# Competing under Information Heterogeneity: Evidence from Auto Insurance

Marco Cosconati  Yi Xin Fan  Yizhou Jin  Wu

October 18, 2025

# This Paper

- We develop and estimate a novel model of imperfect competition in selection markets:
  - Firms have **heterogeneous information** about consumers.
  - Differ in cost structure.
  - Offer differentiated products.
- We apply our method to study the Italian auto insurance industry:
  - **Substantial differences** in the precision of risk rating and cost structures across firms.
  - Insurers with more accurate risk rating algorithms **cream-skim** low-risk consumers, but they tend to have less efficient cost structures.
- Equalizing information access through a centralized bureau:
  - Significantly lowers market prices by increasing competition.
  - Boosts consumer surplus by 15.7%, nearly reaching the efficiency benchmark.
  - Reduces costs by 12 euros per contract through more efficient insurer-insuree matching.

# Related Literature

- The first tractable empirical framework for analyzing imperfect competition when firms have heterogeneous information about consumers.
  - Recent empirical studies assume information is symmetrically distributed.
    (Cabral et al., 2018; Crawford et al., 2018; Nelson, 2018; Decarolis et al., 2020; Jaffe and Shepard, 2020; Curto et al., 2021; Cuesta and Sepúlveda, 2021; Tebaldi, 2024)

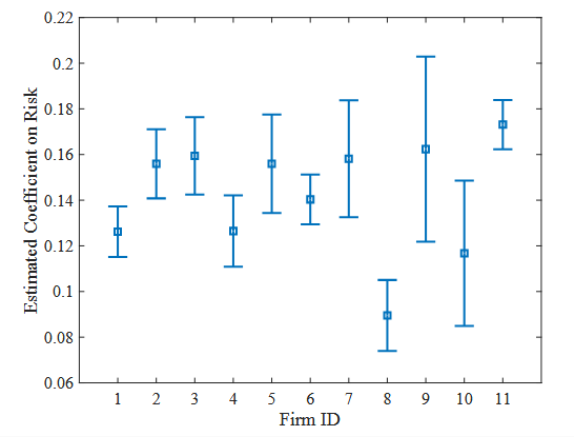
- We extend the classic demand estimation method (Berry, 1994; Berry et al., 1995) to address the common challenge of missing full price menus.
  - Prevalent in many empirical analyses
    **(Goldberg, 1996; Cicala, 2015; Crawford et al., 2018; Allen et al., 2019; D'Haultfoeuille et al., 2019; Salz, 2022; Sagl, 2023)**
  - **Wu and Xin (2024) provide more theoretical results and technical details.**

- On the policy side:
  - 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)**
  - **Financial market regulation**
    **(Einav et al., 2013; Chatterjee et al., 2023; Blattner et al., 2022; Nelson, 2025; Blattner and Nelson, 2021; Liberman et al., 2018; Hertzberg et al., 2011)**

# Data

- A representative sample of matched insurer-insuree panel in Rome (2013-2021).
  - Customers with tenure=0, top 10 firms + fringe firms (Firm 11).
- We observe: **Summary Statistics**
  - Risk factors: age, bonus-malus, residence, vehicle features, driving records, etc.
  - Premium, coverage, and contract clauses.
  - Frequency and severity of claims at each contractual year.
  - **Substantial variation in premiums and claim payouts across firms. Details**
- We don't observe:
  - Full price menu (only transaction prices).
  - Full set of pricing variables (factors in the data explain 50% of price variation). **Details**
  - Insurers incorporate **different** variables into their actuarial models. **Survey**

# Do Insurers Differ in Risk Assessment Precision?




- Regressing premiums on **ex-post** realized claims. **Details**
- Certain firms' premiums are **more responsive to**, or more **accurately reflect**, estimated ex-post risk.

  $\Longrightarrow$ These firms are potentially **more informed** and better at assessing consumer risk.

# Model

- A static model of price competition among $J$ insurers, indexed by $j = 1, 2, \ldots, J$.
- Each firm offers a standardized insurance product.
- Consumer's risk type $\theta$: expected claim payouts.
  - Ex-ante not observed by the firms;
  - Population distribution $f_0(\theta)$ is common knowledge.

# Signal Structure

- For a consumer of type $\theta$, firm $j$ draws a signal $\hat{\theta}_j \sim N(\theta, \sigma_j^2)$, with density $\phi(\hat{\theta}_j; \theta, \sigma_j)$.
- $1/\sigma_j$ measures firm $j$'s information precision.
- Signals are **private** and independent **conditional on** $\theta$.
- Related to common value auctions: signals are noisy estimates of the true but unknown common value, i.e., the cost to insure the consumer.

# Risk Rating

- Firms infer the risk of the consumer upon observing the signal.

$$E(\theta|\hat{\theta}_j, D=j) = \int_\theta \theta f(\theta|\hat{\theta}_j, D=j)d\theta = \frac{\int_\theta \theta \overbrace{\Pr(D=j|\hat{\theta}_j,\theta)}^{\text{Selection Prob.}} \phi(\hat{\theta}_j;\theta,\sigma_j) f_0(\theta) d\theta}{\int_\theta \Pr(D=j|\hat{\theta}_j,\theta)\phi(\hat{\theta}_j;\theta,\sigma_j) f_0(\theta) d\theta}$$

- $E(\theta|\hat{\theta}_j, D=j)$ is an **equilibrium** object.
- $f$ is posterior dist of $\theta$ conditional on the signal and the consumer being **selected** into the firm.
- Selection probability depends on **all firms'** pricing strategies through consumers' demand.

# Pricing Strategy

- Firm $j$ sets the price based on the risk evaluation:

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

- $\alpha_j$ and $\beta_j$ are pricing coefficients **optimally** chosen by the firm.
- $\alpha_j$ relates to the baseline markup; $\beta_j$ relates to the elasticity of price wrt risk.

# Demand

- Firms offer homogeneous insurance plans, but may have **unobserved** (by the econometrician) product attributes $\xi_j$, such as service quality or brand loyalty.
- The utility derived by a consumer $i$ for a product from firm $j$:

$$U_{ij} = -\gamma_i p_j(\hat{\theta}_j) + \xi_j + \varepsilon_{ij}.$$

  $\varepsilon_{ij}$'s are iid utility shocks from type I extreme value dist.

- The probability that a consumer chooses firm $j$ given $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_J)$:

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

- Could also allow demand parameters $(\gamma, \xi)$ to vary with $\theta$.

# Profit Maximization

- Firms simultaneously choose pricing coefficients $(\alpha_j, \beta_j)$ to maximize expected profits, given common knowledge of all firms' primitives (e.g., signal distributions, costs) and $f_0(\theta)$.
- Firm $j$'s profit $\pi_j(\alpha, \beta)$

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

- $c_j$: "net benefits" of contracting with a customer irrespective of the risk.
- $k_j$: efficiency at processing claims.

# Estimation

# Estimation

- **Joint Distribution of Premium and Risk Type:**
  - Using a panel of claim records.
- **Demand:**
  - Market shares $\Longrightarrow$ unobserved product attributes.
  - Consumer sorting patterns $\Longrightarrow$ price sensitivity.
- **Supply:**
  - Joint distribution of premium and risk within firm $\Longrightarrow$ pricing coefficients and signal dist.
  - First-order conditions $\Longrightarrow$ cost parameters.

# Demand

- Key challenge: observe only transaction prices, not offered prices.
  - Auction: observe only winning bids, not submitted bids.
  - Roy models: observe only accepted wages, not potential wages.
- The offered and accepted price distributions are linked through the demand system.
- We propose a novel fixed-point approach that jointly estimates choice probabilities, offered price distributions, and demand parameters.

# Estimation

- Wu and Xin (2024) construct an operator whose fixed point is the offered price distributions and show that it is a functional contraction.
- $\xi$: matched to aggregate market shares; $\gamma$: identified from sorting patterns.
- Possible to allow preference to vary with risk (need parametric assumptions).
- Key insight for supply-side estimation: offered price **monotonically increases** with signal (analogy to auction models: bid is a monotone increasing function of valuation).
- Cost parameters are identified from firms' first-order conditions.

# Demand-Side Results

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

- Senior consumers tend to be less price sensitive.
- We allow preferences for unobserved product attributes to vary with risk type, geography, and time.
- $\xi$'s are generally similar across low- and high-risk groups.

# Supply-Side Results

- Huge heterogeneity along **all** dimensions.

Table 3: Estimates of supply-side parameters

| Firm ID | Pricing Coefficients | | Signal Std. Dev. | Net Benefits | Claim Efficiency |
|---|---|---|---|---|---|
| | $\alpha_j$ | $\beta_j$ | $\sigma_j$ | $c_j$ | $k_j$ |
| 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 |

# Supply-Side Results

- Huge heterogeneity along **all** dimensions.

Table 3: Estimates of supply-side parameters

| Firm ID | Pricing Coefficients | | Signal Std. Dev. | Net Benefits | Claim Efficiency |
|---|---|---|---|---|---|
| | $\alpha_j$ | $\beta_j$ | $\sigma_j$ | $c_j$ | $k_j$ |
| 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 |

# Correlations

- Comparative advantages along different dimensions:
  - Firms suffering lower information precision tend to have lower marginal costs.
  - Firms with higher marginal costs tend to be more efficient at processing claims.

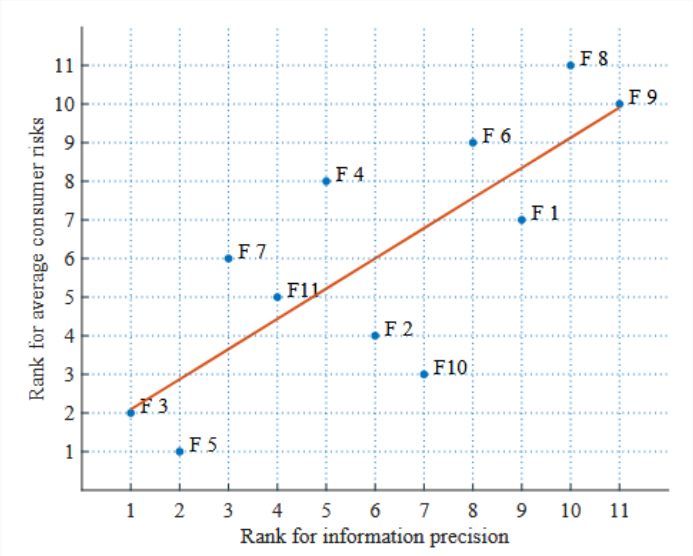
| | $\sigma_j$ | $mc_j$ | $k_j$ |
|---|---|---|---|
| $\sigma_j$ | 1.00 | | |
| $mc_j$ | -0.72 | 1.00 | |
| | (0.01) | | |
| $k_j$ | 0.64 | -0.75 | 1.00 |
| | (0.03) | (0.01) | |

# Correlations

- Comparative advantages along different dimensions:
  - Firms suffering lower information precision tend to have lower marginal costs.
  - Firms with higher marginal costs tend to be more efficient at processing claims.

| | $\sigma_j$ | $mc_j$ | $k_j$ |
|---|---|---|---|
| $\sigma_j$ | 1.00 | | |
| $mc_j$ | -0.72 | 1.00 | |
| | (0.01) | | |
| $k_j$ | 0.64 | -0.75 | 1.00 |
| | (0.03) | (0.01) | |

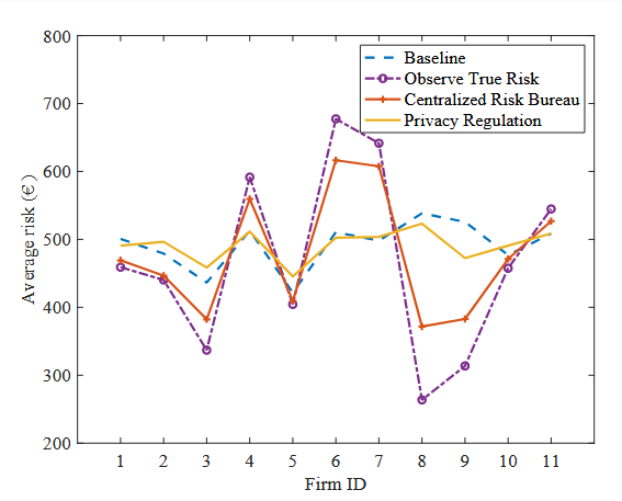
# Who Goes Where?



# Counterfactuals

- **Main policy:** establishing a centralized bureau that collects signals from all firms and aggregates them based on information precision. The posterior estimate of each consumer's risk given **all signals** $E(\theta|\hat{\theta})$ is made public.
- An **efficiency benchmark:** the true risk type of each consumer is observed by all firms $\implies$ information asymmetry is completely eliminated.
- A **privacy benchmark:** firms are required to limit their use of consumer data, $\sigma$ is set to be the largest currently observed in the market $\implies$ reduce overall information availability.

# Sorting Patterns




- Sorting under full information or centralized risk bureau is driven by specialization based on cost advantages.

  $\Longrightarrow$ Improve market efficiency (reduce the avg cost by 3.7%).

- Sorting nearly disappears under privacy regulation.

# Conclusions

- Our paper develops a novel empirical framework for studying competition under information heterogeneity.
- We focus on Italian auto insurance industry: substantial differences in the precision of risk rating across firms; however, firms with lower information precision tend to have lower costs.
- We evaluate the **equilibrium effects** of a public information policy where insurers' risk estimates are aggregated and made public through a centralized bureau:
  - a significant price reduction due to increased competition.
  - boosts consumer surplus by 15.7%, nearly matching the efficiency benchmark.
  - improves the matching efficiency between insurers and insurees.