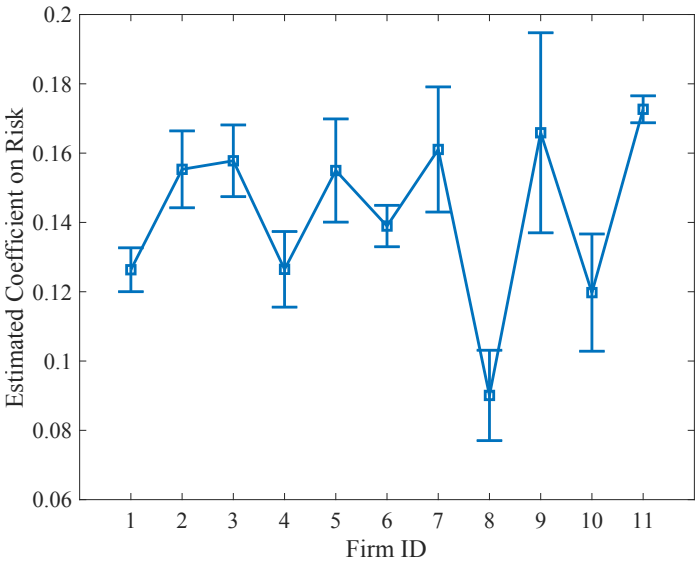


Figure 2: Regressing premiums on consumer risks: Estimated coefficients on risk measure for each firm, with error bars representing the 95% confidence intervals.

nity conditional on involvement in an accident for each consumer. Multiplying the two yields an estimate for the expected cost of insuring consumer i in a year, which we interpret as consumer's true risk type. Since we control for contract features in these regressions, the impact of consumers self-selecting into different clauses has been factored into our risk estimates.[16]

Using the individual-specific risk measure, we run separate linear regressions of premiums paid by consumers on their risk for each firm. Figure 2 shows the coefficients on the risk measure for each firm, with error bars representing the 95% confidence intervals. This figure shows that the extent to which firms adjust their premiums in response to increased true risk varies significantly across firms. For example, Firm 7 charges a premium that is much more sensitive to risk compared to Firm 8, suggesting that Firm 7 potentially has higher precision in evaluating risk.

However, the patterns shown in Figure 2 are complex equilibrium outcomes and the differences in the correlation between premium and risk could be due to other factors. For example, if Firm 8 is more efficient at processing claims, it may design a pricing algorithm to target high-risk consumers. We therefore need a tractable empirical framework that incorporates rich heterogeneity in firms' risk-rating precision, cost structures, and product differentiation to better understand the main drivers of price variation and risk sorting in this market.

[16]Our estimates of risk also take into account the moral hazard effect in a reduced-form way. That is, consumers may change their risky behavior after choosing different contracts.

3 Model

Since all of our analysis can be derived conditional on observable consumer characteristics, such as age, BM class, and vehicle features, we omit these variables from the notation for simplicity when describing the model and estimation strategies in Sections 3 and 4.

We take as given that there are J insurers in the market, indexed by $j = 1, 2, \cdots, J$. Each of them offers a standardized insurance product that shares the same contract features. The insurance contract is exclusive and mandatory for all drivers. In other words, consumers have to purchase the product from one of the companies and there is no outside option. Let θ denote a consumer's true risk type, which can be measured by the expected cost of insuring a consumer within a contract year. The risk type θ is *not* observed by insurers ex ante. The population density of θ is denoted by $f_0(\theta)$, which is common knowledge among all firms.

Signal Structure When a consumer of type θ arrives, firm j draws a signal $\hat{\theta}_j$ from a cumulative distribution function $F_j(\hat{\theta}_j|\theta)$, with its density denoted by $f_j(\hat{\theta}_j|\theta)$. Intuitively, the signal $\hat{\theta}_j$ can be viewed as a risk score assigned to the consumer by firm j based on information it has access to. We assume that the signals are private and independent across firms conditional on θ. This setting is similar to that of common value auctions, where firms' signals can be interpreted as their noisy estimates of the true but unknown common value, i.e., the expected claim payouts to the consumer.

The signal distribution $F_j(\cdot|\theta)$ captures firm j's information technology for risk evaluation. For example, the signal can admit a normal distribution $N(\theta, \sigma_j^2)$, where σ_j measures the dispersion of the signal distribution around the true risk type, and therefore $1/\sigma_j$ captures firm j's information precision. When $\sigma_j = 0$, firm j perfectly observes θ; as σ_j increases, the signal becomes less precise. In reality, the set of variables and algorithms used in the risk-rating process may differ across firms (as documented in Section 2.2); moreover, sales agents in some firms may have more experience and better knowledge about consumers' driving habits. These factors jointly contribute to the heterogeneity of firms' signal distribution $F_j(\cdot|\theta)$.

Risk Rating Firm j infers the risk type of the consumer upon observing the signal $\hat{\theta}_j$. Let $f(\theta|\hat{\theta}_j, j)$ denote the posterior density of θ conditional on the firm receiving a signal $\hat{\theta}_j$ and the consumer being *selected* into firm j. The posterior mean $E(\theta|\hat{\theta}_j, j)$ represents firm j's

risk rating for a consumer conditional on signal $\hat{\theta}_j$ and her choosing j. Specifically,

$$E(\theta|\hat{\theta}_j, j) = \int_\theta \theta f(\theta|\hat{\theta}_j, j)d\theta = \frac{\int_\theta \theta Pr(j|\hat{\theta}_j, \theta) f_j(\hat{\theta}_j|\theta) f_0(\theta)d\theta}{\int_\theta Pr(j|\hat{\theta}_j, \theta) f_j(\hat{\theta}_j|\theta) f_0(\theta)d\theta}, \tag{3.1}$$

where $Pr(j|\hat{\theta}_j, \theta)$ is the probability that a consumer of type θ chooses firm j given the signal $\hat{\theta}_j$. Note that $Pr(j|\hat{\theta}_j, \theta)$ is an equilibrium object because it depends on all firms' signal precision and pricing strategies, as well as consumers' demand responses. Consequently, $E(\theta|\hat{\theta}_j, j)$ is an equilibrium object.

Pricing Strategy Firm j sets the price for a consumer based on their risk evaluation. We assume that the firm's pricing strategy takes the following form:

$$p_j(\hat{\theta}_j) = \alpha_j + \beta_j \underbrace{E(\theta|\hat{\theta}_j, j)}_{\text{risk rating}}, \tag{3.2}$$

where α_j and β_j are pricing coefficients optimally chosen by the firm. Specifically, α_j reflects the firm's baseline markup and β_j relates to the elasticity of price with respect to risk rating.

The price in our model is a function of the signal that the firm receives. Prices may vary with the signals in a complicated nonlinear way, even if we assume that price is a linear function of risk rating. This is because $E(\theta|\hat{\theta}_j, j)$ itself is a complex function of the signal due to self-selection into insurers, as shown in Equation (3.1). The linear structure imposed in Equation (3.2) is *not essential* for our model. More generally, we can assume that prices are monotonically increasing functions of risk rating, i.e., $p_j(\hat{\theta}_j) = \tau_j(E(\theta|\hat{\theta}_j, j))$. Imposing a linear pricing strategy makes our identification argument more transparent and significantly reduces the computational burden, as we will show in Section 4.

Demand We now describe the consumer's choice problem. We follow the standard differentiated product framework and assume that the level of utility that a consumer derives from a product is a function of the price and product characteristics. The insurance plans offered in this market are homogeneous in terms of observable characteristics, but may have unobserved (by the econometrician) product attributes such as service quality or brand loyalty. We assume that the utility derived by consumer i from a product of firm j is given by the scalar value

$$U_{ij} = -\gamma p_j(\hat{\theta}_j) + \xi_j + \varepsilon_{ij}, \tag{3.3}$$

where γ represents the price sensitivity parameter and ξ_j represents the unobserved heterogeneity for product j. It is straightforward to extend our demand model to allow for observable product attributes. Alternatively, we can interpret ξ_j as firm-specific mean utilities, which depend on both observed and unobserved product attributes.

We assume that ε_{ij} follows a type I extreme value distribution and is independent across all firms and individuals. The probability that a consumer chooses firm j given a vector of signals $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_J)$ drawn by all firms is

$$Pr(j|\hat{\boldsymbol{\theta}}) = \frac{\exp(-\gamma p_j(\hat{\theta}_j) + \xi_j)}{\sum_{j'=1}^{J} \exp(-\gamma p_{j'}(\hat{\theta}_{j'}) + \xi_{j'})}. \tag{3.4}$$

In Equation (3.4), a consumer's risk type θ does not directly influence their utilities, but instead affects their choice probabilities by shifting the signal distributions, and ultimately, the price distributions they face. This specification highlights how heterogeneity in firms' pricing strategies impacts consumer sorting patterns, which is a key mechanism emphasized in this paper.

In our empirical estimation, we adopt a more flexible specification by allowing the consumer's preference parameters, such as price sensitivity and unobserved product heterogeneity, to vary with their risk type. This captures the idea that higher-risk consumers may be more price sensitive and have a stronger preference for higher-quality services. The identification and estimation details of the demand model when preference parameters depend on θ are discussed in Section 4.1.

Profit Maximization Each firm is assumed to choose its pricing coefficients (α_j, β_j) to maximize its profit given the pricing strategies of competing firms. We assume that this is a static, complete information game, and therefore the primitives of each firm (e.g., their signal distributions, cost parameters, and unobserved product attributes) are common knowledge.

We use $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_J)$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_J)$ to denote the pricing coefficients of all firms. The profits of firm j are

$$\pi_j(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \int_{\theta} \int_{\hat{\boldsymbol{\theta}}} \Bigg(\underbrace{p_j(\hat{\theta}_j) + c_j - k_j\theta}_{\text{net profit}} \Bigg) \underbrace{Pr(j|\hat{\boldsymbol{\theta}})}_{\text{choice prob.}} \Bigg(\underbrace{\prod_{j'=1}^{J} f_{j'}(\hat{\theta}_{j'}|\theta)}_{\text{signal dist.}} \Bigg) \underbrace{f_0(\theta)}_{\text{type dist.}} d\hat{\boldsymbol{\theta}} d\theta. \tag{3.5}$$

The first term in Equation (3.5) represents firm j's net profit from servicing a consumer with type θ given the price $p_j(\hat{\theta}_j)$. We use c_j to denote the "net benefit" from contracting

with a customer irrespective of their risk type. Insurers need to pay administrative costs (e.g., setting up the profile) when managing contracts, but may also receive future benefits from contracting with new customers due to inertia and dynamic pricing strategies.[17] Our paper focuses on a static price competition for new customers and does not provide a full-fledged dynamic analysis of firms' pricing strategies and consumers' switching behavior. We treat c_j as a primitive of our model. In the estimation section, we calibrate the expected future benefits from contracting with a new customer using the data and then isolate the marginal cost for managing contracts from c_j.

Another cost component in our model is $k_j\theta$, which represents the cost related to claim payouts. In addition to the indemnity paid to the consumers following an accident, which is measured by θ directly, insurers may incur extra costs of processing the claims. For instance, the insurer may need to hire adjusters to inspect the vehicle damage. The scalar k_j essentially measures the efficiency of the firm in processing claims.

The term $Pr(j|\hat{\boldsymbol{\theta}})$ in Equation (3.5) is the probability that a consumer is selected into firm j given the prices and product attributes across all firms. The third term in Equation (3.5) represents the joint distribution of signals from all firms, which, under our conditional independence assumption, can be written as the product of conditional densities. The last term $f_0(\theta)$ denotes the population distribution of consumers' risk type. Each firm integrates over the joint distribution of types and signals $(\theta, \hat{\boldsymbol{\theta}})$ to compute its expected profits.

To summarize, our model allows for rich supply-side heterogeneity. Insurers are differentiated by their cost structures (c_j, k_j), information technology (F_j), and unobserved product attributes (ξ_j). All of these are primitives of the firms and are allowed to be arbitrarily correlated. Throughout the paper, we maintain the assumption that firms' primitives cannot be adjusted within a relatively short period of time. Insurers optimally choose their pricing coefficients (α_j, β_j) to maximize profits, taking into account their opponents' strategies and demand-side responses.[18]

[17] For example, Handel (2013) identifies substantial inertia in health insurance markets. Honka (2014) documents that in the US auto insurance industry, about 70% of consumers stay with their insurance provider after the first year, and this percentage increases with tenure. High switching costs provide incentives for firms to use a "bargain-then-ripoff" pricing strategy (Farrell and Klemperer, 2007; Ericson, 2014). Cosconati (2023) documents that in Italian auto insurance markets, new customers on average receive a discount of 91 euros relative to old customers; premiums increase rapidly in the first five years and gradually become inelastic to tenure.

[18] The payoff functions π_j in Equation (3.5) are continuous in $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. We verify numerically that π_j are concave in (α_j, β_j). Since the strategy spaces are nonempty compact convex subsets of a Euclidean space, a pure-strategy Nash equilibrium of our pricing game exists following Fudenberg and Tirole (1991, Theorem 1.2).

4 Estimation

A consumer's true risk type θ is ex-ante unknown to the insurer but can be estimated ex-post using realized claim data over multiple years. We obtain θ as an individual-specific estimate of the expected cost to insure a consumer over a contract year, as detailed in Section 2.2. For the remainder of the analysis, we take θ as given and known by the econometricians.

On the demand side, we estimate consumers' price sensitivity and the unobserved product attributes for each firm using a novel fixed-point algorithm. The supply-side parameters include the firms' pricing coefficients, signal distributions, and cost parameters. We exploit (1) the joint distribution of premiums and risk types for each firm and (2) the first-order conditions derived from each firm's profit maximization problem to identify and estimate the supply-side parameters. The details of each step are provided in this section.

4.1 Demand

In classic demand models, the price of a product is often assumed to be the same for all consumers. The premiums for insurance contracts, however, depend on various consumer characteristics and discounts are highly likely to vary across individuals. The key challenge in estimating the demand model is that our data do not include the full price menu faced by consumers. Our data include only contracts sold in the market, and therefore we do not have access to the premiums of the unchosen products. This is a common issue in many empirical studies (e.g., Goldberg, 1996; Cicala, 2015; Crawford et al., 2018; Allen et al., 2019; Salz, 2022; Sagl, 2023) and the standard approach is predicting the alternative prices using regressions.

Unfortunately, this standard approach does not apply to our setting. We do not have access to all pricing variables insurers may use, and more importantly, different firms may use different sets of pricing variables and algorithms, as evidenced by Table 7. Crawford et al. (2018) face a similar problem and they estimate the following linear regression model to predict prices of non-chosen products:

$$p_{ij} = \gamma_0 + \boldsymbol{X}_{ij}\gamma_1 + \lambda_j + \omega_i + \varepsilon_{ij},$$

where λ_j is the firm fixed effect and ω_i is the individual fixed effect. The term ω_i includes factors that are observed by all firms in the market, but are unobserved by econometricians (i.e., "soft information" that firms may have about consumers). This regression approach assumes that all firms observe the same set of soft information about consumers (ω_i). This assumption is restrictive particularly in our setting because it essentially erases any infor-

mational asymmetry among competing firms.

Our paper instead adopts a novel fixed-point approach that jointly estimates consumers' sorting probabilities, price distributions, and demand parameters when only transaction prices are available. This approach extends the classic demand estimation techniques developed in the seminal paper of Berry et al. (1995). In addition to matching observed aggregate market share of each firm, our method introduces an outer loop that iterates over consumers' sorting propensities.

Offered vs. Accepted Prices We first highlight that there are two kinds of price distributions in our setting. The price distribution we observe in the data for each firm is the *accepted* price distribution after selection. It is different from the *offered* price distribution that consumers face when making their enrollment decisions. To draw an analogy, the offered prices correspond to submitted bids in auctions and potential wages in Roy models (Roy, 1951), while the accepted prices correspond to winning bids and observed wages. The key observation is that the offered and accepted price distributions are linked through the demand system (or, more generally, the selection rules). Figure 3 provides a graphical illustration of the relationship between offered and accepted price distributions from a simulation exercise. Figure 3a shows that with a lower price, the consumer is more likely to stay. Normalizing the histograms in Figure 3a to plot the density of offered and accepted prices, Figure 3b reveals that after selection, the density of the price distribution shifts to the left. Consequently, lower prices are over-represented, while higher prices are under-represented in the accepted price distribution observed by econometricians.

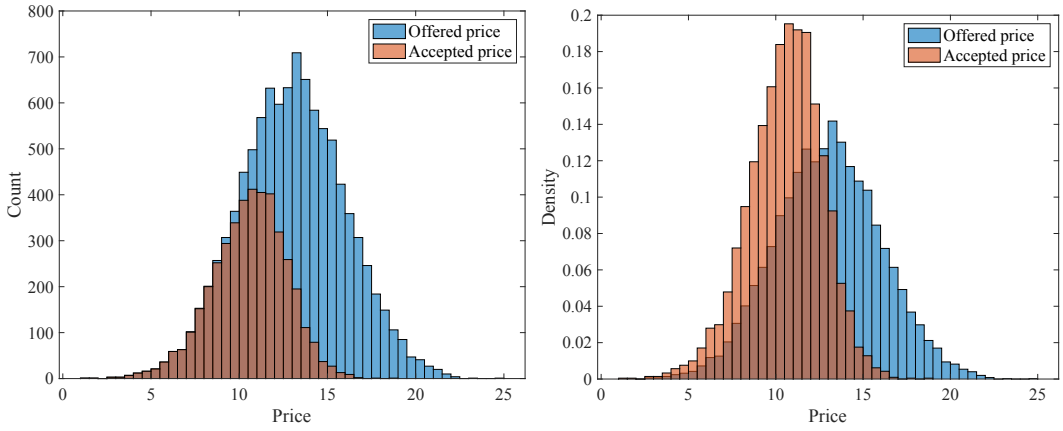
We derive firms' offered price distributions from their accepted price distributions and the choice probabilities using the demand model. Let $g_j(p|\theta)$ and $\tilde{g}_j(p|\theta)$ denote the density of the offered and accepted price in firm j conditional on the true risk type θ. Note that $\tilde{g}_j(p|\theta)$ is the posterior distribution of price *conditional on j being selected* and it can be directly and nonparametrically estimated from the data. By applying Bayes' rule, we can easily observe that:

$$\tilde{g}_j(p|\theta) \propto g_j(p|\theta) Pr(j|p_j = p, \theta),$$

where $Pr(j|p_j = p, \theta)$ represents the likelihood that consumers with type θ choose firm j given a particular price p. To derive the density of the offered price, we divide both sides by the likelihood:

$$g_j(p|\theta) = \frac{\tilde{g}_j(p|\theta)/Pr(j|p_j = p, \theta)}{\int_{p'} \tilde{g}_j(p'|\theta)/Pr(j|p_j = p', \theta)dp'}, \tag{4.1}$$

where the denominator is for normalization.



(a) Count (b) Density

Figure 3: Comparing offered and accepted price distributions for a firm using simulated data. We draw offered prices from $N(13, 3)$, and the probability that the consumer given price p stays with the firm is given by $\exp(10 - p)/(0.1 + \exp(10 - p))$. Panels (a) and (b) plot the histograms of the offered and accepted prices, normalized by count and density, respectively.

In Equation (4.1), the likelihood $Pr(j|p_j = p, \theta)$ is not directly observable from the data, but it can be derived using the demand model and the distributions of offered prices of other firms. Let $\boldsymbol{p}_{-j} = (p_1, \cdots, p_{j-1}, p_{j+1}, \cdots, p_J)$ denote the vector of prices excluding firm j's price.

$$Pr(j|p_j = p, \theta) = \int_{\boldsymbol{p}_{-j}} \frac{\exp(-\gamma p + \xi_j)}{\exp(-\gamma p + \xi_j) + \sum_{j' \neq j} \exp(-\gamma p_{j'} + \xi_{j'})} \left(\prod_{j' \neq j} g_{j'}(p_{j'}|\theta) \right) d\boldsymbol{p}_{-j}. \tag{4.2}$$

Combining Equations (4.1) and (4.2) for all firms yields a system that characterizes the *offered* price distributions as its fixed point for any given value of demand parameters $(\gamma, \boldsymbol{\xi})$, where $\boldsymbol{\xi} = (\xi_1, \xi_2, \cdots, \xi_J)$. Wu and Xin (2024) construct an operator whose fixed point is the offered price distributions and show that it is a functional contraction under mild conditions, which guarantees the existence and uniqueness of the fixed point.[19]

Given the fixed point problem defined in Equations (4.1) and (4.2), we use an iterative procedure to solve for the *offered* price distributions for all firms as functions of demand parameters. We nest the contraction mapping for mean utilities in Berry et al. (1995) in our algorithm to obtain $\boldsymbol{\xi}$ that matches aggregate market shares. At convergence, for a

[19] Wu and Xin (2024) propose a multi-step semiparametric maximum likelihood estimator for the demand parameters and the offered price distributions. The consistency and asymptotic normality of the proposed estimator are established. We refer the readers to Wu and Xin (2024) for more theoretical results and technical details.

fixed price sensitivity parameter γ, we obtain (1) the offered price distribution faced by consumers $g_j(p|\theta;\gamma)$ for all firms, (2) the likelihood that consumers are selected into firm j, $Pr(j|p_j = p, \theta;\gamma)$, and (3) the vector of unobserved product attributes $\boldsymbol{\xi}(\gamma)$ that match the market shares. The details of the iterative algorithm are provided in Appendix B.

Price Sensitivity Finally, we estimate the price sensitivity parameter γ using the sorting pattern observed in the data. To see which moments in the data identify γ, consider the likelihood that a consumer of type θ chooses firm j:

$$Pr(j|\theta;\gamma) = \int_{\boldsymbol{p}} \frac{\exp(-\gamma p_j + \xi_j(\gamma))}{\sum_{j'} \exp(-\gamma p_{j'} + \xi_{j'}(\gamma))} \left(\prod_{j'} g_{j'}(p_{j'}|\theta;\gamma) \right) d\boldsymbol{p}. \tag{4.3}$$

In an extreme case where γ is equal to 0, the probability that a consumer chooses firm j does not vary with their risk type anymore, so that the risk sorting pattern of consumers would disappear. In other words, γ is identified from variations in risk levels across firms.

For a sample of consumers indexed by $i = 1, 2, \cdots, N$, let θ_i denote their risk types and $y_i \in \{1, 2, \cdots, J\}$ denote their choices. We propose a nested fixed point algorithm to estimate γ. In the inner loop, we solve for $\boldsymbol{\xi}(\gamma)$ and $g_j(p|\theta;\gamma)$ using the iterative procedure detailed in Appendix B for each γ and construct the choice probabilities in Equation (4.3). In the outer loop, we search for γ to maximize the following log-likelihood function:

$$LL_d(\gamma) = \sum_{i=1}^{N} \log \Big(Pr(y_i|\theta_i;\gamma) \Big).$$

Risk-Dependent Preference Parameters Up to this point, we have focused on the simplified demand model in Equation (3.4), where the preference parameters do not vary with risk type. However, there may be important correlations between a consumer's risk type and their price sensitivity or preference for unobserved product attributes.

Allowing γ and $\boldsymbol{\xi}$ to depend on θ *does not* pose any issues for our fixed-point algorithm, as Equations (4.1) and (4.2) are derived conditional on θ. In fact, our fixed-point algorithm converges separately for each θ. However, from an identification perspective, it is not possible to identify the demand model when γ and $\boldsymbol{\xi}$ fully nonparametrically depend on θ, without imposing additional restrictions. To see why, note that the key moments used for identification are the choice probabilities conditional on risk type, i.e., $Pr(j|\theta)$. For each θ, the number of available moments is $J - 1$. In contrast, the number of unknown parameters for each θ is J—including $\gamma(\theta)$, $\xi_2(\theta), \ldots, \xi_J(\theta)$, with $\xi_1(\theta)$ normalized to 0. This model is not identified without additional information or restrictions, as the number of unknowns

exceeds the number of constraints.

One straightforward way to impose restrictions is to assume that the preference parameters are the same for consumers within a specific risk group. Note that all of our arguments for identifying and estimating demand parameters $(\gamma, \boldsymbol{\xi})$ remain exactly the same if we partition the sample into multiple subsamples (e.g., high-risk versus low-risk consumers). In other words, we can estimate the demand parameters separately for different risk groups of consumers and allow their preferences to fully vary across these groups. Alternatively, we can impose parametric restrictions on $\gamma(\theta)$ and $\boldsymbol{\xi}(\theta)$, such as $\gamma(\theta) = \gamma_0 + \gamma_1\theta$. Imposing such parametric restrictions significantly reduces the number of unknowns. For example, the linear structure in γ introduces only two unknown parameters, which is far lower than the cardinality of θ. In summary, our approach allows the preference parameters to vary with a consumer's risk type in a flexible parametric way, making it applicable to a wide range of empirical settings.

4.2 Pricing Coefficients

We now turn to the estimation of the supply-side parameters. We first focus on estimating the pricing coefficients (α_j, β_j), which are equilibrium objects optimally chosen by the firm. For each consumer in firm j, we observe their risk type θ and the premium charged by firm j. The difficulty in estimating the pricing coefficients is due to the fact that we do not observe the equilibrium risk rating $(E(\theta|\hat{\theta}_j, j))$ assigned to the consumer by the firm given their private signal $\hat{\theta}_j$.

We show that under the linear structure imposed on the pricing strategy in Equation (3.2), α_j and β_j are identified from the first and second moments of the joint distribution of premiums and risks. To see this, we derive the within-firm mean and covariance of p and θ:

$$E(p_j(\hat{\theta}_j)) = \alpha_j + \beta_j E(E(\theta|\hat{\theta}_j, j)) = \alpha_j + \beta_j \mu_j, \tag{4.4}$$

$$cov(p_j(\hat{\theta}_j), \theta) = \beta_j var(E(\theta|\hat{\theta}_j, j)) = \frac{var(p_j(\hat{\theta}_j))}{\beta_j}, \tag{4.5}$$

where μ_j denotes the expectation of θ among consumers who choose firm j. Equations (4.4) and (4.5) uniquely determine (α_j, β_j) as the solution to a system of two linear equations. Note that all the first- and second-order moments of premiums and risks in firm j can be directly estimated from the data. Plugging the moment estimates back into these equations, we easily obtain the estimators for (α_j, β_j). An important advantage of this approach is that it relies only on data moments and imposes no parametric constraints on the type and signal distributions.

Given the linear structure imposed on firms' pricing strategies, the equilibrium pricing coefficients can be identified and estimated separately from other parts of the model. This significantly reduces the computational burden. If we instead assume that the premium is a nonlinear function of the risk rating, we then need to estimate the pricing coefficients together with other supply-side primitives, such as firms' signal distributions and cost parameters, which is feasible but computationally much more demanding.

4.3 Signal Distribution

Thus far, we have recovered the offered price distribution $g_j(p|\theta)$ faced by consumers and the pricing coefficients (α_j, β_j) for all j. The next step is to pin down the firm-specific signal distribution that generates the observed price dispersion.

The key insight to identify the signal distribution is that there is a one-to-one mapping between the offered price and the signal. This concept is analogous to the auction models, where bids are monotonically increasing in bidders' valuations. The nonparametrically identified bid distribution therefore identifies the underlying valuation distribution under model restrictions (Guerre et al., 2000).

Following similar ideas in auction models, when price is a monotonically increasing function of the signal, the change of variables formula yields

$$G_j(p_j(\hat{\theta}_j)|\theta) = F_j(\hat{\theta}_j|\theta), \tag{4.6}$$

where $G_j(p|\theta)$ denotes the cumulative distribution function of the offered price distribution of firm j for a given risk type θ. Note that $G_j(p|\theta)$ has been recovered from previous steps and can thus be treated as known. Equation (4.6) implies that for a given θ, if we know the signal distribution F_j, we can pin down the corresponding price for a signal $\hat{\theta}_j$ using the inverse of the observed offered price distribution G_j. Specifically,

$$p_j^o(\hat{\theta}_j; F_j, \theta) = G_j^{-1}\Big(F_j(\hat{\theta}_j|\theta)\Big). \tag{4.7}$$

Equation (4.7) identifies the one-to-one mapping between the the price and the signal for a given signal distribution F_j and type θ. This relationship is extremely useful because it allows us to evaluate the equilibrium risk rating given the signal $\hat{\theta}_j$ as a function of the signal distribution F_j:

$$E(\theta|\hat{\theta}_j, j; F_j) = \frac{\int_\theta \theta Pr(j|p_j^o(\hat{\theta}_j; F_j, \theta), \theta) f_j(\hat{\theta}_j|\theta) f_0(\theta) d\theta}{\int_\theta Pr(j|p_j^o(\hat{\theta}_j; F_j, \theta), \theta) f_j(\hat{\theta}_j|\theta) f_0(\theta) d\theta}. \tag{4.8}$$

With the one-to-one mapping, we essentially replace the complicated equilibrium object $Pr(j|\hat{\theta}_j, \theta)$ in Equation (3.1) with the choice probabilities evaluated at the corresponding price $p_j^o(\hat{\theta}_j; F_j, \theta)$, which we have recovered from previous demand estimation.

Once we have obtained $E(\theta|\hat{\theta}_j, j; F_j)$, we derive the model-generated price as a function of the signal, which we denote by $gp_j(\hat{\theta}_j; F_j)$. Specifically,

$$gp_j(\hat{\theta}_j; F_j) = \alpha_j + \beta_j E(\theta|\hat{\theta}_j, j; F_j). \tag{4.9}$$

Note that the pricing coefficients (α_j, β_j) have been recovered from the previous step. Equation (4.9) implies that we can generate the firm's price as a function of the signal $\hat{\theta}_j$ for any fixed signal distribution F_j. This is crucial for deriving the model-implied joint distribution of premiums and risk types for any given F_j. To see this, we first derive the posterior joint distribution of signals and risk types for any given signal distribution F_j using the Bayes rule:

$$f(\hat{\theta}_j, \theta|j; F_j) = \frac{Pr(j|p_j^o(\hat{\theta}_j; F_j, \theta), \theta) f_j(\hat{\theta}_j|\theta) f_0(\theta)}{s_j}.$$

Let gp_j^{-1} denote the inverse function of gp_j in Equation (4.9) to back out the signal corresponding to an observed premium. Then using the change of variables formula, we construct the posterior joint distribution of the premiums and risk types conditional on the consumers being selected into firm j:

$$h(p, \theta|j; F_j) = \frac{f(gp_j^{-1}(p; F_j), \theta|j; F_j)}{gp'(gp_j^{-1}(p; F_j); F_j)}.$$

Matching $h(p, \theta|j; F_j)$ with the data pattern provides the key identifying restrictions for the signal distribution. Intuitively, when the signal distribution is very informative, consumers' risks are better reflected in their premiums, so the correlation between the premium and risk type within a firm would be higher. We formalize this identification intuition for a special case where the signals are drawn from a normal distribution in Appendix C. This is also the empirical specification we use for our application. We show that when the risk rating $E(\theta|\hat{\theta}_j, j; F_j)$ can be approximated by a linear function of the signal $\hat{\theta}_j$, the correlation between the premium and risk type within a firm (which is directly estimable from the data) decreases monotonically with the variance of the signal distribution.

Given the identification argument, we estimate the firm-specific signal distribution by matching the model-implied posterior distribution of the premiums and risk types with the data. Since the identification of the signal distribution relies on within-firm variations, we can estimate it for each firm separately, which significantly reduces the computational bur-

den. Again, we use θ_i, p_i, and y_i to denote consumer i's risk type, premium, and choice of product. We estimate F_j for each firm j by maximizing the following likelihood function:

$$LL_p(F_j) = \sum_{i=1}^{N} \mathbf{1}\{y_i = j\} \log\Big(h(p_i, \theta_i | j; F_j)\Big).$$

4.4 Cost

The final step of our estimation is to recover the cost parameters from the first-order conditions of firms' profit maximization problem. To simplify notation, we use $f(\hat{\boldsymbol{\theta}}|\theta)$ to represent the joint density of signals conditional on the type θ, i.e., $f(\hat{\boldsymbol{\theta}}|\theta) = \prod_{j'=1}^{J} f_{j'}(\hat{\theta}_{j'}|\theta)$. Taking the first-order derivatives of the profit function with respect to (α_j, β_j) yields

$$\begin{aligned}
\frac{\partial \pi_j(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \alpha_j} &= \int_{\theta}\int_{\hat{\boldsymbol{\theta}}} Pr(j|\hat{\boldsymbol{\theta}}) f(\hat{\boldsymbol{\theta}}|\theta) f_0(\theta) d\hat{\boldsymbol{\theta}} d\theta \\
&+ \int_{\theta}\int_{\hat{\boldsymbol{\theta}}} (\alpha_j + \beta_j\theta + c_j - k_j\theta) \frac{\partial Pr(j|\hat{\boldsymbol{\theta}})}{\partial \alpha_j} f(\hat{\boldsymbol{\theta}}|\theta) f_0(\theta) d\hat{\boldsymbol{\theta}} d\theta, \qquad (4.10) \\
\frac{\partial \pi_j(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \beta_j} &= \int_{\theta}\int_{\hat{\boldsymbol{\theta}}} \theta Pr(j|\hat{\boldsymbol{\theta}}) f(\hat{\boldsymbol{\theta}}|\theta) f_0(\theta) d\hat{\boldsymbol{\theta}} d\theta \\
&+ \int_{\theta}\int_{\hat{\boldsymbol{\theta}}} (\alpha_j + \beta_j\theta + c_j - k_j\theta) \frac{\partial Pr(j|\hat{\boldsymbol{\theta}})}{\partial \beta_j} f(\hat{\boldsymbol{\theta}}|\theta) f_0(\theta) d\hat{\boldsymbol{\theta}} d\theta. \qquad (4.11)
\end{aligned}$$

In Equations (4.10) and (4.11), the first terms capture the direct effects of changing the pricing coefficients on profit, while the second terms quantify the marginal changes to the profit through indirect sorting effects. We approximate the derivatives of sorting probabilities with respect to α_j and β_j in the following equations:

$$\frac{\partial Pr(j|\hat{\boldsymbol{\theta}})}{\partial \alpha_j} \approx -\gamma Pr(j|\hat{\boldsymbol{\theta}})(1 - Pr(j|\hat{\boldsymbol{\theta}})), \qquad (4.12)$$

$$\frac{\partial Pr(j|\hat{\boldsymbol{\theta}})}{\partial \beta_j} \approx -\gamma E(\theta|\hat{\theta}_j, j) Pr(j|\hat{\boldsymbol{\theta}})(1 - Pr(j|\hat{\boldsymbol{\theta}})). \qquad (4.13)$$

The approximations in Equations (4.12) and (4.13) ignore the impact of (α_j, β_j) on risk rating $E(\theta|\hat{\theta}_j, j)$. Given the equilibrium concept we use (i.e., Nash equilibrium), changing firm j's pricing coefficients does not affect other firms' ($j' \neq j$) pricing strategy $p_{j'}(\hat{\theta}_{j'})$. In other words, other firms will keep using the same pricing strategies they currently use. Changing α_j and β_j will affect firm j's own expected risk level (i.e., $E(\theta|\hat{\theta}_j, j)$), but the impact of this single term on the selection probability could be small.

With the approximations in Equations (4.12) and (4.13), the first-order conditions in

Equation (4.10) and (4.11) are reduced to a system of two linear equations involving (c_j, k_j). These two equations uniquely pin down (c_j, k_j) as the solution to the system of linear equations, where all other terms are either directly estimable from the data or have been recovered in previous steps.

5 Results

We apply our model and estimation strategy to the Italian auto insurance industry, focusing on the top 10 largest firms and Firm 11, which represents the group of remaining fringe firms. In this section, we first present the estimation results, followed by a discussion of the model's fit to the data.

5.1 Demand-Side Estimation Results

Our demand estimation allows the consumer's preference parameters to depend on their risk type (θ) and observable characteristics (x) through parametric forms. Specifically, we classify consumers with θ above the median as high-risk (HR), and those below as low-risk (LR). We estimate the consumer's preference for unobserved product attributes for these two groups separately, i.e., $\boldsymbol{\xi}(HR)$ and $\boldsymbol{\xi}(LR)$. For the price sensitivity parameter γ, we impose the following parametric assumption:

$$\gamma(x, \theta) = \gamma_0 + \gamma_1 OLD + \gamma_2 HR,$$

where $OLD = 1$ if the consumer's age is above the median, and 0 otherwise.[20]

We report the estimates for price sensitivity parameters $(\gamma_0, \gamma_1, \gamma_2)$ and firm-specific unobserved product heterogeneity in the left panel of Table 3. The standard errors reported in parentheses are computed using bootstraps. We find that senior drivers and lower-risk consumers tend to be less price sensitive, as expected. However, the estimated coefficient on the risk dummy has greater noise and is not statistically significantly different from zero. The estimated preferences for unobserved product attributes are generally similar across low- and high-risk groups. However, a notable difference is that lower-risk consumers appear to have a stronger preference for firms 5 and 11, while higher-risk consumers derive

[20]Conceptually, it is straightforward to include additional observable covariates in x, however, our demand estimation relies on the nonparametric estimation of price distributions conditional on the consumer's risk and observable characteristics for each firm separately. As the number of covariates grows, the amount of data required to accurately estimate the price distributions increases exponentially, which can compromise the precision of the estimates. Given the limited sample size in our data, we therefore chose to include only age and risk dummies in the consumer's preference parameters.

Table 3: Estimates of demand-side parameters

Parameters	Est.	Std. Err.			Est.	Std. Err.		
γ_0	2.43	(0.82)			3.02	(0.16)		
γ_1	-1.25	(0.14)			-1.17	(0.13)		
γ_2	0.71	(0.87)						
	Low-Risk		High-Risk		Low-Risk		High-Risk	
	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.
ξ_1	0		0		0		0	
ξ_2	-0.87	(0.04)	-0.89	(0.03)	-0.85	(0.03)	-0.90	(0.03)
ξ_3	-0.79	(0.04)	-0.86	(0.03)	-0.75	(0.03)	-0.86	(0.03)
ξ_4	-1.52	(0.09)	-1.65	(0.04)	-1.59	(0.03)	-1.65	(0.04)
ξ_5	-1.56	(0.03)	-1.86	(0.04)	-1.57	(0.03)	-1.86	(0.04)
ξ_6	-0.44	(0.08)	-0.53	(0.03)	-0.51	(0.03)	-0.52	(0.03)
ξ_7	-2.12	(0.04)	-2.08	(0.04)	-2.13	(0.04)	-2.08	(0.04)
ξ_8	-1.74	(0.06)	-1.31	(0.03)	-1.70	(0.03)	-1.32	(0.03)
ξ_9	-2.73	(0.06)	-2.52	(0.05)	-2.68	(0.05)	-2.52	(0.05)
ξ_{10}	-1.86	(0.03)	-2.01	(0.04)	-1.86	(0.03)	-2.01	(0.04)
ξ_{11}	0.73	(0.09)	0.60	(0.03)	0.65	(0.03)	0.61	(0.03)

Note: The left panel of this table reports the estimation results when $\gamma(x, \theta) = \gamma_0 + \gamma_1 OLD + \gamma_2 HR$, and the right panel presents the results when γ_2 is set to 0. Standard errors are reported in parentheses. In the estimation, premiums are represented in thousands of euros. The unobserved product heterogeneity for Firm 1 is normalized to 0 for both low- and high-risk groups.

greater utility from Firm 8. Since the price sensitivity parameter γ_2 is not statistically different from 0, we estimate an alternative version of the model with γ_2 set to 0. The results for this specification are presented in the right panel of Table 3. Since the parameter estimates are very similar across both specifications, we adopt the latter model for the remainder of the analysis.

5.2 Supply-Side Estimation Results

In the estimation, we assume that the firms receive Gaussian signals from $N(\theta, \sigma_j^2)$ and we estimate the standard deviation of the signal distribution σ_j. Note that $1/\sigma_j$ measures the precision of risk-rating technology in firm j. We report the estimates for firms' pricing coefficients (α_j, β_j), standard deviations of the signal distributions σ_j, and cost parameters (c_j, k_j) in Table 4, with the standard errors reported in parentheses.

Our estimates display a huge amount of firm heterogeneity along *all* dimensions. Comparing pricing coefficients across the firms shown in the upper panel of Table 4, we find that firms differ dramatically not only in their baseline premium (α_j), but also in how much they adjust premiums with respect to their risk ratings. For example, Firms 1, 8, and 10 charge a relatively lower baseline premium but their prices are much more sensitive to their risk

ratings, while Firms 3, 6, and 11 tend to do the opposite. Pricing coefficients (α_j, β_j) are complex equilibrium objects chosen optimally by firms. The differences in pricing strategies could be due to heterogeneity in firms' cost structures and information technology, as well as competition in the market.

Table 4: Estimates of supply-side parameters

Firm ID	Pricing Coef: Intercept α_j		Pricing Coef: Slope β_j	
	Est.	Std. Err.	Est.	Std. Err.
1	-131.26	(18.54)	1.85	(0.05)
2	-22.56	(23.85)	1.82	(0.07)
3	61.39	(19.70)	1.58	(0.06)
4	-109.61	(25.53)	1.48	(0.07)
5	42.38	(25.39)	1.63	(0.10)
6	54.16	(10.07)	1.21	(0.03)
7	8.95	(30.29)	1.49	(0.10)
8	-481.99	(85.21)	2.20	(0.18)
9	-75.90	(63.33)	1.94	(0.20)
10	-115.26	(54.71)	1.98	(0.17)
11	9.53	(5.60)	1.33	(0.02)

Firm ID	Signal Std. Dev. σ_j		Net Benefits c_j		Claim Efficiency k_j	
	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.
1	1365.66	(42.84)	669.27	(130.16)	1.82	(0.23)
2	978.99	(57.50)	597.00	(92.57)	2.18	(0.17)
3	903.02	(57.23)	494.66	(85.46)	1.90	(0.15)
4	1150.38	(56.27)	535.16	(85.49)	1.44	(0.16)
5	945.94	(74.64)	438.17	(89.49)	1.79	(0.18)
6	993.66	(32.15)	416.17	(86.78)	1.16	(0.12)
7	921.80	(78.00)	509.74	(92.97)	1.77	(0.17)
8	1565.57	(101.78)	1051.79	(167.13)	2.47	(0.32)
9	1042.61	(111.16)	355.07	(126.03)	1.56	(0.30)
10	1183.58	(90.21)	515.25	(128.21)	1.87	(0.28)
11	1023.53	(19.95)	847.46	(130.04)	1.61	(0.17)

We observe substantial differences in the precision of risk rating across firms, as measured by the standard deviation of their signal distributions (see the first column of the lower panel in Table 4). Based on our estimates, Firm 8 has the least advanced risk-rating technology. The standard deviation of its signal distribution is more than 70% higher than that of Firm 3, which has the best risk-rating technology based on our estimates. The heterogeneity in information precision has an important impact on consumers' equilibrium sorting patterns. Figure 4 presents a scatter plot of the ranks for information precision versus average consumer risks across firms. Insurers with low information precision (large σ) tend to attract high-risk consumers. The correlation coefficient between σ and average risk

within firm is 0.58.

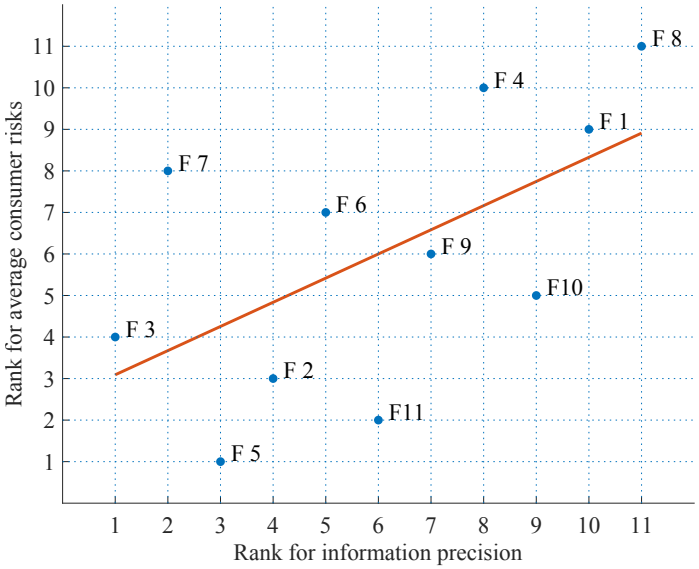


Figure 4: Who goes where: Rank correlation between information precision and average consumer risks across firms. Firms are ranked by information precision (highest to lowest on the x-axis) and average consumer risks (lowest to highest on the y-axis). The red solid line represents a linear fit between the two rankings. Firm IDs are displayed next to each dot.

Another interesting finding from Table 4 is that firms enjoy large net benefits from contracting with new customers. Our estimates of these net benefits range from 355 to 1052 euros per year and per customer across the competing firms. Although this paper does not provide a full-fledged dynamic model to account for consumers' switching behavior and firms' dynamic pricing strategies, we perform a back-of-the-envelope calculation of the future benefits firms may receive, based directly on the data. Let A_j denote the set of consumers who choose firm j in our sample. For each of these consumers, we observe the premium they pay, the actual claim costs, and whether they stay in firm j over the next T years. We can therefore compute the sum of the discounted future premiums averaged across all consumers in A_j, which we denote by $\overline{P}_j$. Specifically,

$$\overline{P}_j = \frac{\sum_{i \in A_j} \sum_t^T \delta^t P_{it} \mathbf{1}\{i \text{ stays in firm } j \text{ at } t\}}{|A_j|}, \tag{5.1}$$

where the discount factor δ is set to 0.95. We use a similar formula as in Equation (5.1) to compute the sum of discounted claim costs and contract sales, which we denote by $\overline{CS}_j$ and $\overline{N}_j$, respectively. Let mc_j denote firm j's marginal cost of managing a contract. The following equation computes the net benefit firms receive from contracting with a new

customer:

$$\underbrace{(\overline{P}_j - \overline{CS}_j k_j - \overline{N}_j mc_j)}_{\text{discounted value of future profits}} - mc_j = c_j. \tag{5.2}$$

Equation (5.2) isolates the firm's marginal cost mc_j from other factors. We report the estimates of marginal costs in Table 5.

Comparing our marginal cost estimates with the average premiums across firms, we find that the marginal cost is around 14%–45% of the premiums charged to consumers.[21] Another important finding from our estimation is that there is a moderate negative correlation between σ and mc, with a correlation coefficient of -0.53. This indicates that firms with lower information precision tend to have lower costs. Firms in equilibrium seemed to have developed different comparative advantages. Some firms rely on their more precise risk-rating technology to cream-skim low-risk consumers, while other firms can afford to serve the high-risk segment of the market due to a more efficient cost structure. Interestingly, we also find a small negative correlation between the two dimensions of the firm's cost structure. The correlation coefficient between mc and k is -0.20, suggesting that firms with higher marginal costs may be more efficient at processing claims.

Table 5: Estimates of marginal costs

Firm ID	Marginal Cost (€)	Perc. of Price (%)
1	185.80	35.74
2	191.90	34.44
3	210.99	37.11
4	125.63	29.37
5	213.58	42.13
6	175.55	39.14
7	199.17	37.80
8	78.99	13.67
9	250.35	45.33
10	208.89	39.88
11	92.66	21.50

5.3 Model Fit

Using the estimates in Table 4, we simulate the premiums offered by each firm and consumers' plan choices given the simulated price menu. Table 9 in the appendix presents the means and standard deviations of risk and premium within each firm, as well as the mar-

[21] This proportion aligns with findings in other markets; for instance, Chiappori et al. (2006) reported a 20% marginal administrative cost in the French auto insurance market. In addition, our estimate is close to the nationwide Italian average administrative cost obtained from balance sheet data of the insurance companies.

ket shares for low- and high-risk consumers, using both real and simulated data. Overall, our model matches the key moments of the data reasonably well. Our model-generated price distribution and market shares are very close to what we observe in the real dataset. As for the risk sorting pattern, we are able to match the mean and variance of consumers' risks for the majority of firms, and the ranking of average risk levels across firms is largely preserved. Our model underestimates the average risk level for Firm 8, which in reality attracts consumers with exceptionally high risks. However, it correctly identifies Firm 8 as having the highest risk level compared to all other firms. This underprediction primarily stems from the limited sample size of consumers with extremely high risks. Consequently, for this subgroup, the price distribution and consumers' sorting patterns are estimated with much lower precision.

Our estimation relies on the assumption that the premium is a monotonically increasing function of the firm's signal. We show that this assumption is self-consistent by validating the monotone relationship between the risk rating $E_j(\theta|\hat{\theta}_j, j)$—which is computed post-estimation—and the signal $\hat{\theta}_j$. In Figure 7 in the appendix, the risk rating, and therefore the premium, indeed monotonically increase with the signal of Firm 1, for both young and senior drivers. Similar patterns are observed for all other firms.

6 Counterfactuals

Using our structural estimates, we assess the equilibrium effects of establishing a centralized risk bureau that equalizes information access among competing firms. Specifically, we assume that the bureau collects signals from all firms. Based on these signals, the bureau would compute a posterior estimate of each consumer's risk,

$$E(\theta|\hat{\boldsymbol{\theta}}) = \int_\theta \theta f(\theta|\hat{\boldsymbol{\theta}}) d\theta = \frac{\int_\theta \theta \Big(\prod_j f(\hat{\theta}_j|\theta) \Big) f_0(\theta) d\theta}{\int_\theta \Big(\prod_j f(\hat{\theta}_j|\theta) \Big) f_0(\theta) d\theta}, \tag{6.1}$$

and make it accessible to all firms equally.

We want to highlight that this information aggregation process through the centralized bureau takes into account that the precision of the signals varies across firms. To see this point, consider a simple case where the prior distribution of θ is $N(\mu, \sigma^2)$, and the signals for firm j are drawn from $N(\theta, \sigma_j^2)$, then the posterior mean in Equation (6.1) can be

expressed in the following closed-form formula:

$$E(\theta|\hat{\boldsymbol{\theta}}) = \frac{\frac{\mu}{\sigma^2} + \sum_{j=1}^{J} \frac{\hat{\theta}_j}{\sigma_j^2}}{\frac{1}{\sigma^2} + \sum_{j=1}^{J} \frac{1}{\sigma_j^2}}.$$

Clearly, signals from different firms are weighted according to their information precision $(1/\sigma_j^2)$. Also note that the posterior mean in Equation (6.1) eliminates the winner's curse, as all available information has been exhausted.

In addition to our main policy experiment, we consider two additional counterfactual scenarios. The first one represents an efficiency benchmark, where the true risk type of each consumer is observed by all firms, so that the information asymmetry is completely eliminated. The second one is motivated by the recent discussions on privacy regulations.[22] To simulate a scenario where firms are required to limit their use of consumer data, we set the standard deviation of their signal distributions to be the largest currently observed in the market. Note that this policy also eliminates information asymmetries between competing firms, but at the same time, *reduces* the overall information availability in the market.

For each counterfactual scenario, we re-solve for the market equilibrium using the iterative procedure described in Appendix D. We also re-solve for the market equilibrium given model estimates in Section 5.2 and use that as the benchmark case for the rest of the analysis. We discuss the impact of counterfactual policies on consumer surplus, firm profits, equilibrium sorting patterns, and the overall market efficiencies in the following subsections.

6.1 Consumer Surplus

We report the average consumer surplus, premium, firm profit, HHI, and average cost under the baseline and three counterfactual scenarios (i.e., centralized risk bureau, observing true risk, and privacy regulation) in Table 6. The percentage changes relative to the baseline are reported in parentheses. To measure consumer surplus, we compute $-\gamma p_j + \xi_j$ for each individual, where j denotes the firm chosen by the consumer in equilibrium and p_j is the transaction price, and then we convert the utility into a monetary value expressed in euros. To evaluate the equilibrium consequence for consumers with different risk types, we report the average consumer surplus for high- and low-risk consumers separately.

[22] Researchers and policy makers have raised concerns regarding the legitimacy of using data from consumers for advertising, targeting, and price discrimination, etc. There is a growing body of literature on the economics of privacy and consumer protection (Campbell et al., 2015; Acquisti et al., 2016; Tucker, 2019; Jin and Wagman, 2021; Johnson, 2022; Johnson et al., 2023; Alcobendas et al., 2023; Goldfarb and Que, 2023).

Table 6: Counterfactual results: The impact of information policies

	Baseline	Centralized Risk Bureau	Observing True Risk	Privacy Regulation
Average CS (€)	-572.28	-497.04 (+13.15%)	-488.61 (+14.62%)	-548.45 (+4.16%)
Average CS: Low risk (€)	-512.69	-188.33 (+63.27%)	11.00 (+102.14%)	-514.03 (-0.26%)
Average CS: High risk (€)	-631.88	-805.74 (-27.52%)	-988.21 (-56.39%)	-582.87 (+7.76%)
Average premium (€)	471.68	392.78 (-16.73%)	381.11 (-19.20%)	449.22 (-4.76%)
Average profit (€)	596.96	552.82 (-7.39%)	550.70 (-7.75%)	579.03 (-3.00%)
HHI	2188.35	2171.98 (-0.75%)	2161.25 (-1.24%)	2198.18 (+0.45%)
Average cost (€)	705.47	687.40 (-2.56%)	676.95 (-4.04%)	707.35 (+0.27%)
Between-firm info asymmetries	Yes	No	No	No
Information availability		Increase	Increase	Decrease

Table 6 shows that eliminating information asymmetries between firms generally benefits consumers. When firms have equal access to the risk scores computed by the centralized risk bureau, consumer surplus increases by 13.15%, nearly matching the 14.62% gain observed under the infeasible benchmark where all firms know the true risk of the consumer. If all firms were to adopt the least advanced risk-rating technology under privacy regulation, consumer surplus would rise by 4.16%.

Examining the distributional effects of these counterfactual policies on consumer surplus, we find that establishing a centralized risk bureau that leverages all available market information primarily benefits low-risk consumers. Under this counterfactual scenario, low-risk individuals experience a 63.27% increase in consumer surplus, while high-risk consumers see a 27.52% reduction. Having access to the aggregated information about consumers' risk levels, firms can more precisely identify and price consumers based on their actual risk profiles. The differential impact on low- and high-risk consumers becomes even more pronounced when firms can observe each consumer's true risk. In this scenario, firms can perfectly price discriminate based on risk type, leading to a sharper increase in consumer surplus for low-risk individuals and a more significant decrease for high-risk consumers.

In contrast, the gains under privacy regulations are primarily enjoyed by high-risk consumers. As firms adopt less precise risk-rating technologies, the overall availability of information in the market declines, making it more difficult to accurately identify high-risk individuals. Consequently, high-risk consumers are less likely to be charged premiums that fully reflect their risk levels, leading to lower premiums and increased consumer surplus for them. This creates a form of cross-subsidization, where lower-risk consumers experience a slight reduction in their average consumer surplus.

6.2 Firm Profit

In all three counterfactual scenarios, we observe a decrease in average premium. When information asymmetries between firms is eliminated, as in all three cases, certain firms lose their market power gained from informational advantages and are forced to compete more aggressively on price. This increased competition contributes to the overall reduction in premiums.

Another key factor driving the price reduction is the elimination of the winner's curse, which occurs either when a centralized risk bureau is in place or when firms can observe consumers' true risk levels. In the baseline scenario, the prices offered to a consumer by different firms are fairly dispersed due to imprecise information. Consumers often receive

low-price offers from certain firms and tend to choose these, which can lead to a winner's curse. From a firm's perspective, lowering its price becomes ineffective in attracting more consumers under the winner's curse, as price dispersion means that even a reduced price is still likely to be higher than the lowest offer the consumer receives. Consequently, firms have little incentive to engage in aggressive Bertrand competition.

However, when a centralized bureau is in place (or when firms can perfectly observe risks), firms share a uniform evaluation of the consumer's risk, leading to more consistent and concentrated price offers across firms. Additionally, firms can accurately predict their competitors' prices. These factors make it easier for firms to undercut each other's prices, thereby intensifying price competition. Overall, we observe a 16.73%–19.20% reduction in the average premium when a centralized risk bureau is introduced or when firms can observe consumers' true risk levels.

As expected, the reduction in premiums leads to an overall decline in average profits for firms (see Table 6). However, the impact on profits varies significantly, depending on whether firms currently use more advanced technology or not. In Figure 5, we report percentage changes in profit under the two counterfactual scenarios (i.e., centralized risk bureau and privacy regulation) relative to the baseline for the 11 firms separately. Firms are ranked from the highest to the lowest risk-rating precision. Firms with less advanced risk-rating technology, such as Firms 1 and 8, benefit the most. In contrast, firms with more advanced risk-rating precision, like Firms 3, 7, 5, and 2, experience significant profit losses under both counterfactual scenarios, regardless of whether the policies improve or reduce overall risk-rating precision in the market.

In our empirical setting, larger firms do not necessarily possess the most advanced information technology. As a result, policies that eliminate heterogeneity in information technology might inadvertently strengthen the market power of these larger players, potentially leading to increased market concentration.

6.3 Market Efficiency

The public information policies we consider significantly alter firms' equilibrium pricing strategies, which in turn affect consumer choice patterns. Who goes where in the new equilibrium? Since our estimation results in Section 5.2 suggest that firms have comparative advantages along different cost dimensions, the matching between different types of consumers and insurers plays a crucial role in determining market efficiency, as we will analyze below.

To examine the effect of eliminating information heterogeneity on consumers' equilib-

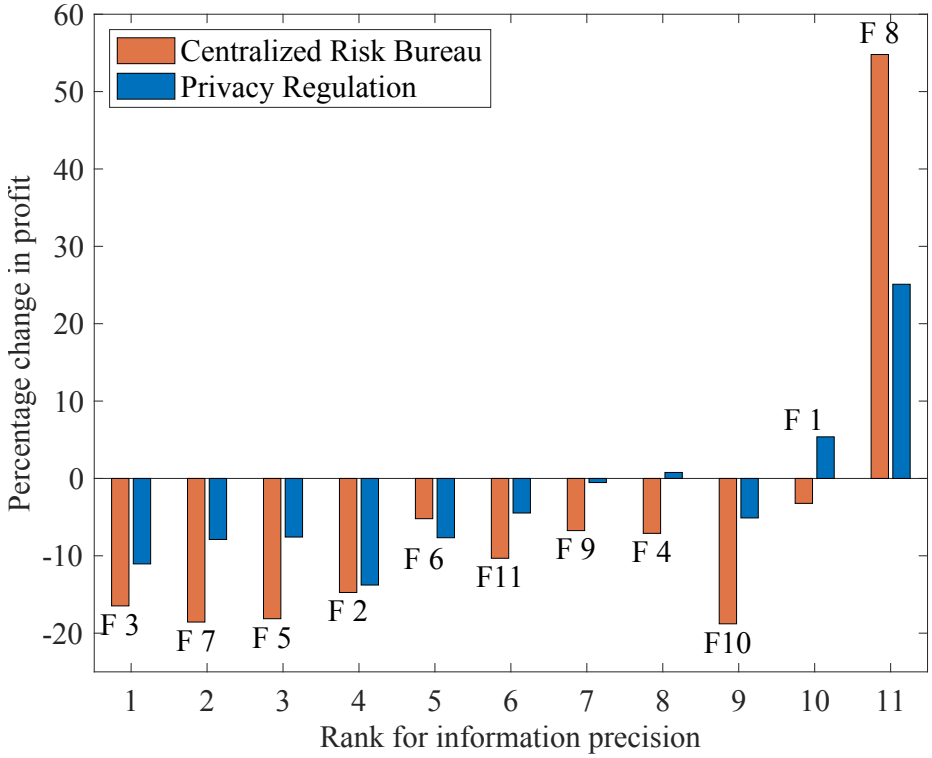


Figure 5: Differential impact on firms: Percentage change in profit relative to the baseline. Firms are ranked from the highest to the lowest risk-rating precision on the x-axis. The orange bars represent the percentage changes in profit under the centralized risk bureau scenario, while the blue bars indicate changes under the privacy regulation. Firm IDs are displayed next to each bar.

rium sorting patterns, we plot the average risk within firms for both the baseline and counterfactual scenarios in Figure 6. When all firms have equal access to risk scores computed by the centralized risk bureau, we observe significant consumer sorting, though the pattern differs from the baseline. Sorting is now primarily driven by specialization based on cost advantages rather than information asymmetries. Firms that attract high-risk consumers, such as Firms 4 and 6, are more efficient at processing claims, as reflected by their lower k_j values. This result is intuitive: when all firms can equally distinguish between high- and low-risk consumers, they adjust their pricing strategies based on comparative advantages in their cost structures. Firms that excel at processing claims can offer more competitive prices to high-risk consumers, thereby attracting more of them in equilibrium. The market segmentation driven by targeted pricing strategies becomes even more salient when firms can observe the true risk, as illustrated by the purple dashed line in Figure 6.

Another noteworthy observation is that consumer sorting nearly disappears under the privacy regulation scenario, as illustrated by the yellow line. The risk levels across firms become much more similar compared to the baseline. This result is again intuitive: when firms are equally ineffective at risk rating, consumers are likely to receive idiosyncratic prices, leading to a more random assignment of consumers to firms.

We compute the cost to insure a consumer (i.e., $mc_j+k_j\theta$), averaged over all individuals, as a measure of market efficiency. The last row of Table 6 shows that under the centralized risk bureau scenario, the average cost is reduced by 2.56% compared to the baseline. This cost reduction from consumer reallocation is approximately 18 euros per contract per year, equivalent to about 4% of the annual premiums paid by consumers. Given the size of the Italian auto insurance market (31 million contracts underwritten in 2018), the overall impact of this cost reduction is substantial. The efficiency gains under the centralized risk bureau scenario nearly match that of the infeasible benchmark, underscoring the effectiveness of this policy.

Our findings highlight that better utilization and aggregation of existing market information can reduce information asymmetries between buyers and sellers. Even in mandatory insurance markets, this reduction can still improve welfare through more efficient matching between insurers and insurees.

7 Conclusions

Our paper develops a novel empirical framework for studying imperfect competition in selection markets when firms have *heterogeneous information* about consumers, differ in cost structures, and offer differentiated products. We build a theoretical model where firms re-

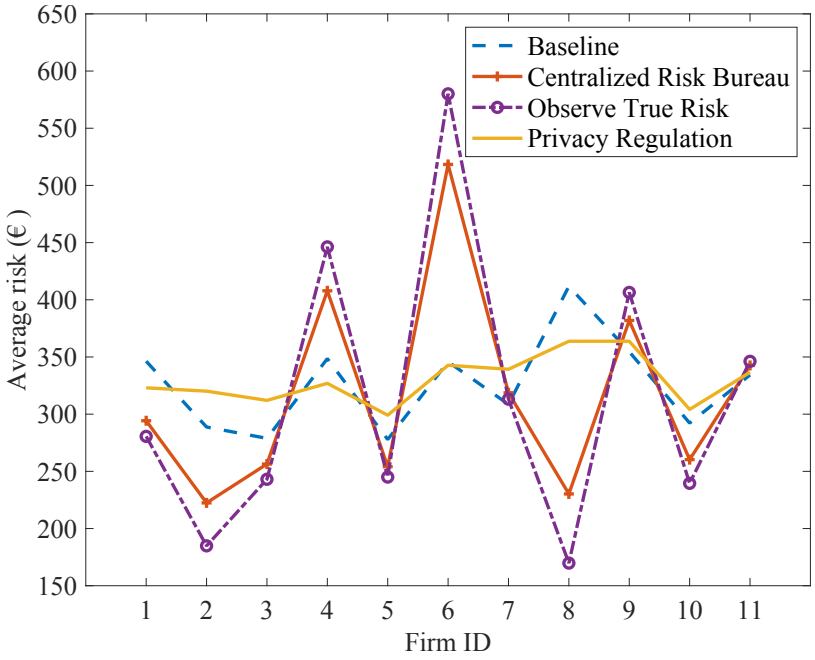


Figure 6: Sorting patterns: Average risk levels among consumers (measured in euros) within each firm under the baseline and three counterfactual scenarios.

ceive private signals about consumers' risk types and choose their optimal pricing strategies to maximize their profit, taking into account their opponents' strategies and demand-side responses. In our model, firms' inferences about consumers' risk types depend not only on their signals but also on equilibrium sorting patterns.

We introduce novel econometric techniques to identify demand parameters, as well as firm-specific information precision and cost structures. Applying the method to a representative sample of Italian auto insurance contracts with associated claims from 2013 to 2021, we find substantial differences in the precision of risk rating across firms; moreover, firms with lower information precision tend to have lower costs. This suggests that firms have developed different comparative advantages. Some firms focus on improving their risk-rating technology so that they can cream-skim low-risk consumers, while other firms are more efficient at underwriting contracts and/or liquidating claims, leading to a competitive equilibrium in which a variety of pricing strategies co-exist.[23] The methodology we develop extends beyond auto insurance; it is applicable to a wide range of selection markets, including credit, other types of insurance, and labor markets.

[23]It is an open and intriguing question to study how insurance companies allocate the complex task of pricing across various subunits within the organization (see discussions in Hortaçsu et al. (2024) regarding the airline industry). For example, different subunits may be tasked with estimating risk, liquidating claims, and underwriting contracts. Exploring how this internal structure relates to cost and information heterogeneity across competing firms presents an interesting avenue for future research.

We evaluate the equilibrium effects of a public information policy where insurers' risk estimates are aggregated and made public through a centralized bureau. This policy significantly reduces prices by increasing competition and eliminating the winner's curse, which in turn boosts consumer surplus by 13%, nearly matching the gains observed under the efficiency benchmark, where firms have full knowledge of consumers' true risk levels. We also observe interesting distributional effects. The centralized risk bureau primarily benefits low-risk consumers, in contrast to the scenario motivated by privacy regulations, where all firms adopt the least-advanced information technology. Lastly, the centralized bureau policy improves the matching efficiency between insurers and insurees, reducing average costs by 2.56%, or equivalently 18 euros per contract.

Our findings have broad implications for public information regulations in the era of big data, artificial intelligence, and machine learning. Firms with superior information can gain market power, which may harm competition and consumer welfare. As technology advances, the gap between firms in both information access and data analysis capabilities widens, creating significant barriers for new entrants. Our results confirm that public policies equalizing information access can enhance competition and improve overall consumer welfare, though with complex distributional effects. However, a caveat is that such policies could erode the competitive advantages of technologically advanced firms, including smaller and newer firms in our sample, potentially diminishing incentives for long-term innovation (Goldfarb and Tucker, 2012; Jia et al., 2018).

Finally, our paper mainly focuses on static price competition with information and cost heterogeneity in selection markets. Our demand model follows a more standard approach and does not explicitly account for search frictions. While earlier studies (Honka, 2014; Honka et al., 2017; Allen et al., 2019) have shown limited consumer search for financial products, the recent rise of price comparison websites for auto insurance (such as US News, Experian, the Zebra), along with the increasing use of online purchasing platforms, suggests that search cost in our context might not be very high. More flexible demand models (e.g., allowing consumer switching, search frictions, a limited consideration set, and multi-dimensional private information), as well as several other important aspects of competition, including dynamic pricing, endogenous contract designs, are promising avenues for future research.

A Additional Figures and Tables

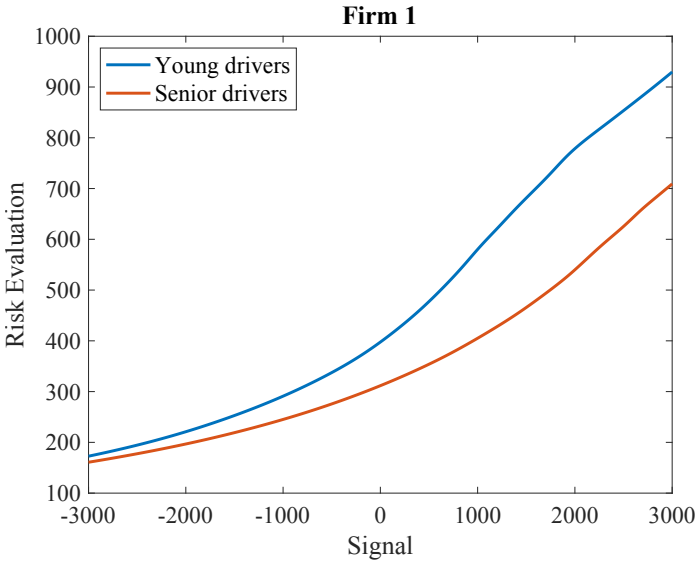


Figure 7: Risk rating versus signal for Firm 1: The y-axis represents the risk rating, and the x-axis denotes the signal. The blue and red curves represent young and senior drivers, respectively.

Firm A	Firm B	Firm C	Firm D	Firm E
Years of non-circulation of the vehicle	Safe Driving & Savings Clause	Brand-model of the vehicle	Credit information of the vehicle owner (data derived from the census cell of residence)	Census cells
Vehicle weight	License Seniority / Type of Person / CU Class	Age of the vehicle at the time of purchase	Use	Special uses (Driving school, Leasing, etc.)
Brand-model of the vehicle	Infocar Code	Occupation	Trailer towing	Vehicle adapted for reduced motor capacity
Years of vehicle ownership	Vehicle Age at Purchase	Marital status	Family parameter (presence of other car insurance policies within the family unit)	Number of driving wheels
Purchase of a new/used vehicle	Increase for vehicles already insured with NA in the Current Year	Seniority in obtaining a driving license	Safe driving	Body type
Occupation	Waiver of Right to Recourse Clauses	Kilometers per year	Discount program for youth	Safety devices
			Temporary insurance	Brand
			Vehicle size category	Transported substances

Table 7: Selected Variables Used by Five Major Auto Insurance Companies (Not Included in Our Data)

VARIABLES	Firm 1	Firm 2	Firm 3	Firm 4	Firm 5	Firm 6	Firm 7	Firm 8	Firm 9	Firm 10	Firm 11
Age	-1.0425***	-0.7190***	-1.4790***	-0.2519	-0.4040**	-0.7867***	-1.4040***	-1.6740***	-0.9280***	-1.0231***	-0.5228***
	(0.0889)	(0.1504)	(0.1313)	(0.1634)	(0.1734)	(0.0743)	(0.2255)	(0.2091)	(0.3133)	(0.2088)	(0.0470)
BM class	19.9390***	24.9547***	26.5323***	21.7531***	31.1112***	21.8711***	23.5842***	13.6431***	31.5233***	11.9072***	23.0984***
	(0.8676)	(1.1211)	(1.1948)	(1.9162)	(1.5436)	(0.4909)	(1.5086)	(2.1490)	(2.3300)	(2.0083)	(0.4230)
Man	-8.9978***	-7.8429*	-3.3121	-21.3611***	-6.4063	4.0793**	7.2525	-1.9419	-10.1827	-9.2817	-1.1208
	(2.3761)	(4.0330)	(3.7137)	(4.6269)	(4.9098)	(1.8624)	(6.5576)	(5.6881)	(8.5708)	(6.1545)	(1.2884)
# of Acc.=1	124.3281***	148.8613***	105.8185***	76.6681***	112.4740***	112.0255***	125.3370***	119.6688***	145.8684***	182.2420***	120.8346***
	(4.0380)	(7.0657)	(5.7095)	(6.5595)	(8.7984)	(3.0135)	(9.9444)	(9.2291)	(16.5050)	(12.4197)	(2.0157)
# of Acc.=2	283.4361***	306.9949***	259.0145***	198.5605***	298.4881***	344.4711***	273.7250***	278.8861***	304.9190***	426.8850***	231.0981***
	(12.3873)	(15.6679)	(16.2553)	(21.9686)	(36.9753)	(24.8148)	(28.0848)	(26.4159)	(61.3592)	(47.9834)	(6.0845)
# of Acc.=3	110.0135***	108.2404***	139.7064***	109.2576***	147.9388***	135.6132***	150.3077***	103.5709***	98.6608***	129.9741***	133.8557***
	(2.8790)	(5.6049)	(5.2093)	(5.2251)	(8.1627)	(3.7884)	(9.5861)	(6.0418)	(11.6335)	(9.1320)	(2.1983)
Medium city	1.5995	-13.2042**	-0.4254	-1.8326	2.5681	21.4227***	2.2266	4.0654	-3.9565	3.2213	14.7086***
	(4.0300)	(6.0853)	(6.2695)	(5.4826)	(7.1698)	(3.3934)	(8.6543)	(9.8588)	(18.1009)	(8.3043)	(2.3882)
Big city	34.2326***	45.3041***	56.9994***	54.9616***	39.4402***	17.0140***	49.5132***	54.8911***	83.8861***	52.4244***	11.6583***
	(2.5860)	(4.3675)	(3.8968)	(4.5677)	(5.0926)	(2.0113)	(7.3400)	(6.3508)	(10.2456)	(6.7350)	(1.4520)
Vehicle age	-1.4443***	-0.4925	2.1354***	2.2253***	2.6130***	5.3059***	2.4714***	-2.6089***	7.4976***	-0.6330	0.9280***
	(0.2401)	(0.4101)	(0.3805)	(0.4779)	(0.4957)	(0.2043)	(0.6580)	(0.5927)	(0.9893)	(0.6800)	(0.1438)
Power	1.3433***	1.0896***	1.5654***	0.9985***	1.2255***	1.0370***	1.0561***	1.3188***	1.8501***	1.2230***	1.0667***
	(0.0704)	(0.1019)	(0.1052)	(0.1630)	(0.1336)	(0.0625)	(0.1482)	(0.1589)	(0.1946)	(0.2192)	(0.0421)
One installment	-85.0581***	-111.0775***	-80.9882***	-98.9312***	-89.2521***	-144.9128***	-79.9680***	-114.8558***	-79.8162***	-96.4682***	-113.2898***
	(2.4533)	(4.0605)	(3.7805)	(4.3576)	(4.7135)	(3.3021)	(6.4574)	(5.9871)	(8.9444)	(5.8693)	(2.0658)
Constant	482.3752***	599.9441***	477.0323***	366.5853***	479.5029***	510.0879***	498.6533***	573.6218***	440.7018***	589.3077***	548.3449***
	(9.6316)	(28.5192)	(14.4716)	(51.7123)	(21.9729)	(20.3601)	(34.4345)	(21.5220)	(73.2541)	(39.1469)	(6.8714)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Optional clauses	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Vehicle char.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	20,837	8,077	8,088	5,509	3,953	15,993	2,531	4,302	1,357	2,996	50,949
R-squared	0.3921	0.3909	0.4739	0.4432	0.4774	0.5887	0.4541	0.3879	0.5062	0.4448	0.5187

Table 8: Predicting premium

Table 9: Model fit: Comparing data moments with simulated results using model estimates

Panel (A) Data Moments						
Firm ID	Risk (€)		Premium (€)		Market Share	
	Mean	Std. Dev.	Mean	Std. Dev.	Low Risk	High Risk
1	352.71	433.72	519.81	209.46	0.16	0.17
2	318.53	416.51	557.12	221.44	0.07	0.06
3	320.07	432.11	568.59	216.07	0.07	0.06
4	363.97	461.81	427.78	199.55	0.04	0.04
5	285.56	401.07	506.97	201.39	0.04	0.03
6	326.86	423.73	448.57	173.50	0.12	0.13
7	346.70	426.08	526.93	209.00	0.02	0.02
8	481.29	511.30	577.66	227.70	0.03	0.04
9	324.17	397.91	552.28	225.59	0.01	0.01
10	322.36	435.80	523.75	212.33	0.03	0.02
11	316.25	412.78	431.04	198.01	0.42	0.40

Panel (B) Simulated Moments Using Model Estimates						
Firm ID	Risk (€)		Premium (€)		Market Share	
	Mean	Std. Dev.	Mean	Std. Dev.	Low Risk	High Risk
1	344.88	435.81	521.58	217.57	0.17	0.18
2	338.47	441.12	562.14	231.39	0.06	0.06
3	321.26	430.77	570.24	225.43	0.07	0.06
4	341.69	435.02	422.61	205.36	0.04	0.04
5	300.94	424.38	506.62	213.28	0.04	0.03
6	329.18	411.70	445.91	180.41	0.12	0.13
7	344.37	443.66	521.90	222.60	0.02	0.02
8	438.56	470.63	577.70	240.07	0.02	0.04
9	335.09	412.38	548.71	245.64	0.01	0.01
10	303.08	420.00	525.59	226.46	0.03	0.02
11	321.11	418.47	432.19	203.23	0.41	0.40

B Iterative Algorithm

Given the fixed point problem defined in Equations (4.1) and (4.2), we propose an iterative procedure to solve for the *offered* price distributions for all firms as functions of demand parameters. We nest the contraction mapping for mean utilities in Berry et al. (1995) in our algorithm to obtain the vector of unobserved product attributes $\boldsymbol{\xi}$ that matches aggregate market shares. Our algorithm works as follows:

1. Fix the value of γ.
2. Set the initial value: $\xi_j^1 = 0$, and $Pr^1(j|p_j = p, \theta) = \exp(-\gamma p)$ for all j.[24]
3. At the r-th iteration, denote the vector of unobserved product attributes by $\boldsymbol{\xi}^r = (\xi_1^r, \xi_2^r, \cdots, \xi_J^r)$. Let $Pr^r(j|p_j = p, \theta)$ denote the probability of consumers being sorted into firm j given price p and type θ for $j = 1, 2, \cdots, J$.
4. Compute the offered price distribution for the r-th iteration using Equation (4.1):

$$g_j^r(p|\theta) = \frac{\tilde{g}_j(p|\theta)/Pr^r(j|p_j = p, \theta)}{\int_p \tilde{g}_j(p|\theta)/Pr^r(j|p_j = p, \theta)dp}. \tag{B.1}$$

5. Update $\boldsymbol{\xi}^r$ using aggregate market shares. Given γ and the offered price distributions $\boldsymbol{g}^r = (g_1^r, g_2^r, \cdots, g_J^r)$, $\boldsymbol{\xi}^{r+1}$ solves the following system of equations:

$$\begin{aligned}
\hat{s}_1 &= s_1(\boldsymbol{\xi}^{r+1}, \gamma, \boldsymbol{g}^r), \\
\hat{s}_2 &= s_2(\boldsymbol{\xi}^{r+1}, \gamma, \boldsymbol{g}^r), \\
&\vdots \\
\hat{s}_J &= s_J(\boldsymbol{\xi}^{r+1}, \gamma, \boldsymbol{g}^r),
\end{aligned}$$

where $\hat{s}_j$ represents the observed market share for firm j. We denote by $s_j(\boldsymbol{\xi}^{r+1}, \gamma, \boldsymbol{g}^r)$ the model-implied market share given by the following equation:

$$s_j(\boldsymbol{\xi}^{r+1}, \gamma, \boldsymbol{g}^r) = \int_\theta \int_{\boldsymbol{p}} \frac{\exp(-\gamma p_j + \xi_j^{r+1})}{\sum_{j'} \exp(-\gamma p_{j'} + \xi_{j'}^{r+1})} \left(\prod_{j'} g_{j'}^r(p_{j'}|\theta) \right) f_0(\theta) d\boldsymbol{p} d\theta. \tag{B.2}$$

[24] The initial value for choice probabilities is the numerator in the demand model (when $\xi_j = 0$). It can be viewed as a coarse approximation to the selection probability, assuming changing firm j's price would only impact the numerator, but not the denominator. The fact that it is not a well-defined probability measure does not matter, because it only serves as a weight and will be normalized when we update the offered price distribution using Equation (4.1). Since the operator is a contraction, the choices of initial values are less important. In our estimation, we try using different initial values for choice probabilities, and they all converge to the same results.

This step is similar to the contraction mapping for the mean utilities in Berry et al. (1995). Using the same iterative procedure as in Berry et al. (1995), we solve for the vector of unobserved product attributes $\boldsymbol{\xi}^{r+1}$ from observed market shares (ξ_1 is normalized to 0 without loss).

6. Given $\boldsymbol{\xi}^{r+1}$, we update the selection probabilities for all firms using Equation (4.2):

$$Pr^{r+1}(j|p_j = p, \theta) \tag{B.3}$$
$$= \int_{\boldsymbol{p}_{-j}} \frac{\exp(-\gamma p + \xi_j^{r+1})}{\exp(-\gamma p + \xi_j^{r+1}) + \sum_{j' \neq j} \exp(-\gamma p_{j'} + \xi_{j'}^{r+1})} \left(\prod_{j' \neq j} g_{j'}^r(p_{j'}|\theta) \right) d\boldsymbol{p}_{-j}.$$

7. Advance to the next iteration with $\boldsymbol{\xi}^{r+1}$ and $Pr^{r+1}(j|p_j = p, \theta)$ for all j.

8. Iterate until $\boldsymbol{g}^r$ converges. In the limit, for a fixed price sensitivity parameter γ, we obtain (1) the offered price distribution faced by consumers $g_j(p|\theta)$ for all firms, (2) the probability that consumers are selected into firm j, $Pr(j|p_j = p, \theta)$, and (3) the vector of unobserved product attributes $\boldsymbol{\xi}$ that match the market shares.

C Identification of Signal Variance

We consider the identification of the variance of the signal distribution when the signals are drawn from a normal distribution. In this case, the density function of the signal distribution for firm j takes the following form:

$$f_j(\hat{\theta}_j|\theta) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(\hat{\theta}_j - \theta)^2}{2\sigma_j^2}\right), \tag{C.1}$$

where σ_j^2 is the variance of the signal distribution to be identified.

We first derive the posterior density of $\hat{\theta}_j$ for those who self-select into firm j. Let $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_J)$ denote the vector of signals received by all firms. Let $\hat{\boldsymbol{\theta}}_{-j}$ denote the vector of signals excluding firm j's signal. To simplify notation, we use $f(\hat{\boldsymbol{\theta}}|\theta)$ and $f(\hat{\boldsymbol{\theta}}_{-j}|\theta)$, respectively, to denote the densities of $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{-j}$ conditional on θ.

$$\begin{aligned} f(\hat{\theta}_j|\theta, j) &= \frac{\int_{\hat{\boldsymbol{\theta}}_{-j}} Pr(j|\hat{\boldsymbol{\theta}}) f(\hat{\boldsymbol{\theta}}|\theta) f_0(\theta) d\hat{\boldsymbol{\theta}}_{-j}}{\int_{\hat{\boldsymbol{\theta}}} Pr(j|\hat{\boldsymbol{\theta}}) f(\hat{\boldsymbol{\theta}}|\theta) f_0(\theta) d\hat{\boldsymbol{\theta}}} \\ &= \frac{\exp(-\gamma p_j(\hat{\theta}_j) + \xi_j) f_j(\hat{\theta}_j|\theta) \left[\int_{\hat{\boldsymbol{\theta}}_{-j}} \frac{f(\hat{\boldsymbol{\theta}}_{-j}|\theta) f_0(\theta)}{\sum_{j'=1}^{J} \exp(-\gamma p_{j'}(\hat{\theta}_{j'}) + \xi_{j'})} d\hat{\boldsymbol{\theta}}_{-j}\right]}{\int_{\hat{\theta}_j} \exp(-\gamma p_j(\hat{\theta}_j) + \xi_j) f_j(\hat{\theta}_j|\theta) \left[\int_{\hat{\boldsymbol{\theta}}_{-j}} \frac{f(\hat{\boldsymbol{\theta}}_{-j}|\theta) f_0(\theta)}{\sum_{j'=1}^{J} \exp(-\gamma p_{j'}(\hat{\theta}_{j'}) + \xi_{j'})} d\hat{\boldsymbol{\theta}}_{-j}\right] d\hat{\theta}_j} \end{aligned} \tag{C.2}$$

$$\approx \frac{\exp(-\gamma p_j(\hat{\theta}_j)) f_j(\hat{\theta}_j|\theta)}{\int_{\hat{\theta}_j} \exp(-\gamma p_j(\hat{\theta}_j)) f_j(\hat{\theta}_j|\theta) d\hat{\theta}_j}. \tag{C.3}$$

Denote the term in square brackets in Equation (C.2) by $\Delta(\hat{\theta}_j, \theta)$. Since it is inside of the integral in the denominator, $\Delta(\hat{\theta}_j, \theta)$ cannot be cancelled out. However, if firm j has a small market share, the effect of $\hat{\theta}_j$ on $\Delta(\hat{\theta}_j, \theta)$ is small, and therefore we can treat this term as if it does not depend on $\hat{\theta}_j$. When firm j is large, then as long as $\hat{\theta}_j$ has a relatively small variance, the overall impact of $\Delta(\hat{\theta}_j, \theta)$ varying with $\hat{\theta}_j$ is small in the denominator. With this approximation, we obtain Equation (C.3).

If the risk-rating term $E(\theta|\hat{\theta}_j, j)$ can be well approximated by a linear function of the signal $\hat{\theta}_j$,[25] i.e., $E(\theta|\hat{\theta}_j, j) \approx a_j + b_j\hat{\theta}_j$, we can further simplify the posterior distribution

[25]We empirically verify this assumption in Figure 7, where we plot the relationship between risk rating $E(\theta|\hat{\theta}_j, j)$ and signal $\hat{\theta}_j$ after we estimate the model. This figure suggests that the assumption that risk rating can be approximated by a linear function of the signal is reasonable.

of $\hat{\theta}_j$ to be

$$f(\hat{\theta}_j|\theta, j) \approx \frac{\exp(-\gamma\beta_j b_j \hat{\theta}_j) f_j(\hat{\theta}_j|\theta)}{\int_{\hat{\theta}_j} \exp(-\gamma\beta_j b_j \hat{\theta}_j) f_j(\hat{\theta}_j|\theta) d\hat{\theta}_j}$$
$$= \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(\hat{\theta}_j - (\theta - \gamma\beta_j b_j \sigma_j^2))^2}{2\sigma_j^2}\right). \tag{C.4}$$

Equation (C.4) implies that after selection, the signal still approximately follows a normal distribution with the same variance (σ_j^2) as before. However, the mean of the posterior normal distribution shifts to $\theta - \gamma\beta_j b_j \sigma_j^2$, which is lower than the original mean θ. This is intuitive as consumers who receive lower prices from j are more likely to self-select into the firm.

The following lemma provides the key identification argument for the signal variance. We focus on the joint distribution of premiums and risk types of consumers selected into firm j. To simplify notation, let p_j and θ_j be the premiums and risk types of consumers selected into firm j. The random variable θ_j admits distribution m_j, which can be directly estimated from the data.

Lemma 1. *Given m_j, the correlation $corr(p_j, \theta_j)$ is monotonically decreasing in σ_j.*

Proof. Since p_j and θ_j are positively correlated (empirically verifiable),

$$corr(p_j, \theta_j) = \frac{cov(p_j, \theta_j)}{\sqrt{var(\theta_j) var(p_j)}} = \sqrt{\frac{cov^2(p_j, \theta_j)}{var(p_j) var(\theta_j)}} = \sqrt{\frac{var(E(\theta_j|\hat{\theta}_j))}{var(\theta_j)}}. \tag{C.5}$$

The last equality in Equation (C.5) holds because, by Equation, (4.5)

$$var(E(\theta_j|\hat{\theta}_j)) = \frac{var(p_j)}{\beta_j^2} = \frac{var(p_j)}{[var(p_j)/cov(p_j, \theta_j)]^2} = \frac{cov^2(p_j, \theta_j)}{var(p_j)}.$$

Thus, for a fixed type distribution m_j within firm j, it is equivalent to show $var(E(\theta_j|\hat{\theta}_j))$ is monotonically decreasing in σ_j.

We first show that $var(E(\theta_j|\hat{\theta}_j)) = var(E(\theta_j|\tilde{\theta}_j))$, where $\tilde{\theta}_j \equiv \hat{\theta}_j + \gamma\beta_j b_j \sigma_j^2$. Let ϕ

denote the density of $\hat{\theta}_j$. We get that

$$\begin{aligned} var(E(\theta_j|\hat{\theta}_j)) &= \int (E(\theta_j|\hat{\theta}_j = x))^2 \phi(x)dx - \mu_j^2 \\ &= \int (E(\theta_j|\tilde{\theta}_j = x + \gamma\beta_j b_j \sigma_j^2))^2 \phi(x)dx - \mu_j^2 \\ &= \int (E(\theta_j|\tilde{\theta}_j = y))^2 \phi(y - \gamma\beta_j b_j \sigma_j^2)dy - \mu_j^2 \\ &= var(E(\theta_j|\tilde{\theta}_j)). \end{aligned}$$

Next, we show that $var(E(\theta_j|\tilde{\theta}_j))$ is monotonically decreasing in σ_j. Note that by Equation (C.4), $\tilde{\theta}_j \sim \mathcal{N}(\theta, \sigma_j^2)$. Consider a new normal random variable $\tilde{\theta}'_j \sim \mathcal{N}(\theta, \sigma_j'^2)$ with a larger variance $\sigma_j'^2 > \sigma_j^2$. Since both $\tilde{\theta}_j$ and $\tilde{\theta}'_j$ follow a normal distribution, we get that

$$\tilde{\theta}'_j = \tilde{\theta}_j + \eta,$$

where η is an independent and normally distributed random variable. By independence, the distribution of $E(\theta_j|\tilde{\theta}_j)$ is the same as the distribution of $E(\theta_j|\tilde{\theta}'_j, \eta)$, so $var(E(\theta_j|\tilde{\theta}_j)) = var(E(\theta_j|\tilde{\theta}'_j, \eta))$. Moreover,

$$var(E(\theta_j|\tilde{\theta}'_j, \eta)) > var(E(\theta_j|\tilde{\theta}'_j)),$$

because on the left-hand side of the inequality we project θ_j onto a larger space. We therefore obtain the desired result that $var(E(\theta_j|\tilde{\theta}_j)) > var(E(\theta_j|\tilde{\theta}'_j))$. □

For a fixed distribution of risk types within a firm, higher variance in the firm's signal distribution leads to a lower correlation between the premiums and risk types. Intuitively, when the signal distribution is very informative, consumers' risk types are better reflected in their premiums, and vice versa.[26] Since the type distribution m_j and $corr(p_j, \theta_j)$ can be easily estimated from the data, the one-to-one mapping between $corr(p_j, \theta_j)$ and σ_j^2 in Lemma 1 uniquely pins down the signal variance.

[26]To see this from another angle, the proof of Lemma 1 shows that $var(E(\theta_j|\hat{\theta}_j))$ decreases with σ_j^2. When σ_j^2 is small, the signals received by the firm are precise, and therefore the posterior mean of θ is very sensitive to the signals. As a result, the variance of the posterior mean is large. By contrast, when σ_j^2 is large, the signals received by the firm are not informative, and therefore the posterior mean is similar across different signals. As a result, the variance of $E(\theta|\hat{\theta}_j)$ is small.

D Solving the Equilibrium for Counterfacual Analysis

To evaluate counterfactual policies, we need to solve the market equilibrium for any given set of model parameters. We propose an iterative procedure to solve for all firms' pricing coefficients $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, the equilibrium expectation $E_j(\theta|\hat{\theta}_j, j)$, and the selection probabilities $Pr(j|\hat{\theta}_j, \theta)$ for all j. The iterative algorithm for solving the equilibrium works as follows:

1. In the outer loop, we solve for firms' pricing coefficients $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Denote the pricing coefficients at the r-th iteration by $(\boldsymbol{\alpha}^r, \boldsymbol{\beta}^r)$.

2. Given the model primitives $(\gamma, \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{c}, \boldsymbol{k})$ and pricing coefficients $(\boldsymbol{\alpha}^r, \boldsymbol{\beta}^r)$, we solve the r-th equilibrium sorting pattern in an inner loop:

   (1) Set the initial value $E_j^{r,0}(\theta|\hat{\theta}_j, j) = E_j^{r-1}(\theta|\hat{\theta}_j, j)$.

   (2) Compute the premium offered by each firm using the following equation:

   $$p_j^{r,0}(\hat{\theta}_j) = \alpha_j^r + \beta_j^r E_j^{r,0}(\theta|\hat{\theta}_j, j).$$

   (3) Compute the choice probabilities:

   $$Pr_j^{r,0}(j|\hat{\theta}_j, \theta) = \int_{\hat{\boldsymbol{\theta}}_{-j}} \frac{\exp(-\gamma p_j^{r,0}(\hat{\theta}_j) + \xi_j)}{\sum_{j'=1}^{J} \exp(-\gamma p_{j'}^{r,0}(\hat{\theta}_{j'}) + \xi_{j'})} f(\hat{\boldsymbol{\theta}}_{-j}|\theta; \boldsymbol{\sigma}_{-j}) d\hat{\boldsymbol{\theta}}_{-j}.$$

   (4) Update the expectation by

   $$E_j^{r,1}(\theta|\hat{\theta}_j, j) = \frac{\int_\theta \theta Pr^{r,0}(j|\hat{\theta}_j, \theta) f_j(\hat{\theta}_j|\theta; \sigma_j) f_0(\theta) d\theta}{\int_\theta Pr^{r,0}(j|\hat{\theta}_j, \theta) f_j(\hat{\theta}_j|\theta; \sigma_j) f_0(\theta) d\theta}.$$

   (5) Iterate this process until the expectation converges. Denote the limits by $E_j^r(\theta|\hat{\theta}_j, j)$ and $Pr_j^r(j|\hat{\theta}_j, \theta)$. These represent the r-th equilibrium sorting pattern.

3. Having obtained $E_j^r(\theta|\hat{\theta}_j, j)$ and $Pr_j^r(j|\hat{\theta}_j, \theta)$ from the inner loop, we now compute the first-order derivatives of the profit function with respect to pricing coefficients

(α_j^r, β_j^r) for all firms. Specifically,

$$\begin{aligned}
\frac{\partial \pi_j^r(\boldsymbol{\alpha}^r, \boldsymbol{\beta}^r)}{\partial \alpha_j^r} &= \int_\theta \int_{\hat{\boldsymbol{\theta}}} Pr^r(j|\hat{\boldsymbol{\theta}}) f(\hat{\boldsymbol{\theta}}|\theta) f_0(\theta) d\hat{\boldsymbol{\theta}} d\theta \\
&+ \quad \int_\theta \int_{\hat{\boldsymbol{\theta}}} (\alpha_j^r + \beta_j^r \theta + c_j - k_j \theta) \frac{\partial Pr^r(j|\hat{\boldsymbol{\theta}})}{\partial \alpha_j^r} f(\hat{\boldsymbol{\theta}}|\theta) f_0(\theta) d\hat{\boldsymbol{\theta}} d\theta, \\
\frac{\partial \pi_j^r(\boldsymbol{\alpha}^r, \boldsymbol{\beta}^r)}{\partial \beta_j^r} &= \int_\theta \int_{\hat{\boldsymbol{\theta}}} \theta Pr^r(j|\hat{\boldsymbol{\theta}}) f(\hat{\boldsymbol{\theta}}|\theta) f_0(\theta) d\hat{\boldsymbol{\theta}} d\theta \\
&+ \quad \int_\theta \int_{\hat{\boldsymbol{\theta}}} (\alpha_j^r + \beta_j^r \theta + c_j - k_j \theta) \frac{\partial Pr^r(j|\hat{\boldsymbol{\theta}})}{\partial \beta_j^r} f(\hat{\boldsymbol{\theta}}|\theta) f_0(\theta) d\hat{\boldsymbol{\theta}} d\theta.
\end{aligned}$$

4. Update the pricing coefficients as follows:

$$\begin{aligned}
\alpha_j^{r+1} &= \alpha_j^r + \Delta\alpha_j^r \mathbf{1}\left\{ \frac{\partial \pi_j^r(\boldsymbol{\alpha}^r, \boldsymbol{\beta}^r)}{\partial \alpha_j^r} \geq 0 \right\} - \Delta\alpha_j^r \mathbf{1}\left\{ \frac{\partial \pi_j^r(\boldsymbol{\alpha}^r, \boldsymbol{\beta}^r)}{\partial \alpha_j^r} < 0 \right\}, \\
\beta_j^{r+1} &= \beta_j^r + \Delta\beta_j^r \mathbf{1}\left\{ \frac{\partial \pi_j^r(\boldsymbol{\alpha}^r, \boldsymbol{\beta}^r)}{\partial \beta_j^r} \geq 0 \right\} - \Delta\beta_j^r \mathbf{1}\left\{ \frac{\partial \pi_j^r(\boldsymbol{\alpha}^r, \boldsymbol{\beta}^r)}{\partial \beta_j^r} < 0 \right\},
\end{aligned}$$

   where $\Delta\alpha_j^r$ and $\Delta\beta_j^r$ are the r-th incremental changes in the pricing coefficients.

5. Finally, we reduce the size of incremental changes $(\Delta\alpha_j^r, \Delta\beta_j^r)$ if sign switching in the first-order conditions is observed in two consecutive iterations.

$$\begin{aligned}
\Delta\alpha_j^{r+1} &= \begin{cases} \frac{\Delta\alpha_j^r}{2} & \text{if} \quad \frac{\partial \pi_j^r(\boldsymbol{\alpha}^r, \boldsymbol{\beta}^r)}{\partial \alpha_j^r} \frac{\partial \pi_j^{r-1}(\boldsymbol{\alpha}^{r-1}, \boldsymbol{\beta}^{r-1})}{\partial \alpha_j^{r-1}} < 0 \\ \Delta\alpha_j^r & \text{otherwise,} \end{cases} \\
\Delta\beta_j^{r+1} &= \begin{cases} \frac{\Delta\beta_j^r}{2} & \text{if} \quad \frac{\partial \pi_j^r(\boldsymbol{\alpha}^r, \boldsymbol{\beta}^r)}{\partial \beta_j^r} \frac{\partial \pi_j^{r-1}(\boldsymbol{\alpha}^{r-1}, \boldsymbol{\beta}^{r-1})}{\partial \beta_j^{r-1}} < 0 \\ \Delta\beta_j^r & \text{otherwise,} \end{cases}
\end{aligned}$$

6. Iterate this process until $(\Delta\alpha_j^r, \Delta\beta_j^r)$ converge to 0 for all j.

Finally, given that the current solution of the pricing coefficient might be a local maximizer, we check whether it is globally optimal. That is, for each firm j, fixing other firms' pricing strategy at the current solution, we search for the value of (α_j, β_j) that maximizes profit π_j. If the global maximizer differs from the current pricing coefficient, we update firm j's pricing coefficient to that global maximizer and repeat steps 1–6. Otherwise, we have found the Nash equilibrium. For the counterfactual scenarios where firms have equal access to either the aggregated risk score from the centralized risk bureau or the true risk type, we solve the equilibrium using a similar iterative procedure, but can skip Step 2.

References

ACQUISTI, A., C. TAYLOR, AND L. WAGMAN (2016): "The economics of privacy," *Journal of Economic Literature*, 54, 442–492.

AKERLOF, G. A. (1970): "The market for 'lemons': Quality uncertainty and the market mechanism," *Quarterly Journal of Economics*, 84, 488–500.

ALCOBENDAS, M., S. KOBAYASHI, K. SHI, AND M. SHUM (2023): "The impact of privacy protection on online advertising Markets," in *Proceedings of the 24th ACM Conference on Economics and Computation*, 62–62.

ALLEN, J., R. CLARK, AND J.-F. HOUDE (2019): "Search frictions and market power in negotiated-price markets," *Journal of Political Economy*, 127, 1550–1598.

AZEVEDO, E. M. AND D. GOTTLIEB (2017): "Perfect competition in markets with adverse selection," *Econometrica*, 85, 67–105.

BARWICK, P. J., H.-S. KWON, AND S. LI (2023): "Attribute-based subsidies and market power: an application to electric vehicles," *Working paper*.

BECKER, J., K. HENDRICKS, J.-F. HOUDE, AND D. RAISINGH (2024): "Asymmetric information and the supply-chain of mortgages: The case of Ginnie Mae loans," Tech. rep., University of Wisconsin-Madison.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile prices in market equilibrium," *Econometrica*, 63, 841–890.

BERRY, S. T. (1994): "Estimating discrete-choice models of product differentiation," *RAND Journal of Economics*, 242–262.

BIGLAISER, G., F. LI, C. MURRY, AND Y. ZHOU (2020): "Intermediaries and product quality in used car markets," *RAND Journal of Economics*, 51, 905–933.

BLICKLE, K., Z. HE, J. HUANG, AND C. PARLATORE (2024): "Information-Based Pricing in Specialized Lending," Tech. rep., National Bureau of Economic Research.

BOOMHOWER, J., M. FOWLIE, J. GELLMAN, AND A. PLANTINGA (2024): "How are insurance markets adapting to climate change? risk selection and regulation in the market for homeowners insurance," Tech. rep., National Bureau of Economic Research.

CABRAL, M., M. GERUSO, AND N. MAHONEY (2018): "Do larger health insurance subsidies benefit patients or producers? Evidence from Medicare Advantage," *American Economic Review*, 108, 2048–2087.

CAMPBELL, J., A. GOLDFARB, AND C. TUCKER (2015): "Privacy regulation and market structure," *Journal of Economics & Management Strategy*, 24, 47–73.

CHEN, N. AND H.-T. TSAI (2023): "Price competition dnder information (dis)advantage," *Available at SSRN 4420175*.

CHIAPPORI, P.-A., B. JULLIEN, B. SALANIÉ, AND F. SALANIE (2006): "Asymmetric information in insurance: General testable implications," *RAND Journal of Economics*, 37, 783–798.

CHIAPPORI, P.-A. AND B. SALANIE (2000): "Testing for asymmetric information in insurance markets," *Journal of Political Economy*, 108, 56–78.

CICALA, S. (2015): "When does regulation distort costs? Lessons from fuel procurement in us electricity generation," *American Economic Review*, 105, 411–444.

COHEN, A. (2005): "Asymmetric information and learning: Evidence from the automobile insurance market," *Review of Economics and Statistics*, 87, 197–207.

COHEN, A. AND L. EINAV (2007): "Estimating risk preferences from deductible choice," *American Economic Review*, 97, 745–788.

COSCONATI, M. (2023): "Market-wide moral hazard and price walking: Evidence from automobile insurance market," *Working paper*.

CRAWFORD, G. S., N. PAVANINI, AND F. SCHIVARDI (2018): "Asymmetric information and imperfect competition in lending markets," *American Economic Review*, 108, 1659–1701.

CRAWFORD, G. S., O. SHCHERBAKOV, AND M. SHUM (2019): "Quality overprovision in cable television markets," *American Economic Review*, 109, 956–995.

CUESTA, J. I. AND A. SEPÚLVEDA (2021): "Price regulation in credit markets: A trade-off between consumer protection and credit access," *Available at SSRN 3282910*.

CURTO, V., L. EINAV, J. LEVIN, AND J. BHATTACHARYA (2021): "Can health insurance competition work? evidence from medicare advantage," *Journal of Political Economy*, 129, 570–606.

DECAROLIS, F. AND A. GUGLIELMO (2017): "Insurers' response to selection risk: Evidence from Medicare enrollment reforms," *Journal of Health Economics*, 56, 383–396.

DECAROLIS, F., M. POLYAKOVA, AND S. P. RYAN (2020): "Subsidy design in privately provided social insurance: Lessons from Medicare Part D," *Journal of Political Economy*, 128, 1712–1752.

EINAV, L. AND A. FINKELSTEIN (2011): "Selection in insurance markets: Theory and empirics in pictures," *Journal of Economic Perspectives*, 25, 115–138.

EINAV, L., A. FINKELSTEIN, AND M. R. CULLEN (2010): "Estimating welfare in insurance markets using variation in prices," *Quarterly Journal of Economics*, 125, 877–921.

EINAV, L., A. FINKELSTEIN, AND N. MAHONEY (2021): "The IO of selection markets," in *Handbook of Industrial Organization*, Elsevier, vol. 5, 389–426.

EINAV, L., M. JENKINS, AND J. LEVIN (2012): "Contract pricing in consumer credit markets," *Econometrica*, 80, 1387–1432.

——— (2013): "The impact of credit scoring on consumer lending," *RAND Journal of Economics*, 44, 249–274.

ERICSON, K. M. M. (2014): "Consumer inertia and firm pricing in the Medicare Part D prescription drug insurance exchange," *American Economic Journal: Economic Policy*, 6, 38–64.

FAN, Y. AND C. YANG (2020): "Competition, product proliferation, and welfare: A study of the US smartphone market," *American Economic Journal: Microeconomics*, 12, 99–134.

FARRELL, J. AND P. KLEMPERER (2007): "Coordination and lock-in: Competition with switching costs and network effects," *Handbook of Industrial Organization*, 3, 1967–2072.

FRIEDRICH, B. U., M. B. HACKMANN, A. KAPOR, S. MORONI, AND A. B. NANDRUP (2023): "Asymmetric information in matching markets: Evidence from medical school programs in Denmark," *Working paper*.

FUDENBERG, D. AND J. TIROLE (1991): *Game Theory*, MIT Press.

GOLDBERG, P. K. (1996): "Dealer price discrimination in new car purchases: Evidence from the consumer expenditure survey," *Journal of Political Economy*, 104, 622–654.

GOLDBURD, M., A. KHARE, D. TEVET, AND D. GULLER (2016): "Generalized linear models for insurance rating," *Casualty Actuarial Society, CAS Monographs Series*, 5.

GOLDFARB, A. AND V. F. QUE (2023): "The economics of digital privacy," *Annual Review of Economics*, 15.

GOLDFARB, A. AND C. TUCKER (2012): "Privacy and innovation," *Innovation policy and the economy*, 12, 65–90.

GRODZICKI, D. (2023): "The evolution of competition in the credit card market," *Available at SSRN 4493211*.

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal nonparametric estimation of first-price auctions," *Econometrica*, 68, 525–574.

HANDEL, B. AND K. HO (2021): "The industrial organization of health care markets," in *Handbook of industrial organization*, Elsevier, vol. 5, 521–614.

HANDEL, B. R. (2013): "Adverse selection and inertia in health insurance markets: When nudging hurts," *American Economic Review*, 103, 2643–2682.

HENDRICKS, K. AND R. H. PORTER (1988): "An empirical study of an auction with asymmetric information," *American Economic Review*, 865–883.

HENDRICKS, K., R. H. PORTER, AND B. BOUDREAU (1987): "Information, returns, and bidding behavior in OCS auctions: 1954–1969," *Journal of Industrial Economics*, 517–542.

HENDRICKS, K., R. H. PORTER, AND C. A. WILSON (1994): "Auctions for oil and gas leases with an informed bidder and a random reservation price," *Econometrica: Journal of the Econometric Society*, 1415–1444.

HONKA, E. (2014): "Quantifying search and switching costs in the US auto insurance industry," *RAND Journal of Economics*, 45, 847–884.

HONKA, E., A. HORTAÇSU, AND M. A. VITORINO (2017): "Advertising, consumer awareness, and choice: Evidence from the US banking industry," *The RAND Journal of Economics*, 48, 611–646.

HORTAÇSU, A., O. R. NATAN, H. PARSLEY, T. SCHWIEG, AND K. R. WILLIAMS (2024): "Organizational structure and pricing: Evidence from a large us airline," *The Quarterly Journal of Economics*, 139, 1149–1199.

JAFFE, S. AND M. SHEPARD (2020): "Price-linked subsidies and imperfect competition in health insurance," *American Economic Journal: Economic Policy*, 12, 279–311.

JEON, D.-S., D. MENICUCCI, AND N. NASR (2023): "Compatibility choices, switching costs, and data portability," *American Economic Journal: Microeconomics*, 15, 30–73.

JEZIORSKI, P., E. KRASNOKUTSKAYA, AND O. CECCARINI (2017): "Adverse selection and moral hazard in the dynamic model of auto insurance," *UC Berkeley, Haas School of Business working paper*.

JIA, J., G. Z. JIN, AND L. WAGMAN (2018): "The short-run effects of GDPR on technology venture investment," Tech. rep., National Bureau of Economic Research.

JIN, G. Z. AND L. WAGMAN (2021): "Big data at the crossroads of antitrust and consumer protection," *Information Economics and Policy*, 54, 100865.

JOHNSON, G. (2022): "Economic research on privacy regulation: Lessons from the GDPR and beyond," *NBER Working paper*.

JOHNSON, G. A., S. K. SHRIVER, AND S. G. GOLDBERG (2023): "Privacy and market concentration: Intended and unintended consequences of the GDPR," *Management Science*, 69, 5695–5721.

KRÄMER, J. (2021): "Personal data portability in the platform economy: economic implications and policy recommendations," *Journal of Competition Law & Economics*, 17, 263–308.

LAM, W. M. W. AND X. LIU (2020): "Does data portability facilitate entry?" *International Journal of Industrial Organization*, 69, 102564.

MAHONEY, N. AND E. G. WEYL (2017): "Imperfect competition in selection markets," *Review of Economics and Statistics*, 99, 637–651.

MATCHAM, W. (2023): "Risk-based quantity limits in credit card markets," Tech. rep.

NELSON, S. (2018): "Private information and price regulation in the US credit card market," *Working paper*.

PANETTA, F., F. SCHIVARDI, AND M. SHUM (2009): "Do mergers improve information? Evidence from the loan market," *Journal of Money, Credit and Banking*, 41, 673–709.

PORTER, R. H. (1995): "The role of information in US offshore oil and gas lease auction," *Econometrica: Journal of the Econometric Society*, 1–27.

ROTHSCHILD, M. AND J. STIGLITZ (1976): "Equilibrium in competitive insurance markets: An essay on the economics of imperfect information," *Quarterly Journal of Economics*, 90, 629–649.

ROY, A. D. (1951): "Some thoughts on the distribution of earnings," *Oxford Economic Papers*, 3, 135–146.

SAGL, S. (2023): "Dispersion, discrimination, and the price of your pickup," *Working paper*.

SALANIÉ, B. (2017): "Equilibrium in insurance markets: An empiricist's view," *Geneva Risk and Insurance Review*, 42, 1–14.

SALZ, T. (2022): "Intermediation and competition in search markets: An empirical case study," *Journal of Political Economy*, 130, 310–345.

SERNA, N. (2023): "Non-price competition, risk selection, and heterogeneous costs in hospital networks," *Working paper*.

SWEETING, A. (2013): "Dynamic product positioning in differentiated product markets: The effect of fees for musical performance rights on the commercial radio industry," *Econometrica*, 81, 1763–1803.

TEBALDI, P. (2024): "Estimating equilibrium in health insurance exchanges: Price competition and subsidy design under the aca," *Review of Economic Studies*, rdae020.

TUCKER, C. (2019): "Digital data, platforms and the usual [antitrust] suspects: Network effects, switching costs, essential facility," *Review of industrial Organization*, 54, 683–694.

WU, F. AND Y. XIN (2024): "Estimating nonseparable selection models: A functional contraction approach," Tech. rep., California Institute of Technology.