Competing under Information Heterogeneity: Evidence from Auto Insurance*

Marco Cosconati
IVASS, Bank of Italy

Yi Xin
Caltech

Fan Wu
Caltech

Yizhou Jin
Toronto

July 14, 2025

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 15.7% boost in consumer surplus, almost reaching the efficiency

*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 Francesca Carapella, Satyajit Chatterjee, José Ignacio Cuesta, Liran Einav, Hanming Fang, Nathan Foley-Fisher, Jeremy Fox, Jean-François Houde, Yingyao Hu, Ginger Jin, Elena Krasnokutskaya, Benjamin Lester, Fei Li, Igor Livshits, Yao Luo, Bob Sherman, Matthew Shum, Paulo Somaini, Andrew Sweeting, Xun Tang, Pietro Tebaldi, Ana-Maria Tenekedjieva, Tiffany Tsai, and Mo Xiao. We thank seminar and conference participants at Caltech, Cornell, CUHK, FRB, JHU, HKU, HKUST, Indiana (Bloomington), Maryland, NTU, Penn State, Philly Fed, PKU (GSM), Rochester (Simon), Stony Brook, UC Davis, UNC-Chapel Hill, U Penn, UVA, Yale, SITE (2024), SHUFE IO Conference (2024), SoCal Structural Econometrics Conference (2024), AEA (2025), and SoCCAM (2025). 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.

benchmark where firms have full knowledge of consumers' true risk. Aggregating information through the bureau favors low-risk consumers and reduces average costs by 12 euros per contract through more efficient insurer-insuree matching.

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 prepayment 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 telematics 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 information asymmetries between firms 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

[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 have 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). In mortgage market leverage regulation, policymakers allow large lenders to use internal rating-based (IRB) models, which are costly to develop and maintain, to calculate risk weights, while smaller lenders typically rely on standardized approaches (Benetton, 2021).

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

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.

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 15.7%, nearly reaching the efficiency benchmark where firms have perfect knowledge of consumers' true risk, and reduces average costs by 12 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 *ex-post* realized consumer risk.[3] Our analysis demonstrates that certain firms' premi-

[3]Similarly, Porter (1995) estimates the ex-post value of oil leases (which is ex-ante unknown to bidders)

ums are more responsive to, or more accurately reflect, realized consumer risk, indicating that these firms are potentially more informed and better at assessing 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 strategies 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 is 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. We show that the way average prices vary with risk identifies firms' pricing coefficients, and the remaining price dispersion 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

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 find that dealers have better information about car quality.

[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

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 those with lower information precision tend to attract more high-risk customers.[6] We also find that insurers enjoy large and different net benefits from contracting with new customers (potentially due to inertia, dynamic pricing strategies, and cross-selling other products), 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 completely eliminating information asymmetry. At the other extreme, motivated by privacy regulations, we simulate a scenario in which firms are restricted to using only basic consumer information, and their signal variance is set to the highest level observed in the market.

We observe a substantial reduction in average premiums, ranging from 21.6% when a centralized risk bureau is introduced to 25.7% 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, more nuanced channel is that 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 15.7%, nearly matching the 16.9% increase 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. The privacy benchmark results in opposite distributional effects, with high-risk consumers enjoying a 6.9% increase in surplus, as it

(2000) and Crawford et al. (2018).

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

becomes more difficult for firms to distinguish them from low-risk consumers.

Eliminating information heterogeneity also significantly impacts consumer sorting patterns, which critically affects market efficiency. We find that when insurers have equal access to consumers' true risk, 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 3.7%. Under the centralized risk bureau scenario, a similar mechanism applies, leading to a reduction in the average cost of 12 euros. Given the size of the Italian auto insurance market, this implies a substantial aggregate impact. Our results highlight 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, 2025; 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.[7] Importantly, our paper also incorporates multidimensional cost heterogeneity on the supply side (Salanié, 2017; Serna, 2023; Nelson, 2025) 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

[7]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).

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). A recent paper by D'Haultfœuille et al. (2019) addresses a related challenge in demand estimation under unobserved price discrimination using supply-side restrictions.[8] Moreover, we identify and estimate the 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).

On the policy side, our paper contributes to ongoing 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. Our work also relates to broader questions in financial market regulation, including the design and implications of credit scoring systems (Einav et al., 2013; Chatterjee et al., 2023; Blattner et al., 2022), and how policy interventions influence lenders' access to and use of information, thereby shaping market outcomes (Nelson, 2025; Blattner and Nelson, 2021; Liberman et al., 2018; Hertzberg et al., 2011). Our paper contributes to this literature by providing a tractable empirical framework for analyzing such policy questions, starting from a baseline in which lenders use heterogeneous risk evaluation models.

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 in Section 3, followed by the identification and estimation strategies in Section 4. Section 5 reports the empirical estimation results and model fit. The details of our counterfactual experiments are provided in Section 6. Section 7 concludes.

[8]The analysis in D'Haultfœuille et al. (2019) focuses on general consumer goods. One of the key assumptions in their framework is that the cost of selling to different consumers is identical, which is reasonable in many settings. However, this assumption is unlikely to hold in selection markets such as insurance. As D'Haultfœuille et al. (2019) notes, "This is the case for insurance providers that offer different prices to consumers based on their observable characteristics (e.g., age, gender, driving experience), because those characteristics imply different risk classes and different costs for insurers."

2 Institutional Background

2.1 Italian Auto Insurance Market

Our paper focuses on the market for mandatory 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 around 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.

Contract Design Several 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. In addition to the baseline contract, consumers can select optional clauses. For example, the "Exclusive Driving" clause restricts vehicle use to the driver named in the contract. The "Protected Bonus" clause limits the premium increase following an at-fault accident.

Premium Contract premiums are determined by insurers' actuarial algorithms based on a range of risk factors, including age, vehicle features, place of residence, and driving history.

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

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

Premiums also vary with optional contractual clauses and the number of payment installments selected by the consumer. For instance, the "Exclusive Driving" clause typically lowers the premium, while the "Protected Bonus" clause increases it. Choosing to pay in multiple installments rather than in a single upfront payment also incurs a surcharge.

In Italy (as in many other countries), a uniform experience rating system relates the history of accidents to a class of risk, known as the 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.

Indeed, insurance companies in Italy use heterogeneous information in their pricing algorithms. A survey of five major insurers reveals that the variables used for pricing differ substantially across firms. For example, some firms incorporate factors such as the presence of safety devices, marital status, annual driving distance, finer-level zip codes, occupation, and other socioeconomic indicators, while others do not.[11] These firm-specific variables are not observed in the data, although our dataset includes risk factors typically used in motor insurance pricing, such as consumer characteristics (e.g., age, bonus-malus (BM) class, number of accidents in the past five years), vehicle features (e.g., engine power, vehicle age), province of residence, and city-level geographic indicators. Overall, the observable characteristics in our data explain approximately 50% of the variation in premiums paid by consumers, with firm-level R^2 values ranging from 0.39 to 0.59. We report the firm-specific regressions of premiums on the observable covariates in Table S2, presented in Section S1 of the Supplementary Materials.

Even if firms use the same set of pricing variables, their actuarial algorithms may still differ significantly. Coefficients applied to these variables are often estimated using proprietary historical data, which can vary across firms in terms of sample size, coverage, and quality. Differences in internal modeling practices, legacy systems, and implementation approaches can further contribute to variation in pricing algorithms. Table S2 shows that 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 1.68 euros, a higher BM class increases premiums by 12 to 32 euros, and having one accident in

[11] A subset of pricing variables used by insurance companies is listed in Table S1 in the Supplementary Materials, Section S1. To protect firm confidentiality and ensure anonymity, we report only a selected subset.

the previous five years raises premiums by 74 to 181 euros across firms. While these results reflect only reduced-form relationships between premiums and observable characteristics and do not account for competitive effects, they nonetheless highlight how firms may differentially translate risk-related information into pricing, providing suggestive evidence of heterogeneity in their pricing algorithms.

In addition to the posted prices determined by actuarial models, discounts offered by sales agents play an important role in determining the final premiums consumers pay. While some smaller insurers sell products exclusively online, the major firms in this market rely heavily on sales agents. In some firms, agents may also have more experience or better knowledge of consumers' driving habits, potentially influencing pricing decisions.

Another potential source of discounts in reality arises from the fact that insurance companies could offer multiple products beyond liability auto insurance, such as comprehensive coverage, property insurance, and life insurance. This might be especially relevant among larger insurers. Firms that offer a broader range of products may have stronger incentives to provide premium discounts, as doing so can encourage consumers to purchase additional policies. Our model partially captures the benefits of cross-selling, which we discuss in more detail in Section 3. Selling multiple products to the same consumer may also provide firms with informational advantages. For example, a consumer who purchases property insurance may inadvertently reveal additional risk-relevant information, such as their neighborhood, allowing the firm to refine its assessment of that consumer's risk.

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 124,428 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.

[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).

Table 1: Summary statistics

Variables	Mean	Std. Dev.	Min	Max	N
Premium (€)	477.68	208.79	133.68	1335.05	124,428
Claim size (€)	260.89	10217.58	0	2521014	124,428
No. of claims (within contract year)	0.08	0.29	0	4	124,428
No. of accidents in last 5 years	0.81	1.22	0	3	124,428
BM class	2.06	2.51	1	15	124,428
Age	48.24	14.11	18	99	124,428
Man	0.56	0.50	0	1	124,428
Median city	0.10	0.30	0	1	124,428
Big city	0.62	0.49	0	1	124,428
Car age	8.30	5.27	0	19	124,428
Horsepower	66.88	26.84	0	493	124,428
Petrol vehicle	0.52	0.50	0	1	124,428
One installment	0.67	0.47	0	1	124,428

Price Variations and Consumer Sorting We plot the average premium and claim payouts for 11 firms separately in Figure 1, where the color of the circle represents the market share of each firm. We observe a substantial price variation across firms in this market, even 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 sorted into different insurers.

Heterogeneity in Information Precision Do insurers differ in the precision of their risk assessments? To explore this question, we examine the extent to which each firm's premiums are correlated with realized consumer risk. A consumer's true risk type is *ex-ante* unknown to insurers. Instead, insurers rely on observable characteristics to *predict* the likelihood of an accident. At the time of contract signing, insurers cannot observe whether an accident will occur during the coverage period because these outcomes are in the future and are yet to be realized. In contrast, we observe the *ex-post* realized accident records for each consumer over multiple periods. This allows us to construct an estimate of the consumer's underlying risk type, which is the object insurers attempt to forecast when setting premiums. Intuitively, if a firm's premiums are more responsive to (or more accurately reflect) the realized consumer risk, this suggests that the firm is better at assessing risk, and vice

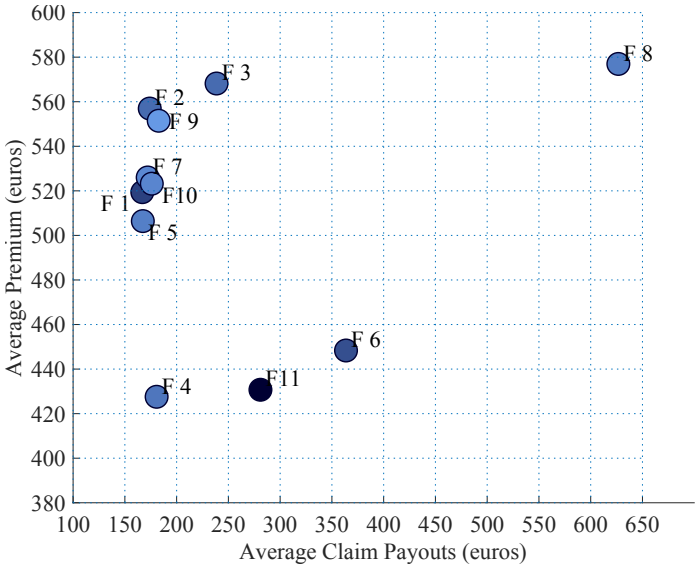


Figure 1: Heterogeneity across firms: Average premium and claim payouts (in euros). Darker colors indicate larger market share, and firm IDs are labeled next to each circle.

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 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.[13] 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,

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

We then estimate a log-normal regression of claim size conditional on the consumer being

[13]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).

involved in an accident:[14]

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

With the estimated regression coefficients and the individual fixed effects from Equations (2.1) and (2.2), we predict the expected number of accidents and the expected indemnity 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.[15] 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]

We estimate firm-specific linear regressions of the initial premiums paid by consumers at the time of signing the contract on their estimated risk. Figure 2 shows the coefficients on the risk measure for each firm, with error bars representing the 95% confidence intervals.[17] This figure shows that the extent to which firms adjust their premiums in response to increases in realized consumer risk varies significantly across firms.[18] 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. Following a similar intuition, we also use premiums to directly predict claim counts, controlling for individual and vehicle characteristics as well as contract features, using a Poisson regression. The results are reported in Table S4 in the Supplementary Material, Section S1. Again, we find that the extent to which premiums predict realized claim counts within the contract period varies

[14]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.

[15]Our approach to estimating the expected cost of insuring a consumer is related to the two methods discussed in Abaluck and Gruber (2016). The first is a "realized cost" model, which constructs out-of-pocket costs using claims incurred during the year. The second is a "rational expectations" model, which predicts expected drug spending based on claims from the prior year. Our approach can be viewed as a hybrid of these two methods. We estimate the expected cost of insuring a consumer using ex-post realized claim records, allowing the estimation to depend not only on rich consumer characteristics, including accident histories, but also on individual fixed effects. The key advantage of our setting is that we observe each consumer over multiple years, which enables us to control for individual fixed effects and better capture heterogeneity that may not be fully explained by observables alone.

[16]Specifically, we control for contract features including coverage, repair restrictions, exclusive driving, expert driving, free driving, protected bonus, and the presence of a monitoring device. Our estimates of risk can be interpreted as incorporating the effects of moral hazard in a reduced-form way. That is, consumers may change their risky behavior after choosing different contracts.

[17]We compute the standard error of the linear regression coefficient using 200 bootstrap replications, accounting for the sampling variability introduced by the generated regressor from the first-step estimations of Equations (2.1) and (2.2).

[18]The positive coefficients on risk level and their considerable variation across firms remain robust even after controlling for a comprehensive set of individual and vehicle characteristics, as well as contract features. The regression results are reported in Table S3 in the Supplementary Material, Section S1.

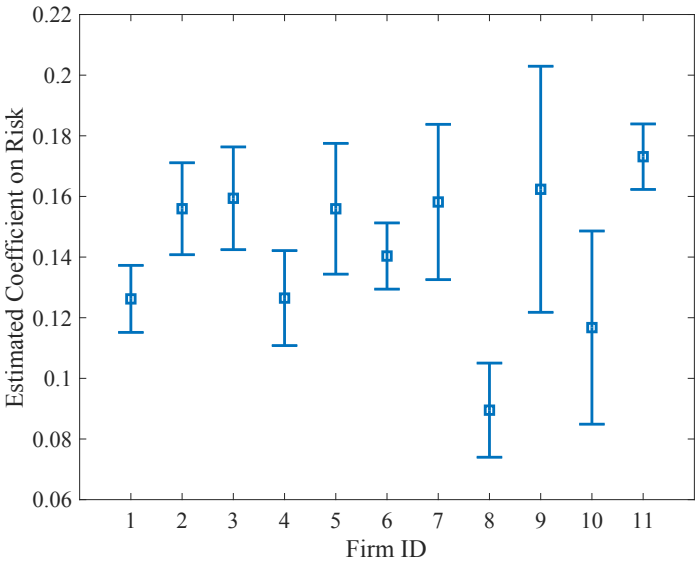


Figure 2: Regressing premiums on consumer risks: Estimated coefficients on risk measure for each firm, with error bars representing the 95% confidence intervals. The standard errors are computed using 200 bootstrap replications, accounting for the sampling variability introduced by the generated regressor from the first-step estimations.

significantly across firms, potentially reflecting heterogeneity in their risk-rating precision.

In general, the patterns shown in Figure 2 are complex equilibrium outcomes and the differences in the correlation between premiums and risk could be due to other factors. For example, if a firm is more efficient at processing claims, it may design its pricing algorithm to attract high-risk consumers, leading to a weaker correlation between premiums and risk. 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.

3 Model

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.[19] 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

[19]Our model abstracts away from consumers' choices of additional contract features; instead, we treat them as observable consumer characteristics that can directly affect premiums and risks.

all firms. Let D denote the contract chosen by a consumer.

For notational simplicity, we omit the individual index i throughout this section. Moreover, 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.

Signal Structure When a consumer of type θ arrives, firm j draws a signal $\hat{\theta}_j$ from $\mathcal{N}(\theta, \sigma_j^2)$, with its density denoted by $\phi(\hat{\theta}_j; \theta, \sigma_j)$. The parameter 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.[20] When $\sigma_j = 0$, firm j perfectly observes θ; as σ_j increases, the signal becomes less precise. 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 $\mathcal{N}(\theta, \sigma_j^2)$ captures firm j's information technology for risk evaluation. In reality, the set of variables and algorithms used in the risk-rating process may differ across firms (as documented in Section 2.2); sales agents in some firms may have more experience and better knowledge of consumers' driving habits; and some firms may obtain additional information through cross-selling other products. In this paper, we treat these factors as part of a broader "black box" that contributes to firm-level differences in information precision. We do not model these mechanisms explicitly, primarily due to data limitations, but they represent promising directions for future research.

Risk Rating Firm j infers the risk type of the consumer upon observing the signal $\hat{\theta}_j$. Let $f(\theta|\hat{\theta}_j, D = j)$ 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 $E(\theta|\hat{\theta}_j, D = j)$ represents firm j's risk rating for a consumer conditional on signal $\hat{\theta}_j$ and her choosing j. Specifically,

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

[20] More generally, we can relax the normality assumption and assume that the signal is drawn from some cumulative distribution function $F_j(\hat{\theta}_j|\theta)$. The signal $\hat{\theta}_j$ can be viewed as a risk score and the location and scale of it do not have economic meanings. In other words, any affine transformation of the signal distribution will play the same role in our model. To compare the information precision across firms, some concentration measures of the signal distribution are necessary.

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

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

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 $E(\theta|\hat{\theta}_j, D = j)$ 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., $p_j(\hat{\theta}_j) = \tau_j(E(\theta|\hat{\theta}_j, D = j))$. 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

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

where $\gamma(\theta)$ represents the price sensitivity parameter and $\xi_j(\theta)$ represents the unobserved heterogeneity for product j.[21] Our demand model allows consumer preference parameters

[21] Non-financial attributes of the insurance product may vary across geography and time due to differences in branch presence and advertising efforts. To capture this possibility, we allow ξ_j to vary by location and time in our demand estimation. More detailed discussion is provided in Section 5.1.

to vary with risk type θ, capturing the idea that higher-risk consumers may be more price sensitive and place greater value on higher-quality services. It can also be readily extended to incorporate observable product attributes and allow preference parameters to vary with individual characteristics, geography, and time.

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 and the consumer's risk type θ is

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

In Equation (3.4), a consumer's risk type θ influences choice behavior through two channels. First, it directly affects the consumer's preference parameters. Second, it indirectly affects choice probabilities through the signal distributions, and ultimately, the price distributions the consumer faces. This latter channel highlights how heterogeneity in firms' pricing strategies impacts consumer sorting patterns, which is a key mechanism emphasized in this paper.

Profit Maximization We assume that firms simultaneously choose their pricing coefficients (α_j, β_j) to maximize expected profits, taking into account competitors' pricing strategies and consumer demand responses in a static game. At the time of setting their strategies, firms do not observe the specific realizations of consumer types or signals (including both their own and those of competitors). However, the signal distributions, cost parameters, unobserved product attributes, and the risk distribution are all common knowledge. Once a pricing strategy is chosen, it is applied to consumers as they arrive. This modeling assumption reflects industry practice, where insurance companies develop and maintain pricing algorithms that are consistently applied, rather than adjusting prices through real-time bidding for each consumer.

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

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

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, dynamic pricing strategies, and cross-selling other products.[22] We treat c_j as a primitive of our model.

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 $Pr(D = j|\hat{\boldsymbol{\theta}}, \theta)$ 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.

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). The equilibrium of our pricing game may not be unique. We describe how the game is re-solved under counterfactual scenarios in Section 6.

Remarks We conclude the section with a few remarks on our modeling choices. First, our model allows for rich supply-side heterogeneity. Insurers are differentiated by their cost structures (c_j, k_j), information technology (σ_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. Our model does not capture firms' long-term investment decisions, such as improving information technology or reducing costs, that may have led to the observed equilibrium. Incorporating a first-stage investment decision is a promising

[22]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.

direction for future research.

Second, 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 also restrict attention to liability auto insurance contracts. In practice, insurance companies may also offer other products (such as comprehensive auto insurance or property insurance), and may price them as part of a bundle, which is not observed in our data. We partially capture the effects of consumer retention and bundling through the net benefit parameter c_j. Once a consumer purchases an auto insurance plan from firm j, they may remain with the firm for several years and purchase additional products, giving the firm an incentive to offer lower prices on auto insurance. If consumers' switching behavior or preference for single-homing varies with risk type, the net benefit may also depend on risk type, creating identification challenges for the two cost components (c_j, k_j) in our model. Addressing this would require a detailed analysis of consumer switching behavior and multi-product purchases. We therefore impose the simplifying assumption that the net benefit does not vary with risk type.

Lastly, we abstract away from consumer search behavior and the possibility of limited consideration sets in our demand model to maintain tractability. As a robustness check, we re-estimate demand under the assumption that consumers only consider contracts offered by the top companies. A detailed discussion is provided in Section S2 of the Supplementary Materials.

4 Identification and Estimation

There are three sets of parameters we need to estimate for the model. We first estimate the distribution of consumers' risk types (i.e., the expected claim payouts to consumers) using a panel of claim records. Second, on the demand side, we estimate consumers' price sensitivity and the unobserved product attributes for each firm using a novel fixed-point algorithm. Finally, 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.

In this section, we discuss the key intuition and data variation used to identify and estimate the three sets of parameters. A detailed step-by-step estimation procedure is provided in Appendix B. Again, all identification and estimation strategies can be derived conditional on observable characteristics; we omit them here for notational simplicity.

4.1 Risk Type Distribution

For a consumer i, we assume that the true risk type is given by $\theta_i = \mu\lambda_i$, where μ denotes the expected claim size conditional on having an accident, and λ_i represents the individual-specific Poisson rate of accident occurrence within a contract year. This specification is motivated by industry practice, where claim frequency tends to be more persistent across individuals, while the monetary value of a claim, once an accident occurs, is more unpredictable and less systematically tied to individual traits. We estimate the expected claim size using the log-normal regression specified in Equation (2.2), controlling for a rich set of observable characteristics. The remaining key challenge in identifying the distribution of consumer risk types lies in recovering the distribution of the Poisson arrival rate of accidents.

We observe realized claim records over multiple periods for each consumer. Let y_{it} denote the number of accidents for consumer i in period t. We assume that the $y_{it} \sim Poisson(\lambda_i)$ and that accident counts are independent across periods. The joint distribution of accident records over T periods, conditional on consumer i choosing firm j and paying price p_i, is given by:

$$f(y_{i1}, y_{i2}, \ldots, y_{iT}|p_i, D_i = j) = \int \prod_{t=1}^{T} \left(\frac{\lambda_i^{y_{it}} \exp(-\lambda_i)}{y_{it}!} \right) f(\lambda_i|p_i, D_i = j) d\lambda_i. \quad (4.1)$$

Equation (4.1) can be interpreted as a measurement error model, where the Poisson rate λ_i is a latent variable and the observed accident counts over multiple periods serve as repeated noisy measurements. The joint distribution of accident counts identifies the distribution of the Poisson rate, conditional on the premium and the consumer's choice of firm.[23]

In the estimation, we discretize the Poisson rate and estimate the probability of each type by matching the model-implied joint distribution of accident counts in Equation (4.1) to the data using maximum likelihood. Details of the first-step estimation procedure are provided in Appendix B. Once we obtain an estimate of $f(\lambda_i|p_i, D_i = j)$, we combine it with the estimated expected claim size to derive the joint distribution of true risk type

[23]This type of measurement error model has been widely used in the literature to address unobserved heterogeneity. For example, Krasnokutskaya (2011) uses Kotlarski's identity to recover the distribution of unobserved auction heterogeneity from the joint distribution of multiple bids. Hu and Shum (2012) show that, under certain conditional independence assumptions, the joint distribution of observed state variables across four periods can identify the distribution of unobserved heterogeneity in dynamic discrete choice models. Xin (2023) uses the distribution of loan outcomes to identify the distribution of borrowers' risk types in online lending markets. Hu (2008) and Hu and Schennach (2008) provide general theoretical results on the nonparametric identification of this class of measurement error models. See Hu (2017) as well for a thorough review of the theoretical literature and related empirical applications.

and premium, conditional on the consumer's contract choice. We denote by $\hat{g}(p|\theta, D = j)$ the estimated density of premiums conditional on risk type θ for consumers who select into firm j. This is a direct output of the first-step estimation and is taken as given in the subsequent analysis.

4.2 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 challenge in many empirical studies (e.g., Goldberg, 1996; Cicala, 2015; Crawford et al., 2018; Allen et al., 2019; D'Haultfœuille et al., 2019; Salz, 2022; Sagl, 2023). Existing approaches typically address it either by predicting alternative prices using regression models or by recovering them through supply-side restrictions.

Unfortunately, the regression 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. Crawford et al. (2018) face a similar problem and they estimate the following linear regression model to predict prices of non-chosen products:

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

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 informational asymmetry among competing firms. Alternatively, D'Haultfœuille et al. (2019) propose using supply-side restrictions, specifically firms' first-order conditions, to recover unobserved prices. This approach relies on assumptions about firm conduct (for example, Bertrand competition) and assumes that the cost of selling to different consumers is identical. The latter assumption is unlikely to hold in selection markets such as insurance.

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 de-

veloped 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. For clarity, we first present the estimation strategy for the demand model under the assumption that consumer preference parameters do not vary with risk type, and then discuss the additional restrictions required when this assumption is relaxed.

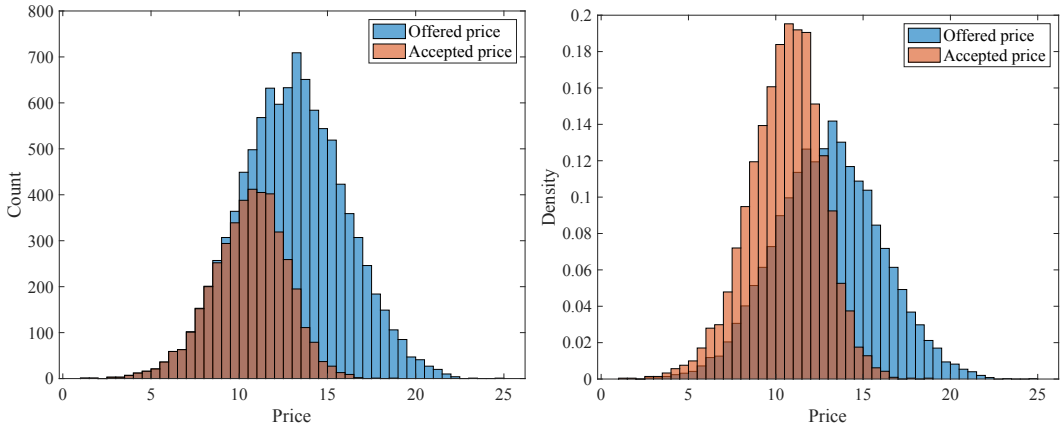
Offered vs. Accepted Prices We first highlight that there are two kinds of price distributions in our setting. The price distribution we observe in the data for each firm is the *accepted* price distribution after selection. It is different from the *offered* price distribution that consumers face when making their contract choices. 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 $g(p|\theta, D = j)$ denote the density of the offered and accepted prices in firm j conditional on the true risk type θ. Note that $g(p|\theta, D = j)$ is the posterior distribution of price *conditional on j being selected* and we have obtained an estimate of this distribution in the previous step, denoted by $\hat{g}(p|\theta, D = j)$. By applying Bayes' rule, we can easily observe that:

$$g(p|\theta, D = j) \propto g_j(p|\theta) Pr(D = j|p_j = p, \theta),$$

where $Pr(D = j|p_j = p, \theta)$ 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:

$$g_j(p|\theta) = \frac{g(p|\theta, D = j)/Pr(D = j|p_j = p, \theta)}{\int_{p'} g(p'|\theta, D = j)/Pr(D = j|p_j = p', \theta)dp'}, \tag{4.2}$$



(a) Count (b) Density

Figure 3: Comparing offered and accepted price distributions for a firm using simulated data. We draw offered prices from $\mathcal{N}(13, 3)$, 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.

where the denominator is for normalization.

In Equation (4.2), the likelihood $Pr(D = j|p_j = p, \theta)$ 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.

$$\begin{aligned} &Pr(D = j|p_j = p, \theta) \\ &= \int_{\boldsymbol{p}_{-j}} \frac{\exp(-\gamma p + \xi_j)}{\exp(-\gamma p + \xi_j) + \sum_{j' \neq j} \exp(-\gamma p_{j'} + \xi_{j'})} \left(\prod_{j' \neq j} g_{j'}(p_{j'}|\theta) \right) d\boldsymbol{p}_{-j}. \end{aligned} \tag{4.3}$$

Combining Equations (4.2) and (4.3) 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.[24]

Iterative Algorithm Given the fixed point problem defined in Equations (4.2) and (4.3), we propose an iterative procedure to solve for the *offered* price distributions for all firms

[24]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.

as functions of the demand parameters and the first-step estimates of the accepted price distributions. 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 $Pr^1(D = j|p_j = p, \theta) = \exp(-\gamma p)$ for all j.[25]

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 $Pr^r(D = j|p_j = p, \theta)$ 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 (4.2), plugging in the estimated accepted price distribution $\hat{g}(p|\theta, D = j)$:

$$g_j^r(p|\theta) = \frac{\hat{g}(p|\theta, D = j)/Pr^r(D = j|p_j = p, \theta)}{\int_{p'} \hat{g}(p'|\theta, D = j)/Pr^r(D = j|p_j = p', \theta)dp'}. \tag{4.4}$$

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:

$$s_j(\boldsymbol{\xi}^{r+1}, \gamma, \boldsymbol{g}^r) = \int_\theta \int_{\boldsymbol{p}} \frac{\exp(-\gamma p_j + \xi_j^{r+1})}{\sum_{j'} \exp(-\gamma p_{j'} + \xi_{j'}^{r+1})} \left(\prod_{j'} g_{j'}^r(p_{j'}|\theta) \right) f_0(\theta) d\boldsymbol{p} d\theta. \tag{4.5}$$

[25]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 (4.2). 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 (4.3):

$$Pr^{r+1}(D = j|p_j = p, \theta) \tag{4.6}$$
$$= \int_{\boldsymbol{p}_{-j}} \frac{\exp(-\gamma p + \xi_j^{r+1})}{\exp(-\gamma p + \xi_j^{r+1}) + \sum_{j' \neq j} \exp(-\gamma p_{j'} + \xi_{j'}^{r+1})} \left(\prod_{j' \neq j} g_{j'}^r(p_{j'}|\theta) \right) d\boldsymbol{p}_{-j}.$$

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

8. Iterate until $\boldsymbol{g}^r$ converges.

At convergence, for a fixed price sensitivity parameter γ and the estimated accepted price distributions $\hat{g}(p|\theta, D = j)$ for all j, 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, $Pr(D = j|p_j = p, \theta; \gamma)$, and (3) the vector of unobserved product attributes $\boldsymbol{\xi}(\gamma)$ that match the market shares.

Price Sensitivity Finally, we estimate the price sensitivity parameter γ using the risk sorting pattern observed in the data. When consumers' preference parameters do not vary with their risk types, the only channel through which risk type affects choice probabilities is by influencing the price distributions consumers face. 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 would disappear. In other words, γ is identified from variations in risk distributions observed across firms.

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 described above for each γ and construct the choice probabilities for selecting firm j conditional on each risk type. In the outer loop, we estimate γ by matching the model-implied choice probabilities for each type with the risk sorting patterns recovered in the first-step estimation. We describe the likelihood function used to estimate γ in Appendix B. Once we obtain an estimate $\hat{\gamma}$ for the price sensitivity parameter, the vector of unobserved product attributes $\boldsymbol{\xi}(\hat{\gamma})$ and the offered price distribution $g_j(p|\theta; \hat{\gamma})$ are recovered as by-products of the iterative procedure.

Risk-Dependent Preference Parameters Up to this point, we have focused on the simplified demand model, 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 (4.2) and (4.3) 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., $Pr(D = j|\theta)$. 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.

4.3 Supply

We now turn to the estimation of the supply-side parameters. The key insight is that there is a one-to-one mapping between the offered price and the signal, and we have already recovered the offered price distribution $g_j(p|\theta)$ for all j in the previous step. This idea 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).

The main challenge in identifying the supply-side parameters is disentangling the pric-

ing coefficients from the signal variance. Given the identified offered price distribution $g_j(p|\theta)$, we observe price dispersion at a fixed risk level. This dispersion may arise from both noisy signals and the pricing coefficients. However, by focusing on average prices conditional on each risk level, the noise introduced by signals averages out. The way average prices vary with risk then identifies the pricing coefficients. The remaining price dispersion is attributed to variation in signals, allowing us to identify the signal variance. Once these parameters are identified, we recover the firms' cost parameters using the first-order conditions derived from their profit maximization problem.

Pricing Coefficients Building on the identification intuition above, 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:

$$E(p|D=j) = \alpha_j + \beta_j E(E(\theta|\hat{\theta}_j, D=j)|D=j) = \alpha_j + \beta_j E(\theta|D=j), \tag{4.7}$$

$$cov(p, \theta|D=j) = \beta_j var(E(\theta|\hat{\theta}_j, D=j)|D=j) = \frac{var(p|D=j)}{\beta_j}. \tag{4.8}$$

Equations (4.7) and (4.8) uniquely determine (α_j, β_j) as the solution to a system of two linear equations, where all the first- and second-order moments of premiums and risks in firm j have been recovered.

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.

Signal Distribution Motivated by the identification argument in auction models, when price is a monotonically increasing function of the signal, the change-of-variables formula yields:

$$G_j(p_j(\hat{\theta}_j)|\theta) = \Phi\left(\frac{\hat{\theta}_j - \theta}{\sigma_j}\right), \tag{4.9}$$

where $G_j(p|\theta)$ denotes the cumulative distribution function of the offered price distribution of firm j for a given risk type θ, which has been identified in the previous step, and Φ denotes the cumulative distribution function of the standard normal distribution. Equation (4.9) implies that for a given θ, if we know the signal standard deviation σ_j, we can pin

down the corresponding price for a signal $\hat{\theta}_j$ using the inverse of the recovered offered price distribution G_j. Specifically,

$$p_j^o(\hat{\theta}_j; \theta, \sigma_j) = G_j^{-1}\left(\Phi\left(\frac{\hat{\theta}_j - \theta}{\sigma_j}\right)\right). \tag{4.10}$$

Equation (4.10) identifies the one-to-one mapping between the price and the signal for a given (θ, σ_j). 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 σ_j:

$$E(\theta|\hat{\theta}_j, D = j; \sigma_j) = \frac{\int_\theta \theta Pr(D = j|p_j^o(\hat{\theta}_j; \theta, \sigma_j), \theta)\phi(\hat{\theta}_j; \theta, \sigma_j) f_0(\theta) d\theta}{\int_\theta Pr(D = j|p_j^o(\hat{\theta}_j; \theta, \sigma_j), \theta)\phi(\hat{\theta}_j; \theta, \sigma_j) f_0(\theta) d\theta}. \tag{4.11}$$

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

Once we have obtained $E(\theta|\hat{\theta}_j, D = j; \sigma_j)$, we can easily derive the model-implied joint distribution of premiums and risk types for a given σ_j, because the pricing coefficients (α_j, β_j) have been recovered from the previous step. Matching this distribution to the observed empirical pattern provides the key identifying restriction for the signal distribution. Intuitively, when the signal distribution is very informative (i.e., σ_j is small), consumers' risks are more accurately reflected in their premiums, so the correlation between the premium and risk type within a firm would be higher. We formalize this identification intuition in Appendix C. We show that when the risk rating $E(\theta|\hat{\theta}_j, D = j; \sigma_j)$ 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 has been recovered from previous steps) decreases monotonically with σ_j.

Given the identification argument, we estimate the firm-specific signal distribution by matching the model-implied joint distribution of premiums and risk types within each firm to the empirical distribution recovered from 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 burden. Details of the estimation procedure are provided in Appendix B.

Cost Parameters 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., $f(\hat{\boldsymbol{\theta}}|\theta) = \prod_{j'=1}^{J} \phi(\hat{\theta}_{j'}; \theta, \sigma_{j'})$. Taking the first-order derivatives of the profit function with respect to

(α_j, β_j) yields

$$\frac{\partial \pi_j(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \alpha_j} = \int_\theta \int_{\hat{\boldsymbol{\theta}}} Pr(D = j|\hat{\boldsymbol{\theta}}, \theta) f(\hat{\boldsymbol{\theta}}|\theta) f_0(\theta) d\hat{\boldsymbol{\theta}} d\theta$$
$$+ \quad \int_\theta \int_{\hat{\boldsymbol{\theta}}} (\alpha_j + \beta_j \theta + c_j - k_j \theta) \frac{\partial Pr(D = j|\hat{\boldsymbol{\theta}}, \theta)}{\partial \alpha_j} f(\hat{\boldsymbol{\theta}}|\theta) f_0(\theta) d\hat{\boldsymbol{\theta}} d\theta, \quad (4.12)$$
$$\frac{\partial \pi_j(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \beta_j} = \int_\theta \int_{\hat{\boldsymbol{\theta}}} \theta Pr(D = j|\hat{\boldsymbol{\theta}}, \theta) f(\hat{\boldsymbol{\theta}}|\theta) f_0(\theta) d\hat{\boldsymbol{\theta}} d\theta$$
$$+ \quad \int_\theta \int_{\hat{\boldsymbol{\theta}}} (\alpha_j + \beta_j \theta + c_j - k_j \theta) \frac{\partial Pr(D = j|\hat{\boldsymbol{\theta}}, \theta)}{\partial \beta_j} f(\hat{\boldsymbol{\theta}}|\theta) f_0(\theta) d\hat{\boldsymbol{\theta}} d\theta. \quad (4.13)$$

In Equations (4.12) and (4.13), 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. These two first-order conditions provide a system of equations to identify (c_j, k_j), where all other terms are either directly estimable from the data or have been recovered in previous steps. We provide details on how to solve (c_j, k_j) in Appendix B.

5 Results

We apply our model and estimation strategy to the Italian auto insurance industry. Our sample consists of 124,428 liability insurance contracts sold to new customers in Rome's metropolitan area between 2013 and 2021. We focus 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 and supporting external validation.

5.1 Demand-Side Estimation Results

Our demand estimation allows consumer preference parameters to depend on both risk type and observable characteristics through parametric forms. Specifically, to capture potential variation in preferences for unobserved product attributes across risk types, geography, and time, we estimate ξ_j separately for each firm across eight demographic groups. These groups are defined based on whether a consumer's risk type is above or below the median, whether they are located in a big city, and whether the contract was signed in the first or second half of the sample period. We also allow γ to vary with age and location to capture potential heterogeneity in responsiveness to price across demographic groups.

We report the estimates of the demand parameters in Table 2. Standard errors, shown in parentheses, are computed using bootstrap methods, with details provided in Appendix B. We find that senior drivers tend to be less price sensitive, whereas drivers located in major cities are more price sensitive. Estimated preferences for unobserved product attributes are generally similar across low- and high-risk groups. However, one notable difference is that lower-risk consumers appear to have a stronger preference for Firms 3 and 5, while higher-risk consumers derive greater utility from Firm 8. Consumers in smaller cities tend to prefer Firms 4 and 10, potentially reflecting differences in advertising intensity across geographic areas. Over time, we also observe an increase in consumer preference for unobserved attributes associated with Firm 11, which represents a group of fringe firms.

5.2 Supply-Side Estimation Results

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 3, 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 first two columns of Table 3, we find that firms differ dramatically not only in their baseline markups, but also in how much they adjust premiums with respect to their risk ratings. For example, Firms 8 and 9 charge a relatively lower baseline premium but their prices are much more sensitive to their risk ratings, while Firms 3, 5, 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.

We observe substantial variation in the precision of risk rating across firms, as measured by the standard deviation of their signal distributions (see the third column of Table 3). Based on our estimates, Firms 8 and 9 have the least advanced risk-rating technologies, with signal standard deviations more than 50–55% higher than that of Firm 3, which exhibits the highest precision in risk assessment. 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 0.78.

Another notable finding from Table 3 is that firms derive substantial net benefits from contracting with new customers, potentially due to consumer inertia, dynamic pricing

Table 2: Estimates of demand-side parameters

(A) Price sensitivity parameter	
Constant	2.11
	(0.26)
Old	-1.21
	(0.20)
Big city	0.45
	(0.23)

(B) Preferences for unobserved product attributes

Firm ID	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8
2	-0.54	-0.52	-0.88	-0.76	-0.95	-0.90	-1.41	-1.36
	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
3	-0.38	-0.62	-0.51	-0.58	-1.49	-1.60	-1.58	-1.64
	(0.05)	(0.06)	(0.05)	(0.05)	(0.07)	(0.07)	(0.06)	(0.07)
4	-1.18	-1.19	-1.90	-1.81	-1.15	-1.05	-1.91	-1.84
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
5	-1.20	-1.50	-1.82	-1.95	-1.35	-1.73	-1.60	-2.01
	(0.06)	(0.07)	(0.06)	(0.08)	(0.07)	(0.08)	(0.06)	(0.06)
6	-0.31	-0.21	-0.36	-0.43	-0.70	-0.56	-0.51	-0.57
	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
7	-1.94	-2.01	-2.23	-2.09	-1.79	-2.01	-2.39	-2.20
	(0.08)	(0.08)	(0.08)	(0.07)	(0.08)	(0.08)	(0.09)	(0.07)
8	-1.82	-1.19	-1.56	-1.08	-2.12	-1.39	-2.16	-1.51
	(0.07)	(0.05)	(0.07)	(0.05)	(0.08)	(0.06)	(0.08)	(0.06)
9	-2.56	-2.66	-2.61	-2.26	-2.86	-3.29	-2.90	-3.17
	(0.10)	(0.14)	(0.09)	(0.09)	(0.11)	(0.16)	(0.11)	(0.14)
10	-1.56	-1.75	-2.18	-2.28	-1.48	-1.74	-1.97	-2.18
	(0.06)	(0.08)	(0.08)	(0.09)	(0.07)	(0.08)	(0.07)	(0.08)
11	0.32	0.57	0.54	0.71	0.93	0.72	1.04	0.80
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)
High risk	N	Y	N	Y	N	Y	N	Y
Big city	N	N	Y	Y	N	N	Y	Y
Later periods	N	N	N	N	Y	Y	Y	Y

Note: 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 zero across all groups. Demographic characteristics for each group are summarized in the bottom panel.

Table 3: Estimates of supply-side parameters

Firm ID	Pricing Coefficients		Signal Std. Dev.	Net Benefits	Claim Efficiency
	α_j	β_j	σ_j	c_j	k_j
1	-342.22	1.72	1339.05	1165.31	1.90
	(48.56)	(0.10)	(56.12)	(346.43)	(0.34)
2	-333.44	1.81	1217.08	1087.50	1.98
	(80.47)	(0.17)	(83.71)	(349.50)	(0.37)
3	-163.18	1.65	1053.16	1110.60	2.29
	(43.31)	(0.10)	(72.66)	(280.35)	(0.26)
4	-315.64	1.45	1178.77	1034.33	1.60
	(58.59)	(0.12)	(72.82)	(237.37)	(0.20)
5	-194.72	1.65	1117.57	922.39	1.86
	(61.41)	(0.16)	(85.78)	(290.10)	(0.31)
6	-310.08	1.47	1301.49	943.08	1.37
	(43.47)	(0.09)	(60.30)	(273.87)	(0.24)
7	-220.93	1.50	1118.80	858.29	1.54
	(90.81)	(0.19)	(115.56)	(293.74)	(0.33)
8	-1404.66	3.00	1580.41	2132.07	3.16
	(252.91)	(0.39)	(119.79)	(371.91)	(0.49)
9	-688.85	2.15	1637.52	1440.84	2.45
	(312.62)	(0.57)	(172.54)	(424.11)	(0.74)
10	-246.68	1.59	1245.79	972.73	1.79
	(138.25)	(0.28)	(107.35)	(317.01)	(0.39)
11	-158.71	1.19	1139.30	1435.99	1.56
	(23.58)	(0.05)	(47.07)	(435.10)	(0.36)

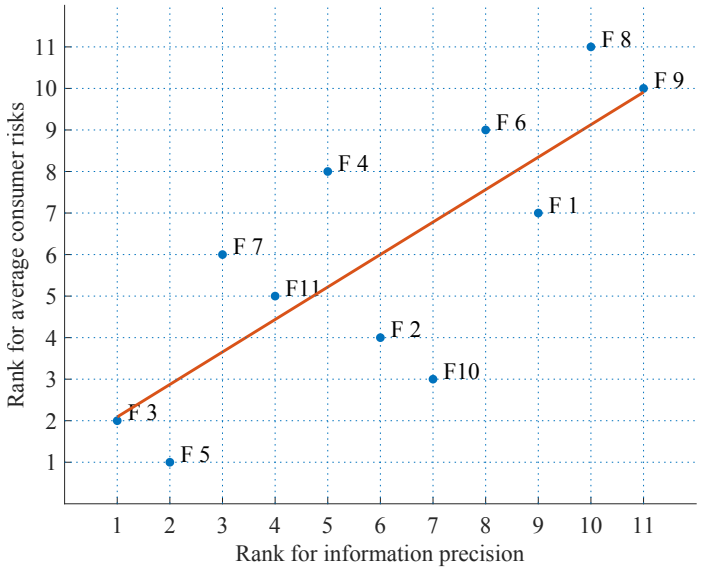


Figure 4: Who goes where: Rank correlation between information precision and average consumer risks across firms. Firms are ranked by information precision (highest to lowest on the x-axis) and average consumer risks (lowest to highest on the y-axis). The red solid line represents a linear fit between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

strategies, and opportunities for cross-selling other products. The net benefits received by Firm 8 are more than twice those of several other firms, including Firms 4–7 and 10. Firms also vary significantly in their efficiency in processing claims, as shown in the last column of Table 3, with Firm 6 exhibiting the lowest claim processing cost multiplier among all firms. Although this paper does not develop a fully dynamic model to capture consumers' switching behavior and firms' dynamic pricing strategies, we perform a back-of-the-envelope calculation based directly on the data to approximate the future benefits firms may receive. We then use this to isolate each firm's marginal cost mc_j. Details of the procedure for recovering mc_j are provided in Appendix B.

To examine the relationship between different dimensions of supply-side primitives, Table 4 reports the correlation coefficients between firms' signal standard deviations, marginal costs of servicing customers, and claim processing efficiency. The corresponding p-values are reported in parentheses. A key finding from our estimation is a statistically significant negative correlation between σ and mc, indicating that firms with lower information precision tend to have lower marginal 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 negative correlation between the two dimensions of firms' cost structures, suggesting that

firms with higher marginal costs may be more efficient at processing claims.

Table 4: Correlation coefficients between firms' signal standard deviations, marginal costs, and claim processing efficiency.

	σ_j	mc_j	k_j
σ_j	1.00		
mc_j	-0.72	1.00	
	(0.01)		
k_j	0.64	-0.75	1.00
	(0.03)	(0.01)	

5.3 Model Fit

Using the estimates in Tables 2 and 3, we simulate the premiums offered by each firm and consumers' plan choices given the simulated price menu. Table 6 in the appendix presents the means and standard deviations of risk and premium within each firm, using both real and simulated data. Overall, our model matches the key moments of the data reasonably well. Our model-generated price distributions 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.[26]

We also evaluate the out-of-sample fit of our model. Specifically, we randomly select 80% of the observations to estimate the model parameters, then simulate premiums and consumer choices for the remaining 20% of the sample. Table 7 in the appendix compares observed moments in the testing dataset with those simulated using our estimates. Again, we find that the model replicates key out-of-sample patterns well.

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 $E_j(\theta|\hat{\theta}_j, D = j)$—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 increases with the signal of Firm 1, for both young and senior drivers. Similar patterns are observed for all other firms.

Finally, we provide external validation of our estimation results. Since firms' risk-rating technologies are proprietary and generally unobservable, we turn to indirect measures to assess the plausibility of our findings. To proxy for the sophistication of a firm's actuarial

[26]The simulated market shares for low- and high-risk consumers also closely match the patterns observed in the data.

team, we use LinkedIn data to count the number of employees at major insurers who list expertise in machine learning, data science, or artificial intelligence. The idea is that a larger number of such engineers indicates more advanced pricing capabilities. In Figure 8a in the appendix, we plot the firms' information precision rankings against the rankings based on the number of these engineers and find a moderately positive correlation.

Second, to proxy for service quality, we hand-collect data on the number of service centers each major company operates in the Rome metropolitan area, using information from their websites. We compare this measure to our estimates of firms' unobserved product quality (ξ) and find a positive rank correlation, as shown in Figure 8b.

Third, we validate our marginal cost estimates using firms' financial statements. We collect data from the balance sheets of seven major insurers on their reported expenditures related to customer service and claim liquidation. These cost components closely match our estimates of marginal cost and claim processing efficiency. Our average estimated marginal cost (62 euros) for these companies aligns closely with the reported average (68 euros). Furthermore, the firm-level rankings based on balance sheet data are highly correlated with the rankings derived from our estimated marginal costs and claim efficiency parameters, as shown in Figures 8c and 8d.

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,

$$E(\theta|\hat{\boldsymbol{\theta}}) = \int_\theta \theta f(\theta|\hat{\boldsymbol{\theta}}) d\theta = \frac{\int_\theta \theta \left(\prod_j \phi(\hat{\theta}_j; \theta, \sigma_j) \right) f_0(\theta) d\theta}{\int_\theta \left(\prod_j \phi(\hat{\theta}_j; \theta, \sigma_j) \right) f_0(\theta) d\theta}, \tag{6.1}$$

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 $\mathcal{N}(\mu_0, \sigma_0^2)$, then the poste-

rior mean in Equation (6.1) can be expressed in the following closed-form formula:

$$E(\theta|\hat{\boldsymbol{\theta}}) = \frac{\frac{\mu_0}{\sigma_0^2} + \sum_{j=1}^{J} \frac{\hat{\theta}_j}{\sigma_j^2}}{\frac{1}{\sigma_0^2} + \sum_{j=1}^{J} \frac{1}{\sigma_j^2}}.$$

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 firms' concerns about 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.[27] 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 asymmetries in information technology between competing firms, but at the same time, *reduces* the overall information availability in the market.[28]

For each counterfactual scenario, we re-solve for the market equilibrium. We propose an iterative procedure to solve for all firms' pricing coefficients, beginning from the status quo, with the details described in Appendix D. While the equilibrium of the pricing game may not be unique, we argue that our approach is informative and likely reflects how firms adjust their strategies in practice. 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., observing true risk, centralized risk bureau, and privacy regulation) in Table 5. The percentage changes relative to the baseline

[27] 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).

[28] We also conduct a counterfactual experiment to evaluate market outcomes when one firm's information technology is improved, thereby assessing the value of information. We separately quantify (1) the direct effect, where better information allows a firm to improve its risk assessment and pricing, and (2) the equilibrium effect. Detailed results and discussion are provided in Section S3 of the Supplementary Materials.

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.

Table 5: Counterfactual results: The impact of information policies

	Baseline	Observing True Risk	Centralized Risk Bureau	Privacy Regulation
Average CS (€)	-542.83	-451.05 (+16.91%)	-457.59 (+15.70%)	-523.42 (+3.57%)
Average CS: Low risk (€)	-477.24	153.74 (+132.21)	-103.63 (+78.28%)	-480.25 (-0.63%)
Average CS: High risk (€)	-608.42	-1055.83 (-73.54%)	-811.55 (-33.39%)	-566.60 (+6.87%)
Average premium (€)	461.25	342.56 (-25.73%)	361.61 (-21.60%)	441.79 (-4.22%)
Average profit (€)	849.58	802.11 (-5.59%)	799.29 (-5.92%)	834.38 (-1.79%)
HHI	2297.26	2241.40 (-2.43%)	2264.29 (-1.44%)	2303.38 (+0.27%)
Average cost (€)	882.39	849.85 (-3.69%)	869.95 (-1.41%)	884.15 (+0.20%)
Between-firm info asymmetries	Yes	No	No	No
Information availability		Increase	Increase	Decrease

Table 5 shows that eliminating information asymmetries between firms generally benefits consumers. When all firms observe consumers' true risk types, consumer surplus increases by 16.91%. Granting all firms equal access to risk scores from a centralized risk bureau raises consumer surplus by 15.70%, nearly matching the efficiency benchmark. If all firms were to adopt the least advanced risk-rating technology under privacy regulation, consumer surplus would rise by 3.57%.

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 78.28% increase in consumer surplus, while high-risk consumers see a 33.39% 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.

We also evaluate the impact of the three counterfactual policies across different consumer demographic groups. Table 8 in Appendix A shows that eliminating information asymmetries between firms has a similar effect across age groups but systematically provides greater benefits to consumers located in major cities.

6.2 Firm Profit

In all three counterfactual scenarios, we observe a decrease in the average premium. When information asymmetries between firms are 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.

A more nuanced channel behind the price reduction is the dampening effect of price dispersion on competition. In the baseline scenario, imprecise information leads to wide variation in the prices offered to a given consumer across firms. This weakens individual firms' incentives to compete aggressively on price, since even a price cut may not make their offer the most attractive. However, when firms can perfectly observe risks or when a centralized bureau is in place, 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 21.60%–25.73% 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 5).[29] 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 8 and 9, benefit the most. In contrast, firms with more advanced risk-rating precision, like Firms 3, 5, and 7, experience significant profit losses under both counterfactual scenarios, regardless of whether the policies improve or reduce overall risk-rating precision in the market.

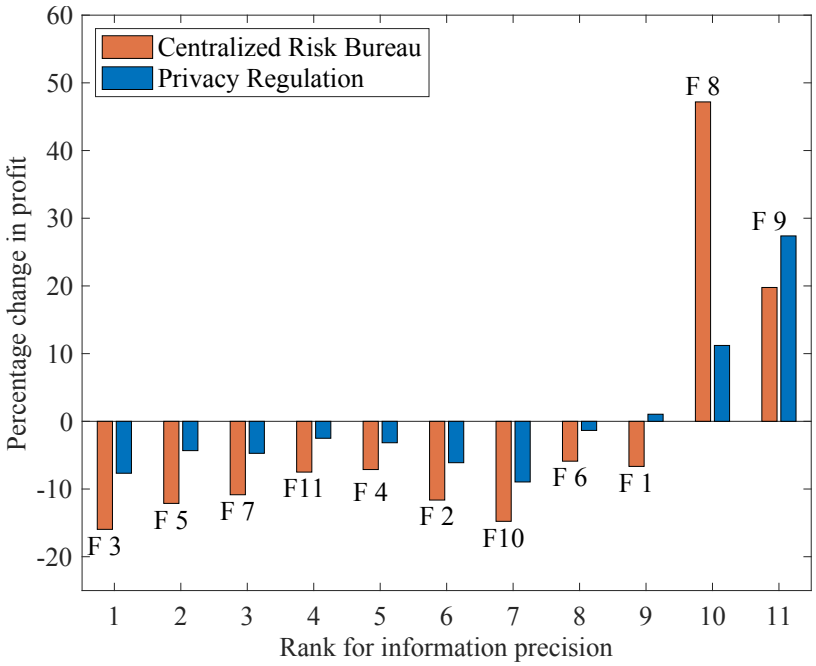


Figure 5: Differential impact on firms: Percentage change in profit relative to the baseline. Firms are ranked from the highest to the lowest risk-rating precision on the x-axis. The orange bars represent the percentage changes in profit under the centralized risk bureau scenario, while the blue bars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

[29]Our counterfactual assumes that firms share identical risk evaluations, effectively eliminating informational differences across firms, whereas ANIA's proposal (c.f. Footnote 2) appears to focus on expanding access to more detailed information from individual driving records. In practice, however, even if firms observe the same underlying driving record, they may differ substantially in how they interpret and model this information, depending on their actuarial methods, machine learning tools, and overall information-processing capabilities. Moreover, ANIA's support for data sharing may reflect broader industry objectives, such as improving market transparency, reducing fraud, or meeting regulatory requirements. Our counterfactual analysis serves as a benchmark that complements the policy proposal by showing how information heterogeneity can shape firm behavior and market outcomes.

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' equilibrium sorting patterns, we plot the average risk within firms for both the baseline and counterfactual scenarios in Figure 6.[30] When firms can observe the true risk, 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, 6 and 7, 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. Market segmentation driven by targeted pricing strategies remains similarly pronounced when all firms have equal access to risk scores from the centralized risk bureau, as illustrated by the red 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 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 5 shows that under the efficiency benchmark, average costs decrease by 3.69% relative to the baseline. This reduc-

[30] We do not find substantial changes in market shares for most firms across counterfactual scenarios relative to the baseline. The notable exception is Firm 8, whose market share nearly doubles when a centralized risk bureau is implemented or when firms can observe true consumer risk.

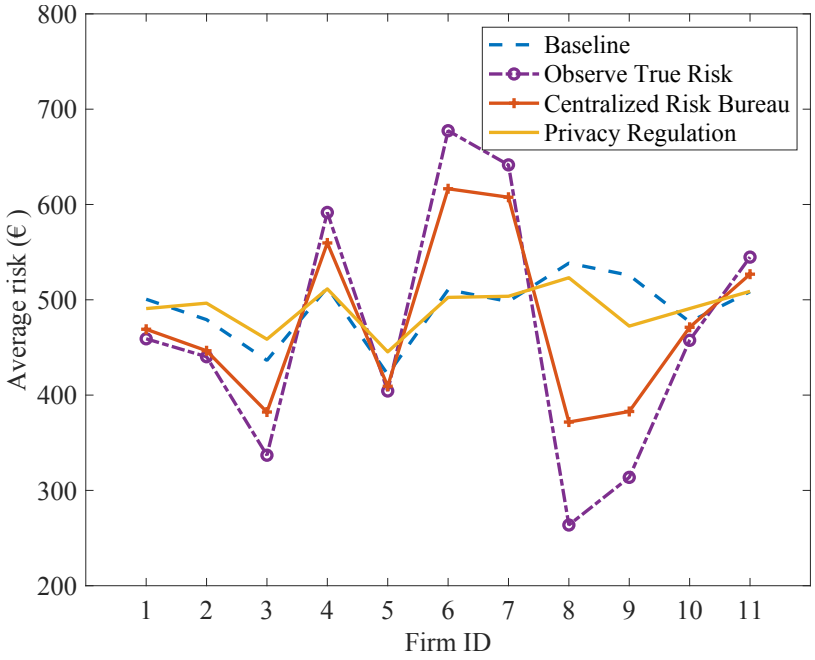


Figure 6: Sorting patterns: Average risk levels among consumers (measured in euros) within each firm under the baseline and three counterfactual scenarios.

tion, driven by improved consumer reallocation, amounts to approximately 32 euros per contract per year, equivalent to about 7% of the annual premiums paid by consumers. Under the centralized risk bureau scenario, the average cost is reduced by 12 euros per contract, capturing around 40% of the savings achieved under the infeasible efficiency benchmark. Given the size of the Italian auto insurance market (31 million contracts underwritten in 2018), this cost reduction translates into a substantial aggregate impact.

Our findings highlight that in mandatory insurance markets, where inefficiencies primarily stem from the misallocation of consumers across firms, better utilization and aggregation of existing market information can substantially reduce information asymmetries across firms and 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 receive 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.[31] 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.

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, which in turn boosts consumer surplus by 15.7%, 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 12 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).

[31] 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.

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. We also abstract from modeling consumers' active choice of contract features and treat insurers' contract designs as given. In the counterfactual analysis where information policies are changed, firms may have incentives to adjust their contract offerings, which could further complicate the policy implications.[32] In this paper, we do not explicitly model the extensive margin of consumer selection into or out of the market. However, since insurance companies are legally prohibited from rejecting applicants and driving without mandatory liability insurance is illegal, we believe the magnitude of this effect is likely limited.

There are several promising avenues for future research. On the demand side, incorporating more flexible features—such as consumer switching, search frictions, limited consideration sets, and multidimensional private information—would allow for a more realistic representation of consumer behavior. On the supply side, modeling additional aspects of competition, such as dynamic pricing strategies and endogenous contract design, would provide a deeper understanding of firm behavior and the implications of policy interventions.

[32]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 (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).

A Additional Tables and Figures

Table 6: Model fit: Comparing data moments with simulated results using model estimates

(A): Data Moments				
Firm ID	Risk (€)		Premium (€)	
	Mean	Std. Dev.	Mean	Std. Dev.
1	499.86	452.28	519.41	209.20
2	492.41	482.88	556.93	221.56
3	443.83	440.64	568.17	215.58
4	512.60	466.64	427.57	199.30
5	424.22	457.41	506.42	201.00
6	517.59	491.72	448.34	173.35
7	498.45	491.92	526.00	208.25
8	659.83	504.36	576.96	227.48
9	576.66	585.70	551.46	224.70
10	483.60	505.35	523.21	211.78
11	493.60	488.22	430.72	197.85

(B): Simulated Moments Using Model Estimates				
Firm ID	Risk (€)		Premium (€)	
	Mean	Std. Dev.	Mean	Std. Dev.
1	509.84	485.84	509.01	219.60
2	500.17	488.72	539.52	235.64
3	460.28	460.80	550.99	230.75
4	515.58	483.77	423.76	218.02
5	452.92	485.62	508.31	222.33
6	510.80	482.12	434.19	182.62
7	521.76	499.82	514.10	217.81
8	605.84	500.19	536.30	242.47
9	514.71	456.28	513.21	233.13
10	488.03	480.98	507.44	218.85
11	489.59	476.88	433.29	202.53

Table 7: Out-of-Sample Model fit: Comparing data moments with simulated results using model estimates

(A): Data Moments				
Firm ID	Risk (€)		Premium (€)	
	Mean	Std. Dev.	Mean	Std. Dev.
1	491.54	433.31	518.21	209.08
2	469.38	473.34	551.70	227.42
3	439.84	437.73	569.71	215.55
4	519.26	518.50	422.11	193.13
5	413.86	463.96	509.75	208.92
6	522.25	513.97	453.88	176.49
7	508.08	523.94	523.45	205.30
8	657.99	528.12	560.24	225.17
9	520.82	537.65	549.24	218.20
10	481.89	542.93	520.13	215.83
11	502.96	497.22	431.61	196.74

(B): Simulated Moments Using Model Estimates				
Firm ID	Risk (€)		Premium (€)	
	Mean	Std. Dev.	Mean	Std. Dev.
1	519.76	483.72	517.13	222.65
2	500.00	498.94	545.04	238.90
3	452.54	453.30	544.16	228.07
4	517.04	485.32	426.80	222.97
5	454.59	498.71	507.76	224.42
6	497.24	473.24	432.61	186.95
7	501.33	509.23	516.32	217.94
8	617.42	512.31	539.57	247.00
9	499.14	434.08	506.36	227.00
10	449.04	451.49	518.96	228.27
11	495.87	497.66	439.21	205.51

Table 8: Counterfactual results: Distributional Effects on Consumer Surplus

	Baseline	Observing True Risk	Centralized Risk Bureau	Privacy Regulation
Average CS: Young drivers (€)	-732.31	-609.75 (+16.74%)	-603.37 (+17.61%)	-706.81 (+3.48%)
Average CS: Senior drivers (€)	-366.90	-303.70 (+17.23%)	-322.24 (+12.17%)	-353.16 (+3.75%)
Average CS: Small city (€)	-539.12	-483.16 (+10.38%)	-476.27 (+11.66%)	-525.56 (+2.52%)
Average CS: Big city (€)	-545.10	-431.35 (+20.87%)	-446.14 (+18.16%)	-522.11 (+4.22%)

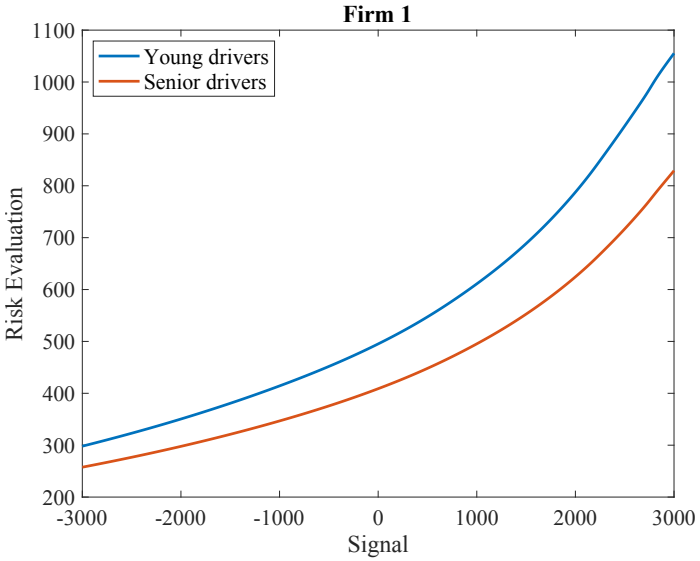


Figure 7: Risk rating versus signal for Firm 1: The y-axis represents the risk rating, and the x-axis denotes the signal. The blue and red curves represent young and senior drivers, respectively.

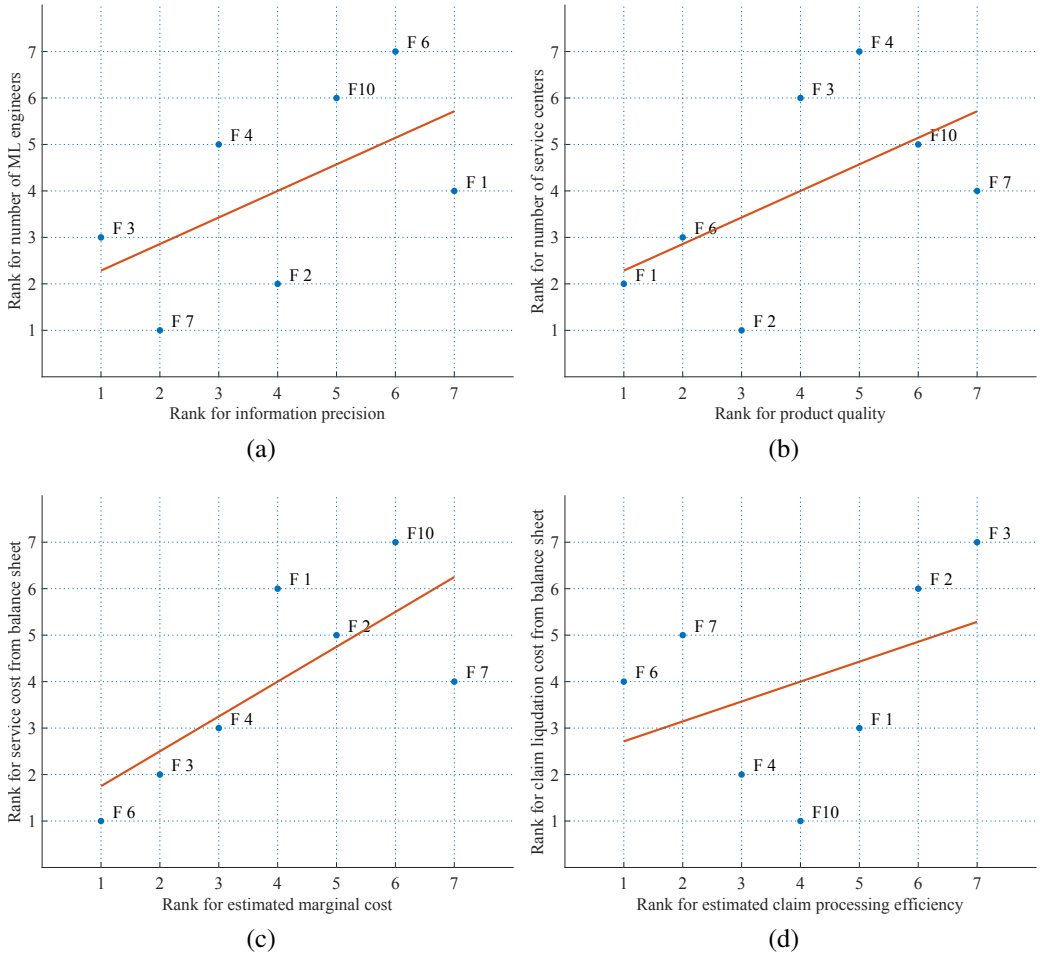


Figure 8: Rank correlations between model estimates and external measures for seven major insurers. Panel (a): Firms ranked by information precision (x-axis) and number of ML engineers (y-axis), both from highest to lowest. Panel (b): Firms ranked by average product quality (x-axis) and number of service centers (y-axis), both from highest to lowest. Panel (c): Firms are ranked on the x-axis by estimated marginal cost and on the y-axis by service-related expenses (as a percentage of total cost, based on balance sheet data), with both axes ordered from lowest to highest. Panel (d): Firms are ranked on the x-axis by estimated claim processing efficiency and on the y-axis by claim liquidation costs (measured as a percentage of total cost from balance sheet data), with both axes ordered from lowest to highest. The red solid line in each figure represents a linear fit between the two rankings. Firm IDs are displayed next to each dot; firms with incomplete data are excluded from the analysis.

B Estimation Details

We describe the step-by-step estimation procedure for the model. Let $i = 1, 2, \ldots, N$ index individual consumers. For each consumer i, we observe the number of accidents in period t, denoted by y_{it} for $t = 1, 2, \ldots, T$. Let p_i denote the premium paid, and $D_i \in \{1, 2, \ldots, J\}$ represent the contract choice of consumer i. The estimation procedure described in this section can be carried out conditional on observable consumer characteristics, which we omit for notational simplicity.

Step 1: Recovering Risk Type Distribution We assume that the Poisson rate can take on a finite set of values $\{\lambda_1, \lambda_2, \ldots, \lambda_L\}$. Let $\boldsymbol{q} = (q_1, q_2, \ldots, q_L)$ denote the vector of probabilities associated with each Poisson rate such that $\sum_{l=1}^{L} q_l = 1$ and $q_l \geq 0, \forall l$. Define $I_j(\underline{p}, \bar{p})$ as the set of consumer indices such that for all $i \in I_j(\underline{p}, \bar{p})$, we have $p_i \in [\underline{p}, \bar{p}]$ and $D_i = j$. We estimate the probabilities of each Poisson rate for this group of consumers by maximizing the following log-likelihood function.[33]

$$LL_\lambda(\boldsymbol{q}; I_j(\underline{p}, \bar{p})) = \sum_{i \in I_j(\underline{p}, \bar{p})} \log \left(\sum_{l=1}^{L} \frac{\lambda_l^{(\sum_{t=1}^{T} y_{it})} \exp(-\lambda_l T)}{\prod_{t=1}^{T} y_{it}!} q_l \right).$$

Given the estimated probability distribution of the Poisson rate for consumers within each group, we randomly draw a Poisson rate $\tilde{\lambda}_i$ for each consumer i. We then construct a simulated risk type for that consumer as $\tilde{\theta}_i = \hat{\mu}\tilde{\lambda}_i$, where $\hat{\mu}$ is the estimated expected claim size based on the log-normal regression in Equation (2.2). Using the sample $(p_i, \tilde{\theta}_i, D_i)$, we estimate the density of premiums conditional on risk type and contract choice. Specifically, we discretize risk types into 20 bins and apply kernel density estimation to obtain $\hat{g}(p|\theta, D = j)$, the estimated premium density conditional on each risk type and the consumer's chosen contract. This serves as the output of the first-step estimation.

Step 2: Estimating Demand Parameters We use a nested fixed-point algorithm to estimate the demand parameters, taking the first-step estimates $\hat{g}(p|\theta, D = j)$ as an input. In the inner loop, we fix the price sensitivity parameter γ and apply the iterative procedure described in Section 4.2 to solve for the vector of unobserved product attributes $\boldsymbol{\xi}(\gamma)$ and the offered price distribution $g_j(p|\theta; \gamma)$. We evaluate the offered price distribution at 2,500 grid points. Convergence of the fixed-point algorithm requires that, for all firms and at all grid

[33] In practice, we discretize the observed premiums into 100 bins, each with a width of approximately 23 euros. We experiment with allowing the Poisson rate to take between 10 and 100 discrete values and find that the results are very similar. We ultimately choose to use 50 discrete values for the Poisson rate.

points, the difference between successive iterations falls below a pre-specified tolerance level.

In the outer loop, we estimate γ by maximizing the following log-likelihood function.

$$LL_d(\gamma) = \sum_{i=1}^{N} \sum_{j=1}^{J} \mathbf{1}\{D_i = j\} \log \Big(Pr(D_i = j|\tilde{\theta}_i; \gamma) \Big),$$

where $Pr(D_i = j|\tilde{\theta}_i; \gamma)$ denotes the model-implied probability that consumer i chooses a contract from firm j conditional on risk type $\tilde{\theta}_i$, and is constructed as follows:

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

To compute the choice probability $Pr(D = j|\theta; \gamma)$, we simulate offered prices from $g_j(p|\theta; \gamma)$ and evaluate the probabilities using numerical integration. Once we obtain an estimate $\hat{\gamma}$ for the price sensitivity parameter, the vector of unobserved product attributes $\boldsymbol{\xi}(\hat{\gamma})$ and the offered price distribution $g_j(p|\theta; \hat{\gamma})$ are recovered as by-products.

Step 3: Estimating Pricing Coefficients Equations (4.7) and (4.8) uniquely determine α_j and β_j as follows:

$$\alpha_j = E(p|D = j) - \frac{var(p|D = j)}{cov(p, \theta|D = j)} E(\theta|D = j),$$
$$\beta_j = \frac{var(p|D = j)}{cov(p, \theta|D = j)}.$$

Using the sample $(p_i, \tilde{\theta}_i, D_i)$, we estimate $E(p|D = j)$, $E(\theta|D = j)$, $var(p|D = j)$ and $cov(p, \theta|D = j)$ and use these estimates to compute $\hat{\alpha}_j$ and $\hat{\beta}_j$.

Alternatively, α_j and β_j can be identified and estimated using a linear regression of θ on prices p_j. To see this, note that the pricing equation (3.2) is equivalent to the following:

$$\theta = -\frac{\alpha_j}{\beta_j} + \frac{p_j}{\beta_j} + \underbrace{(\theta - E(\theta|\hat{\theta}_j, D = j))}_{\text{error term}},$$

where $cov(p_j, \theta - E(\theta|\hat{\theta}_j, D = j)|D = j) = 0$. Regressing θ on p_j for consumers within firm j identifies the coefficients $-\frac{\alpha_j}{\beta_j}$ and $\frac{1}{\beta_j}$.

Step 4: Estimating Signal Variance We first derive the posterior joint distribution of signals and risk types for any given σ_j using the Bayes rule:

$$f(\hat{\theta}_j, \theta | D = j; \sigma_j) = \frac{Pr(D = j | p_j^o(\hat{\theta}_j; \theta, \sigma_j), \theta) \phi(\hat{\theta}_j; \theta, \sigma_j) f_0(\theta)}{s_j},$$

where $Pr(D = j | p_j^o(\hat{\theta}_j; \theta, \sigma_j), \theta)$ represents the likelihood of selecting firm j conditional on the signal $\hat{\theta}_j$ (or equivalently the corresponding price $p_j^o(\hat{\theta}_j; \theta, \sigma_j)$) and type θ. The denominator s_j represents the market share of firm j and serves as a normalization factor.

Next, we derive the model-generated price as a function of the signal, which we denote by $gp_j(\hat{\theta}_j; \sigma_j)$, taking the estimated pricing coefficients $\hat{\alpha}_j$ and $\hat{\beta}_j$ as given. Specifically,

$$gp_j(\hat{\theta}_j; \sigma_j) = \hat{\alpha}_j + \hat{\beta}_j E(\theta | \hat{\theta}_j, D = j; \sigma_j), \tag{B.1}$$

where $E(\theta | \hat{\theta}_j, D = j; \sigma_j)$ represents the equilibrium risk rating given signal $\hat{\theta}_j$ as a function of σ_j (see Equation (4.11)).

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

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

We estimate σ_j for each firm j by maximizing the following likelihood function:

$$LL_p(\sigma_j) = \sum_{i=1}^{N} \mathbf{1}\{D_i = j\} \log \Big(h(p_i, \tilde{\theta}_i | D = j; \sigma_j) \Big).$$

Step 5: Estimating Cost Parameters The key challenge in evaluating the first-order conditions lies in estimating the derivatives of the sorting probabilities with respect to α_j and β_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. However, changing firm j's pricing coefficients affects its own prices through two channels: (1) the direct effect, and (2) the indirect effect through the equilibrium risk rating $E(\theta | \hat{\theta}_j, D = j)$.

To see this,

$$\begin{aligned}\frac{\partial Pr(D=j|\hat{\boldsymbol{\theta}},\theta)}{\partial \alpha_j} &= -\gamma(\theta)\frac{\partial p_j(\hat{\theta}_j)}{\partial \alpha_j} Pr(D=j|\hat{\boldsymbol{\theta}},\theta)(1-Pr(D=j|\hat{\boldsymbol{\theta}},\theta)) \\ &= -\gamma(\theta)\left[1+\beta_j\frac{\partial E(\theta|\hat{\theta}_j,D=j)}{\partial \alpha_j}\right] Pr(D=j|\hat{\boldsymbol{\theta}},\theta)(1-Pr(D=j|\hat{\boldsymbol{\theta}},\theta)) \\ &\approx -\gamma(\theta)Pr(D=j|\hat{\boldsymbol{\theta}},\theta)(1-Pr(D=j|\hat{\boldsymbol{\theta}},\theta)). \qquad \text{(B.2)}\end{aligned}$$

The approximation in Equation (B.2) assumes that the impact of α_j on its own expected risk level is small. Similarly, we approximate the derivatives of sorting probabilities with respect to β_j in the following equation:

$$\frac{\partial Pr(D=j|\hat{\boldsymbol{\theta}},\theta)}{\partial \beta_j} \approx -\gamma(\theta)E(\theta|\hat{\theta}_j,D=j)Pr(D=j|\hat{\boldsymbol{\theta}},\theta)(1-Pr(D=j|\hat{\boldsymbol{\theta}},\theta)). \quad \text{(B.3)}$$

We verify the approximations in these two equations numerically by varying α_j and β_j for a single firm while holding all other firms' pricing strategies fixed. We then iterate to compute the new equilibrium risk rating $E(\theta|\hat{\theta}_j, D=j)$ using Step 2 of the iterative procedure described in Appendix D. We find that a 1% increase in α_j and β_j changes $E(\theta|\hat{\theta}_j, D=j)$ by an average of 0.01% and 0.04%, respectively. Our results confirm that the effect of α_j and β_j on the firm's own expected equilibrium risk rating is indeed small and is dominated by the direct effect of changing these pricing coefficients.

With the approximations in Equations (B.2) and (B.3), the first-order conditions in Equation (4.12) and (4.13) 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.

Recovering Marginal Cost Let I_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 I_j, which we denote by $\overline{P}_j$. Specifically,

$$\overline{P}_j = \frac{\sum_{i\in I_j}\sum_t^T \delta^t p_{it}\mathbf{1}\{i \text{ stays in firm } j \text{ at } t\}}{|I_j|}, \qquad \text{(B.4)}$$

where the discount factor δ is set to 0.95. We use a similar formula as in Equation (B.4) 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:

$$\underbrace{(\overline{P}_j - \overline{CS}_j k_j - \overline{N}_j mc_j)}_{\text{discounted value of future profits}} - mc_j = c_j. \tag{B.5}$$

Equation (B.5) isolates the firm's marginal cost mc_j partially from the dynamic factors. In practice, if insurance companies offer additional products to consumers, they may gain extra benefits from bundling or cross-selling, making our marginal cost estimates a lower bound.

Computing Standard Errors We compute standard errors for the demand- and supply-side parameter estimates using 200 bootstrap replications. In each replication, we resample individuals with replacement from the original dataset, preserving all observations associated with each selected individual. For each bootstrap sample, we repeat the full estimation procedure (Steps 1–5 described above) to recover the model primitives. We then compute the standard deviation of the resulting parameter estimates across replications to obtain the bootstrap standard errors.

C Identification of Signal Variance

We consider the identification of the variance of the signal distribution. The pricing coefficients have been recovered in the previous step and are thus treated as known. For simplicity, we consider the case where the demand parameters do not vary with risk type.

Under the assumption of normally distributed signals, the density function of firm j's signal distribution takes the following form:

$$\phi(\hat{\theta}_j;\theta,\sigma_j)=\frac{1}{\sqrt{2\pi\sigma_j^2}}\exp\left(-\frac{(\hat{\theta}_j-\theta)^2}{2\sigma_j^2}\right).$$

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.

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

$$\approx\frac{\exp(-\gamma p_j(\hat{\theta}_j))\phi(\hat{\theta}_j;\theta,\sigma_j)}{\int_{\hat{\theta}_j}\exp(-\gamma p_j(\hat{\theta}_j))\phi(\hat{\theta}_j;\theta,\sigma_j)d\hat{\theta}_j}.\tag{C.2}$$

Denote the term in square brackets in Equation (C.1) 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 $\hat{\theta}_j$. With this approximation, we obtain Equation (C.2).

If the risk-rating term $E(\theta|\hat{\theta}_j,D=j)$ can be well approximated by a linear function of the signal $\hat{\theta}_j$,[34] i.e., $E(\theta|\hat{\theta}_j,D=j)\approx a_j+b_j\hat{\theta}_j$, we can further simplify the posterior

[34]We empirically verify this assumption in Figure 7, where we plot the relationship between risk rating $E(\theta|\hat{\theta}_j,D=j)$ 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.

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

$$f(\hat{\theta}_j|\theta, D=j) \approx \frac{\exp(-\gamma\beta_j b_j \hat{\theta}_j)\phi(\hat{\theta}_j;\theta,\sigma_j)}{\int_{\hat{\theta}_j}\exp(-\gamma\beta_j b_j \hat{\theta}_j)\phi(\hat{\theta}_j;\theta,\sigma_j)d\hat{\theta}_j}$$
$$= \frac{1}{\sqrt{2\pi\sigma_j^2}}\exp\left(-\frac{(\hat{\theta}_j-(\theta-\gamma\beta_j b_j\sigma_j^2))^2}{2\sigma_j^2}\right). \tag{C.3}$$

Equation (C.3) 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. Let m_j denote the distribution of θ_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),

$$corr(p_j,\theta_j) = \frac{cov(p_j,\theta_j)}{\sqrt{var(\theta_j)var(p_j)}} = \sqrt{\frac{cov^2(p_j,\theta_j)}{var(p_j)var(\theta_j)}} = \sqrt{\frac{var(E(\theta_j|\hat{\theta}_j))}{var(\theta_j)}}. \tag{C.4}$$

The last equality in Equation (C.4) holds because, by Equation (4.8)

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

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

denote the density of $\hat{\theta}_j$ conditional on $D = j$. We have

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

where μ_j is the mean of θ_j, i.e., the average risk type of those self-select into firm j.

Next, we show that $var(E(\theta_j|\hat{\theta}_j^*))$ is monotonically decreasing in σ_j. Note that by Equation (C.3), $\hat{\theta}_j^* \sim \mathcal{N}(\theta, \sigma_j^2)$. Take an independent and normally distributed random variable η. Define

$$\hat{\theta}_j^{*\prime} = \hat{\theta}_j^* + \eta.$$

Since $\hat{\theta}_j^*$ and η are independent and normally distributed, $\hat{\theta}_j^{*\prime} \sim \mathcal{N}(\theta, \sigma_j^{\prime 2})$ with a larger variance $\sigma_j^{\prime 2} > \sigma_j^2$. By independence, the distribution of $E(\theta_j|\hat{\theta}_j^*)$ is the same as the distribution of $E(\theta_j|\hat{\theta}_j^{*\prime}, \eta)$, so $var(E(\theta_j|\hat{\theta}_j^*)) = var(E(\theta_j|\hat{\theta}_j^{*\prime}, \eta))$. Moreover,

$$var(E(\theta_j|\hat{\theta}_j^{*\prime}, \eta)) > var(E(\theta_j|\hat{\theta}_j^{*\prime})),$$

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

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.[35] 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.

[35]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 $E(\theta_j|\hat{\theta}_j)$ 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 $E(\theta|\hat{\theta}_j, D = j)$, and the selection probabilities $Pr(D = j|\hat{\theta}_j, \theta)$ 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 $E^{r,0}(\theta|\hat{\theta}_j, D = j) = E^{r-1}(\theta|\hat{\theta}_j, D = j)$.

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

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

   (3) Compute the choice probabilities:

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

   (4) Update the expectation by

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

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

3. Having obtained $E^r(\theta|\hat{\theta}_j, D = j)$ and $Pr^r(D = j|\hat{\theta}_j, \theta)$ 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,

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

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

ABALUCK, J. AND J. GRUBER (2016): "Evolving choice inconsistencies in choice of prescription drug insurance," *American Economic Review*, 106, 2145–2184.

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.

BENETTON, M. (2021): "Leverage regulation and market structure: A structural model of the UK mortgage market," *The Journal of Finance*, 76, 2997–3053.

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.

BLATTNER, L., J. HARTWIG, AND S. NELSON (2022): "Information design in consumer credit markets," *Working paper*.

BLATTNER, L. AND S. NELSON (2021): "How costly is noise? Data and disparities in consumer credit," *arXiv preprint arXiv:2105.07554*.

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.

CHATTERJEE, S., D. CORBAE, K. DEMPSEY, AND J.-V. RÍOS-RULL (2023): "A quantitative theory of the credit score," *Econometrica*, 91, 1803–1840.

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

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.

D'HAULTFŒUILLE, X., I. DURRMEYER, AND P. FÉVRIER (2019): "Automobile prices in market equilibrium with unobserved price discrimination," *The Review of Economic Studies*, 86, 1973–1998.

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.

HERTZBERG, A., J. M. LIBERTI, AND D. PARAVISINI (2011): "Public information and coordination: evidence from a credit registry expansion," *The Journal of Finance*, 66, 379–412.

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.

HU, Y. (2008): "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution," *Journal of Econometrics*, 144, 27–61.

——— (2017): "The Econometrics of Unobservables–Latent Variable and Measurement Error Models and Their Applications in Empirical Industrial Organization and Labor Economics," .

HU, Y. AND S. M. SCHENNACH (2008): "Instrumental variable treatment of nonclassical measurement error models," *Econometrica*, 76, 195–216.

HU, Y. AND M. SHUM (2012): "Nonparametric identification of dynamic models with unobserved state variables," *Journal of Econometrics*, 171, 32–44.

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.

KRASNOKUTSKAYA, E. (2011): "Identification and estimation of auction models with unobserved heterogeneity," *The Review of Economic Studies*, 78, 293–327.

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

LIBERMAN, A., C. NEILSON, L. OPAZO, AND S. ZIMMERMAN (2018): "The equilibrium effects of information deletion: Evidence from consumer credit markets," Tech. rep., National Bureau of Economic Research.

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. (2025): "Private information and price regulation in the US credit card market," *Econometrica*.

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.

XIN, Y. (2023): "Asymmetric information, reputation, and welfare in online credit markets," *Available at SSRN: https://ssrn.com/abstract=3580468*.

Supplementary Materials for "Competing under Information Heterogeneity: Evidence from Auto Insurance"

S1 Additional Empirical Evidence

Table S1: Selected Variables Used by Five Major Auto Insurance Companies (Not Included in Our Data)

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 S2: Regression of Premiums on Observable Characteristics

	Premium	Premium	Premium	Premium	Premium	Premium
VARIABLES	Firm 1	Firm 2	Firm 3	Firm 4	Firm 5	Firm 6
Age	-1.0561***	-0.7269***	-1.5116***	-0.2459	-0.4106**	-0.7838***
	(0.0888)	(0.1510)	(0.1312)	(0.1609)	(0.1734)	(0.0744)
BM	19.7425***	24.9416***	26.4326***	20.9453***	31.0295***	21.8787***
	(0.8648)	(1.1237)	(1.1970)	(1.9128)	(1.5464)	(0.4911)
Man	-8.9443***	-7.8892*	-2.6870	-14.7678***	-6.2815	3.9914**
	(2.3646)	(4.0410)	(3.6989)	(4.6169)	(4.9169)	(1.8632)
1 Acc.	124.1271***	149.2277***	106.1311***	74.1313***	112.2941***	111.7719***
	(4.0282)	(7.0702)	(5.6987)	(6.3382)	(8.8072)	(3.0169)
Big city	34.4064***	45.6070***	56.5082***	55.2540***	39.8449***	16.9238***
	(2.5688)	(4.3682)	(3.8889)	(4.4827)	(5.0987)	(2.0101)
Constant	639.0548***	620.9940***	575.3849***	515.6197***	594.6249***	511.3076***
	(16.5555)	(35.6947)	(26.2550)	(59.5533)	(86.1046)	(20.6469)
Observations	20,800	8,070	8,071	5,504	3,945	15,971
R-squared	0.3974	0.3903	0.4750	0.4644	0.4763	0.5887
VARIABLES	Firm 7	Firm 8	Firm 9	Firm 10	Firm 11	All Firms
Age	-1.3800***	-1.6772***	-0.9907***	-1.0137***	-0.5548***	-0.7635***
	(0.2277)	(0.2072)	(0.3104)	(0.2076)	(0.0468)	(0.0330)
BM	23.6438***	12.9392***	31.5349***	12.0464***	23.1305***	21.9900***
	(1.5264)	(2.1503)	(2.3238)	(2.0040)	(0.4227)	(0.2984)
Man	6.8855	-1.2936	-9.7677	-9.4804	-0.6410	-4.8530***
	(6.5698)	(5.6599)	(8.5327)	(6.1557)	(1.2818)	(0.8964)
1 Acc.	125.2567***	118.8831***	147.2334***	181.3849***	119.4853***	121.7660***
	(9.9382)	(9.1527)	(16.4997)	(12.3331)	(2.0065)	(1.4266)
Big city	48.7960***	56.1094***	84.5230***	52.8687***	12.3459***	28.6369***
	(7.3501)	(6.2824)	(10.1850)	(6.7416)	(1.4467)	(0.9829)
Constant	499.9007***	697.3515***	157.8634**	694.9461***	449.3726***	516.3972***
	(34.4620)	(26.2524)	(75.7489)	(60.5639)	(8.6075)	(7.1540)
Observations	2,525	4,289	1,355	2,992	50,906	124,428
R-squared	0.4515	0.3992	0.5101	0.4458	0.5234	0.4834

Note: All regressions reported in this table control for vehicle characteristics, time fixed effects, and optional contract clauses.

Table S3: Regression of Premiums on Estimated Risk Type and Observable Characteristics

	Premium	Premium	Premium	Premium	Premium	Premium
VARIABLES	Firm 1	Firm 2	Firm 3	Firm 4	Firm 5	Firm 6
Est. Risk	0.0642***	0.0936***	0.0572***	0.0430***	0.0515***	0.0731***
	(0.0042)	(0.0068)	(0.0064)	(0.0073)	(0.0100)	(0.0028)
Age	-0.8955***	-0.4840***	-1.3915***	-0.1352	-0.3352*	-0.5922***
	(0.0889)	(0.1500)	(0.1315)	(0.1611)	(0.1737)	(0.0731)
BM	19.1538***	24.5213***	26.0305***	20.6462***	30.7392***	21.6734***
	(0.8671)	(1.1160)	(1.1907)	(1.9027)	(1.5478)	(0.4715)
Man	-10.7472***	-7.3841*	-3.9804	-15.4388***	-7.0770	2.9258
	(2.3454)	(3.9867)	(3.6796)	(4.6002)	(4.8897)	(1.8135)
1 Acc.	113.0035***	129.5094***	94.8281***	66.1749***	102.6605***	99.7637***
	(4.0816)	(7.1153)	(5.8039)	(6.3675)	(8.8333)	(2.9952)
Big city	34.3785***	46.6072***	56.8048***	55.6861***	39.8474***	16.7550***
	(2.5445)	(4.3043)	(3.8697)	(4.4581)	(5.0768)	(1.9533)
Constant	632.2263***	619.6198***	571.1734***	514.9453***	607.4646***	496.8311***
	(16.4609)	(35.4253)	(26.3254)	(60.2578)	(85.1550)	(18.5252)
Observations	20,800	8,070	8,071	5,504	3,945	15,971
R-squared	0.4057	0.4077	0.4814	0.4690	0.4814	0.6117
VARIABLES	Firm 7	Firm 8	Firm 9	Firm 10	Firm 11	All Firms
Est. Risk	0.0767***	0.0459***	0.0784***	0.0387***	0.0823***	0.0731***
	(0.0111)	(0.0090)	(0.0173)	(0.0118)	(0.0023)	(0.0015)
Age	-1.1958***	-1.5677***	-0.9027***	-0.9363***	-0.3337***	-0.5777***
	(0.2298)	(0.2083)	(0.3089)	(0.2083)	(0.0464)	(0.0329)
BM	23.4924***	12.8527***	31.4627***	11.6583***	22.2223***	21.3837***
	(1.5431)	(2.1425)	(2.3038)	(2.0099)	(0.4208)	(0.2974)
Man	6.1553	-1.9849	-9.5761	-9.4334	-2.0086	-6.0691***
	(6.5007)	(5.6374)	(8.4552)	(6.1263)	(1.2572)	(0.8847)
1 Acc	113.1750***	110.7144***	131.3323***	173.2216***	106.8438***	109.3707***
	(10.0975)	(9.2865)	(16.4076)	(12.6389)	(1.9965)	(1.4293)
Big city	48.1035***	56.8488***	85.4458***	53.4525***	12.9113***	28.9435***
	(7.2178)	(6.2390)	(10.1483)	(6.7341)	(1.4208)	(0.9704)
Constant	494.5030***	697.9200***	167.8205**	692.1357***	443.5997***	511.0770***
	(33.9842)	(26.2245)	(75.6721)	(60.4907)	(8.4378)	(7.0944)
Observations	2,525	4,289	1,355	2,992	50,906	124,428
R-squared	0.4640	0.4037	0.5210	0.4489	0.5403	0.4957

Note: All regressions reported in this table control for vehicle characteristics, time fixed effects, and optional contract clauses.

Table S4: Poisson Regression of Claim Count on Premium and Observable Characteristics

	Claim Count	Claim Count	Claim Count	Claim Count	Claim Count	Claim Count
VARIABLES	Firm 1	Firm 2	Firm 3	Firm 4	Firm 5	Firm 6
Premium	0.8936***	1.1735***	0.4965**	1.6590***	0.7366*	0.5847***
	(0.1401)	(0.2152)	(0.2375)	(0.2927)	(0.3768)	(0.2105)
Age	-0.0047**	-0.0027	-0.0018	-0.0104**	-0.0105**	-0.0051**
	(0.0019)	(0.0031)	(0.0028)	(0.0042)	(0.0043)	(0.0022)
BM	-0.0093	-0.0102	0.0243	-0.0330	-0.0005	-0.0040
	(0.0108)	(0.0180)	(0.0164)	(0.0226)	(0.0287)	(0.0112)
Man	-0.0340	-0.1903**	-0.0257	-0.1226	0.2131*	-0.1083*
	(0.0529)	(0.0892)	(0.0816)	(0.1134)	(0.1247)	(0.0574)
1 Acc.	0.2601***	0.2624**	0.1289	0.1168	0.3383*	0.2681***
	(0.0821)	(0.1260)	(0.1208)	(0.1586)	(0.1819)	(0.0805)
Big city	0.0671	0.0450	0.2042**	-0.0558	0.2835**	0.1360**
	(0.0594)	(0.0960)	(0.0924)	(0.1147)	(0.1328)	(0.0640)
Constant	-2.1891***	-3.3079***	-1.9759***	-2.7247***	-2.9402***	-1.8888***
	(0.2750)	(1.1186)	(0.5409)	(0.4606)	(0.5679)	(0.6591)
Observations	20,800	8,064	8,071	5,499	3,941	15,971
VARIABLES	Firm 7	Firm 8	Firm 9	Firm 10	Firm 11	All Firms
Premium	1.1908***	0.8987***	1.8645***	1.0451***	0.8972***	0.8753***
	(0.4378)	(0.2864)	(0.5203)	(0.3929)	(0.1042)	(0.0611)
Age	-0.0015	-0.0000	0.0122*	0.0054	-0.0056***	-0.0046***
	(0.0049)	(0.0038)	(0.0066)	(0.0048)	(0.0013)	(0.0008)
BM	0.0210	-0.0502**	-0.0507	-0.0276	0.0036	-0.0024
	(0.0285)	(0.0252)	(0.0401)	(0.0287)	(0.0066)	(0.0042)
Man	-0.1769	-0.1955*	0.0228	-0.0971	-0.0825**	-0.0784***
	(0.1448)	(0.1046)	(0.1888)	(0.1423)	(0.0345)	(0.0216)
1 Acc.	-0.2171	0.1641	-0.0928	0.0366	0.1480***	0.1861***
	(0.2198)	(0.1702)	(0.3303)	(0.2549)	(0.0494)	(0.0315)
Big city	-0.0218	0.0968	-0.3329	0.0082	0.1000**	0.0977***
	(0.1691)	(0.1217)	(0.2055)	(0.1605)	(0.0403)	(0.0243)
Constant	-2.6005***	-2.1402***	-2.9924***	-2.8259**	-2.5547***	-2.7121***
	(0.6043)	(0.4632)	(0.5778)	(1.1680)	(0.2066)	(0.1296)
Observations	2,525	4,289	1,325	2,977	50,906	124,428

Note: All regressions reported in this table control for vehicle characteristics, time fixed effects, and optional contract clauses. Premiums are represented in thousands of euros.

S2 Robustness Check: Limited Product Consideration

Our demand model assumes that consumers consider all available insurance products. In practice, however, consumers may face limited consideration sets due to search frictions or other cognitive costs. As a robustness check of our main results, we consider an alternative scenario in which consumers have limited consideration sets. Specifically, we focus on a subsample of consumers who purchase insurance products from the top four firms in the market (Firms 1, 2, 3, and 6), which together account for approximately 42% of total market share. We then re-estimate the demand parameters and the offered price distributions for these firms.

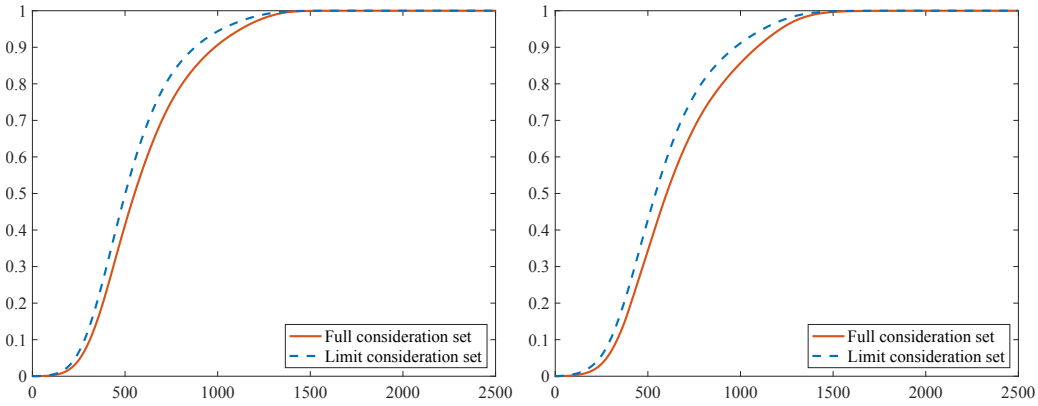
In Table S5, we compare the demand estimates—including price sensitivity parameters and preferences for unobserved product quality—under full and limited consideration sets. In the estimation, we allow price sensitivity to vary with consumer demographics, such as age and whether the consumer lives in a major city. We also estimate preferences for unobserved product attributes separately for eight demographic groups. The table shows that the demand parameters are broadly similar across the two specifications. However, under the limited consideration set, consumers appear to be slightly less sensitive to price increases.

Another key output from our demand estimation is the distribution of offered prices. Figure S1 plots the CDF of offered prices for the top four firms under both full and limited consideration assumptions. The distributions again appear quite similar across the two specifications. Taken together, these results suggest that while consumers may not consider all available insurance products in practice, the impact of this limitation on our estimation results is modest.

Table S5: Comparing demand estimates under full or limited consideration set

(A) Full consideration set								
γ_0	2.11							
Age	-1.21							
Big city	0.45							
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8
ξ_2	-0.54	-0.52	-0.88	-0.76	-0.95	-0.90	-1.41	-1.36
ξ_3	-0.38	-0.62	-0.51	-0.58	-1.49	-1.60	-1.58	-1.64
ξ_6	-0.31	-0.21	-0.36	-0.43	-0.70	-0.56	-0.51	-0.57
(B) Limited consideration set								
γ_0	1.59							
Age	-2.69							
Big city	0.29							
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8
ξ_2	-0.51	-0.61	-0.87	-0.82	-0.92	-0.98	-1.43	-1.49
ξ_3	-0.33	-0.69	-0.47	-0.68	-1.51	-1.74	-1.63	-1.75
ξ_6	-0.19	-0.17	-0.23	-0.29	-0.54	-0.51	-0.40	-0.41
High risk	N	Y	N	Y	N	Y	N	Y
Big city	N	N	Y	Y	N	N	Y	Y
Later periods	N	N	N	N	Y	Y	Y	Y

Note: Panel (A) presents the demand estimation results under the assumption that consumers consider products from all firms. These results are part of the main estimates reported in Table 2, with the exception that we omit the estimates of unobserved product heterogeneity for the remaining firms. Panel (B) shows the results under the assumption that consumers consider only products from the top four insurers. In the estimation, premiums are represented in thousands of euros. The unobserved product heterogeneity for Firm 1 is normalized to zero across all groups. Demographic characteristics for each group are summarized in the bottom panel.



(a) Firm 1 (b) Firm 2

(c) Firm 3 (d) Firm 6

Figure S1: CDF of offered prices for the top four firms. The red solid lines correspond to the case where consumers consider products from all firms, while the blue dashed lines correspond to the case where consumers consider only the top four firms. The CDFs are averaged over consumer characteristics and risk levels.

S3 Counterfactuals: Value of Information

Another interesting counterfactual experiment is to evaluate market outcomes when one firm's information technology is improved, thereby assessing the value of information. A key strength of our model lies in its ability to evaluate the *equilibrium effects* when certain (or all) firms improve their information technology. This equilibrium channel operates in addition to the direct effect, where better information allows a firm to improve its risk assessment and pricing. To disentangle these two channels, we consider two additional counterfactual exercises in this section:

- In the first exercise, we improve Firm 1's information precision to match the best in the market, while holding other firms' pricing strategies fixed. This off-equilibrium scenario isolates the direct effect of better information on Firm 1's pricing and performance.
- In the second exercise, we again improve Firm 1's information precision, but now allow all other firms to adjust their pricing strategies in response. This setup captures both the direct effect and the general equilibrium effect through firm interactions in the market.

Table S6: Counterfactual Results: Off-Equilibrium vs. Equilibrium Outcomes Following an Improvement in Firm 1's Information Precision

	Baseline	Off Equilibrium	Equilibrium Response
Average CS (€)	-542.83	-535.07	-539.14
		(+1.43%)	(+0.68%)
Average CS: Low risk (€)	-477.24	-466.09	-468.54
		(+2.34%)	(+1.82%)
Average CS: High risk (€)	-608.42	-604.04	-609.74
		(+0.72%)	(-0.22%)
Average premium (€)	461.25	454.39	461.80
		(-1.49%)	(+0.12%)
Average profit (€)	849.58	842.16	850.59
		(-0.87%)	(+0.12%)
HHI	2297.26	2286.63	2306.55
		(-0.46%)	(+0.40%)
Average cost (€)	882.39	882.71	881.08
		(+0.04%)	(-0.15%)

Table S6 summarizes the off-equilibrium and equilibrium market outcomes following an improvement in Firm 1's information precision. The results show that enhancing one

firm's information technology leads to only modest changes in overall market outcomes. To better understand firm-level implications, Table S7 reports the percentage change in total profit for each firm under the two counterfactual scenarios.

We highlight several key findings from Table S7. First, improving information precision for a single firm significantly increases that firm's profit, underscoring the value of better information. At the same time, nearly all competing firms experience profit losses. However, when competing firms adjust their pricing in response, their profit losses are mitigated, illustrating the role of strategic responses. Interestingly, Firm 1's profit increases even more under the equilibrium scenario. This may occur because rivals, anticipating Firm 1's improved targeting of low-risk consumers, shift their pricing strategies away from these segments, thereby reducing direct competition.

Table S7: Percentage Changes in Profit: Off-Equilibrium vs. Equilibrium Outcomes Following an Improvement in Firm 1's Information Precision

Firm ID	Off Equilibrium	Equilibrium Response
1	5.85	7.08
2	-4.16	-2.92
3	-2.60	-2.01
4	-1.67	-0.52
5	-0.93	-0.47
6	-1.86	-0.59
7	-5.90	-4.55
8	-3.46	-4.67
9	3.27	5.39
10	-4.21	-3.01
11	-1.61	-0.59