Competing under Information Heterogeneity: Evidence from Auto Insurance

Marco Cosconati Yi Xin Fan Yizhou Jin Wu
July 8, 2025

IVASS, Bank of Italy

Caltech

Caltech

University of Toronto

Motivation

- Firms increasingly differ in *information precision* (data access/analytics) and in *cost structures*.
- This creates information asymmetries *between* firms (beyond classic buyer–seller asymmetry).
- Policy interest: regulations that equalize or share consumer risk information (e.g., centralized "risk bureau").

Research Questions

- How does heterogeneous information across insurers shape pricing, sorting, and market power?
- What happens to prices, surplus, profits, and sorting if information is shared/standardized?
- Distributional effects: who gains (low vs. high risk)? Efficiency effects: matching and costs?

Contributions

- A tractable model of imperfect competition with firm-specific information precision and costs.
- New identification/estimation strategy using offered-price distributions and demand to recover signals.
- Evidence from Italian auto liability insurance with rich panel linking consumers across insurers.
- Counterfactuals: centralized risk bureau, full information, and privacy/high-variance restrictions.

Institutional Background: Italian Auto Liability (RCA)

- Mandatory, annual, exclusive contracts; insurers cannot reject consumers.
- Large market: ≈31M contracts in 2018; ≈50 national competitors.
- Key contract features widely standardized; little use of deductibles.

Data: IVASS IPER Microdata

- Nationally representative matched insurer–insuree panel with claims frequency/severity, premiums, coverage.
- Tracks policyholders across insurers and time $\Rightarrow$ measure risk using ex-post claims panel.
- Focus sample: new customers in Rome (2013–2021); top 10 firms + fringe group.

Descriptive evidence

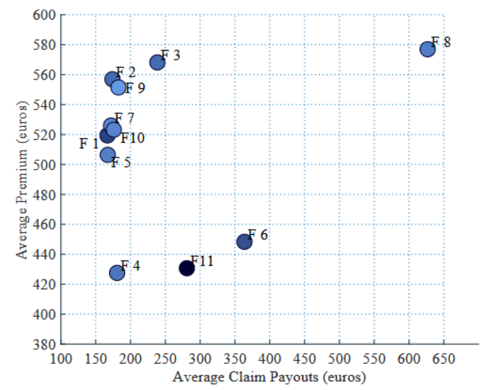
Sample & Summary Statistics

- $N \approx 124{,}428$ contracts; avg premium $\approx$ 478; within-year claim rate $\approx$ 0.08.
- Demographics/vehicle: 56% male; avg age 48; BM class $\approx$ 2; car age $\approx$ 8.3 years.

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

Stylized Facts: Price Variation & Sorting



- Large cross-firm variation in average premiums even at similar average risks/market shares.
- Firms with higher average claim costs attract riskier consumers $\Rightarrow$ sorting across firms.

How we built Figure 2 (Savings distribution)

Goal. Show the distribution of individuals' savings (in UF) for the analysis sample, with extreme values trimmed to improve readability.

Data. Administrative SCOMP records. Savings come from the request file (*Solicitudes*): the AFP-reported balance in UF at the time of request. We restrict to the paper's analysis sample used elsewhere in the slides.

Sample used for the figure.

- Start from the accepted certificates used in the analysis (one row per *certificado de saldo*).
- Keep annuity modality RV inmediata only; drop observations with ELD and with months guaranteed $\neq 0$ to match the core sample used in the paper.
- Merge in savings from *Solicitudes* (balance in UF). If multiple requests exist for a certificate, keep the request closest (in time) to acceptance.

Cleaning and construction.

- Remove missing/implausible balances (nonpositive UF).
- Compute the 99th percentile of savings and *truncate* the right tail at that value (values above p99 are set

Measuring Information Precision: Methodology

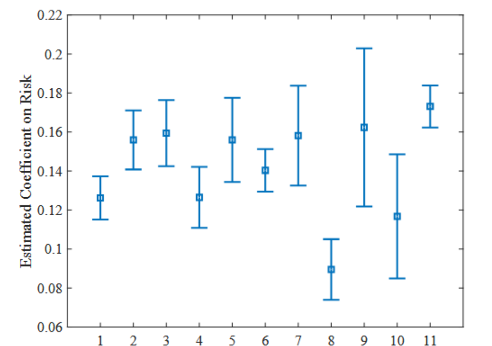
How was Figure 2 created?

1. **Step 1: Construct individual risk measures**
   - Panel regression of claim counts with individual fixed effects
   - Log-normal regression of claim severity conditional on accident
   - Risk measure = Expected frequency × Expected severity
2. **Step 2: Firm-specific premium-risk regressions**
   - For each firm j: $\text{Premium}_{ij} = \alpha_j + \beta_j \cdot \text{Risk}_i + \varepsilon_{ij}$
   - Higher β_j suggests firm's prices are more responsive to actual risk
3. **Step 3: Bootstrap standard errors**
   - 200 bootstrap replications accounting for generated regressor
   - Produces 95% confidence intervals shown as error bars

Key insight: Variation in β_j across firms indicates heterogeneous risk-rating precision

Heterogeneity in Information Precision

- ADD WHY WE NEED A MODEL
- Measure how strongly each firm's premium responds to realized consumer risk (ex-post panel-based risk).
- Strong cross-firm differences in premium–risk slopes ⇒ heterogeneous precision.



Model & Estimation

Conceptual Framework (Overview)

- J insurers; standardized product; no outside option.
- Consumer true risk θ (expected cost/year) unobserved ex ante.
- Firm j observes a private signal $\hat{\theta}_j$ with precision that differs across firms.

Signal Structure

$$\hat{\theta}_j \sim \mathcal{N}(\theta,\ \sigma_j^2), \quad \text{independent across } j \mid \theta, \tag{1}$$

- Lower $\sigma_j^2 \Rightarrow$ higher information precision for firm j.
- Signals are used to form posterior beliefs about θ *conditional on selection*.

Risk Rating & Pricing

$$p_j(\hat{\theta}_j) \;=\; \alpha_j \;+\; \beta_j\, \mathbb{E}[\theta \mid \hat{\theta}_j,\, D = j], \tag{2}$$

- α_j: baseline markup; β_j: pass-through/sensitivity to risk rating.
- $\mathbb{E}[\theta \mid \hat{\theta}_j, D = j]$ embeds selection $\Rightarrow$ nonlinearity in $\hat{\theta}_j$.

Demand

- Consumers choose one insurer; utility depends on price and observable characteristics.
- No outside option (mandatory purchase) $\Rightarrow$ shares across J firms sum to 1.
- Preference parameters allowed to vary with observables and risk type.

Identification: Intuition

- Offered price is (strictly) increasing in the firm's private signal (auction-style monotonicity).
- Use observed *transaction* prices + demand model to invert to *offered-price* distributions by firm.
- How average prices move with risk identifies (α_j, β_j); residual dispersion $\Rightarrow \sigma_j^2$.

Estimation Steps

1. Estimate demand and map transaction prices/shares to offered-price CDFs (firm-specific, nonparametric).
2. Recover pricing coefficients (α_j, β_j) from price–risk relationships.
3. Use price dispersion to identify signal variance σ_j^2 (information precision).
4. Back out firm cost parameters from first-order conditions.

Data Construction of Risk (Two-Part Model)

- Panel regressions to estimate individual risk:
    - Frequency (accident counts with FE) and severity (conditional paid amount).
    - Multiply predicted frequency $\times$ predicted severity $\Rightarrow$ expected cost per year.
- Controls for contract features (coverage, restrictions, devices) mitigate moral-hazard confounds.

Results: Firm Heterogeneity

- Large differences across firms in information precision (σ_j^2), pricing sensitivity (β_j), and costs.
- Firms with less accurate risk-rating tend to have *more efficient* claim-processing costs.
- Baseline sorting: higher-risk consumers concentrate at firms with higher average claim payouts.

Results: Price Sensitivity & Markups

- Estimated β_j varies markedly: some firms' prices are much more responsive to risk.
- Baseline markups (α_j) differ, consistent with market power from information advantages.

[Insert plot: distribution of β_j and α_j across firms.]

Counterfactuals: Information Policies

- **Centralized Risk Bureau**: aggregate firms' signals (weighted by precision), share equally with all.
- **Full Information Benchmark**: firms observe true θ (eliminate information asymmetry).
- **Privacy/Restriction**: firms can only use basic information; set σ_j^2 to the worst observed.

Counterfactual Results: Prices & Welfare

- Average premiums fall by ∼21.6% (bureau) to ∼25.7% (full information).
- Consumer surplus rises by ∼15.7% (bureau), close to ∼16.9% (full information).
- Firm profits decline on average; losses largest for firms with advanced risk-rating tech.

[Insert bar chart: Δ premium, Δ CS, Δ profit under each scenario.]

Distributional Effects by Risk Type

- Bureau/full-info mainly benefit *low-risk* consumers via sharper risk-based pricing.
- Privacy/high-variance benefits *high-risk* consumers (harder to distinguish from low-risk).

[Insert plot: CS changes by risk decile under each scenario.]

Mechanism: Competition & Undercutting

- Equalizing information weakens incumbents' info-based market power.
- Common risk evaluation $\Rightarrow$ more effective undercutting $\Rightarrow$ stronger price competition.

Sorting & Efficiency

- With equal access to risk, firms more efficient at processing claims re-target higher-risk consumers.
- Sorting shifts from info advantages to cost specialization.
- Efficiency gains: avg cost $\downarrow$ by $\sim$3.7% (full info) and by ~ 12 per contract (bureau).

[Insert figure: change in sorting patterns (risk $\times$ firm) vs. baseline.]

Robustness (Selected)

- Alternative risk measures and controls for contract features.
- Bootstrapped uncertainty accounting for generated regressors.
- Poisson checks: premiums predicting claim counts; similar cross-firm heterogeneity.

[Insert table/figure: robustness summaries.]

Policy Implications

- Centralized information can materially lower prices and raise consumer surplus.
- Distributional trade-offs: low-risk consumers gain more under information sharing; high-risk under privacy.
- Industry composition effects: advanced-screening firms lose profits; potential dynamic innovation effects.

Limitations

- Abstract from dynamic pricing/learning and multi-product cross-selling mechanisms.
- Treat signals as reduced-form precision differences (black box of algorithms/data).
- Focus on new customers (tenure $= 0$) to avoid dynamics $\Rightarrow$ external validity caveats.

Paths for Future Work

- Dynamic extensions with learning and switching costs.
- Endogenous investment in information precision and costs (innovation incentives under policy).
- Generalization to other selection markets (credit, health, annuities) under heterogeneous information.

Takeaways

- Information heterogeneity shapes pricing power, sorting, and efficiency.
- Centralized sharing can compress prices and reorient sorting toward cost efficiency.
- Welfare gains are sizable, with clear distributional patterns across risk types.

Appendix: Risk Construction Details

- Frequency model with individual fixed effects; severity model (log amounts).
- Predicted risk $= \widehat{\text{freq}} \times \widehat{\text{severity}}$; controls for contract features.

[Insert table/figure: frequency & severity regression summaries.]

Appendix: Identification Sketch

- Monotonicity of offers in signals $\Rightarrow$ order-preserving mapping to signal quantiles.
- Demand-implied mapping from transactions to offers recovers firm-specific offer CDFs.
- Price–risk slope pins down (α_j, β_j); residual dispersion identifies σ_j^2.

Q&A

Questions?