

Privacy Policies and Consumer Data Extraction: Evidence From U.S. Firms

By Tarun Ramadorai, Antoine Uettwiller, Ansgar Walther

Discussion by Mireia Giné, IESE Business School

Three big questions

- 1. How are data consumer markets? How firms engage in these markets?**
- 2. As firms engage in data markets, how are they affected by cyber risk?**
- 3. How effective are regulatory frameworks on consumer data privacy?**

Research Question & Key Findings

1. Absence of data flows between firms → Indirect way ... **Novel data** & use of AI-driven sophisticated tools.

- 5,000 firms' privacy policies, 2016-2023
- Use of advanced LLMs to extract data points from policies – update relative to literature (from 4 2016 to 6 2023).
- How firms track consumers? Data extraction via cookies using OpenWPM crawler & Wayback Machine.
- Mandatory disclosures in 10K risk factors & Cyberbreach costs.

2. **Four stylized facts** that speaks to basic structure of consumer data.

3. Implications for **modelling market for data & supply of privacy**.

4 Key Findings

- **What can we learn from privacy policies? Heterogeneity & Hedge legal risk**
 1. Substantial within-industry variation rather than boiler plate.
 2. Association btw. policies & tracking activity: privacy policies signal compliance rather than enhance consumer protection.
 3. Extraction activity is linked to cybersecurity risk: more cookies & data points are linked to ex-ante perceived risks (position in 10k) and higher likelihood cyberbreach.
 4. Policy & Extraction is linked to size and technical sophistication (Acc. R&D / Total capital). Large & Intermediate technical sophistication most extractive. More concerned with cyber risk.

Implications

- **What are the Effect of Privacy Regulations on these patterns?**
 - Regulatory frameworks (GDPR, CCPA) influence privacy policies and data practices.
 - Regulations lead to increased **transparency, policy length, and reduced obfuscation in disclosures.**
- **These facts ... Two-tier structure in consumer data markets**
 - Intermediate-sophistication firms (“collect and share”) amass and distribute data to high-sophistication firms (“receive and process”).

In sum, [...] data flows from medium to high technical sophistication firms but financial risks are borne by medium sophistication firms.

#1. Hedging Risk vs Consumer Protection in GDPR context

- **How to differentiate between these two outcomes?**
 - **What would be direct evidence for consumer protection?**
 - Here my lenses from the bank... through compliance the consumer is protected.
 - Is consumer protection a by-product of hedging-risk?
 - If policies are shaped by regulation and they are more transparency, less obfuscation ... wouldn't that protect consumer?

1. Data policies in Banking (tip of ice-berg) & GDPR

- **Chief Protection Officer - 2 Line of Defense**

Oversee all data flows (collection, archive, processing, use) - that is legitimate and users are informed.

GDPR general framework + Spanish Agency of Data Protection specifics on activity.

Fines a very hefty – based on annual revenues. Ex: 6M banks vs low th. for SME.

Does it constrain feature design?

- OK: For digital onboarding : yes selfie & photo of DNI authentication.
- Not allowed so far: biometrics to change certain features ie phone number.
- **Tension between containing fraud & privacy protection (use).**
- **Minimize risk re/ Data Protection w/o legitimization or informed. Trust.**
- My lense from the bank... **through compliance the consumer is protected.**

#2. Data Extraction & Cyber risk

- **Why is Extraction connected to Cyber risk?**
 - **Is Extraction a proxy for exposure (of critical assets)?** Digital services?
 - How valuable / sensitive / critical for operations ... is the extracted data?

If Banks are Tier2...

Is Cyber risk stronger for those that collect than those that process?

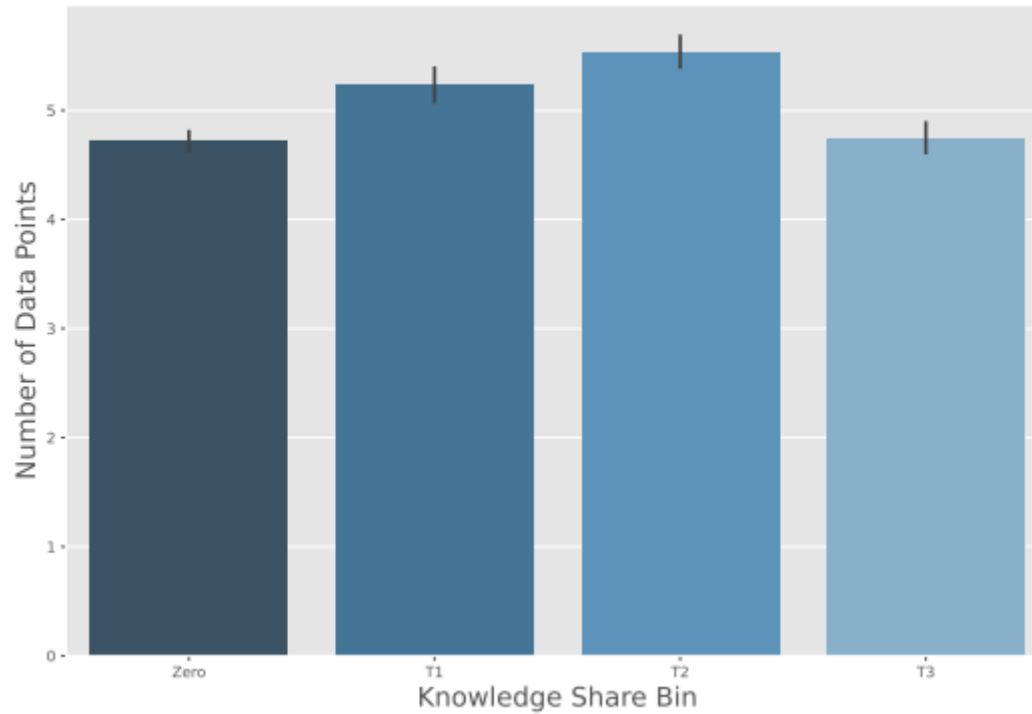
- Protecting from Cyber risk: **Operational Resilience depends on IT investment rather than R&D.**
- Critical services (recovery within 2hr max): payments, access to accounts, digital channels, self-service models (all w/ consumer data).
→ high transactionality, volumes, direct contact with client, automated servicing...
- Vs <5 days: International, Corporate, Mortgages...

#3: The Two-tier data model is very appealing

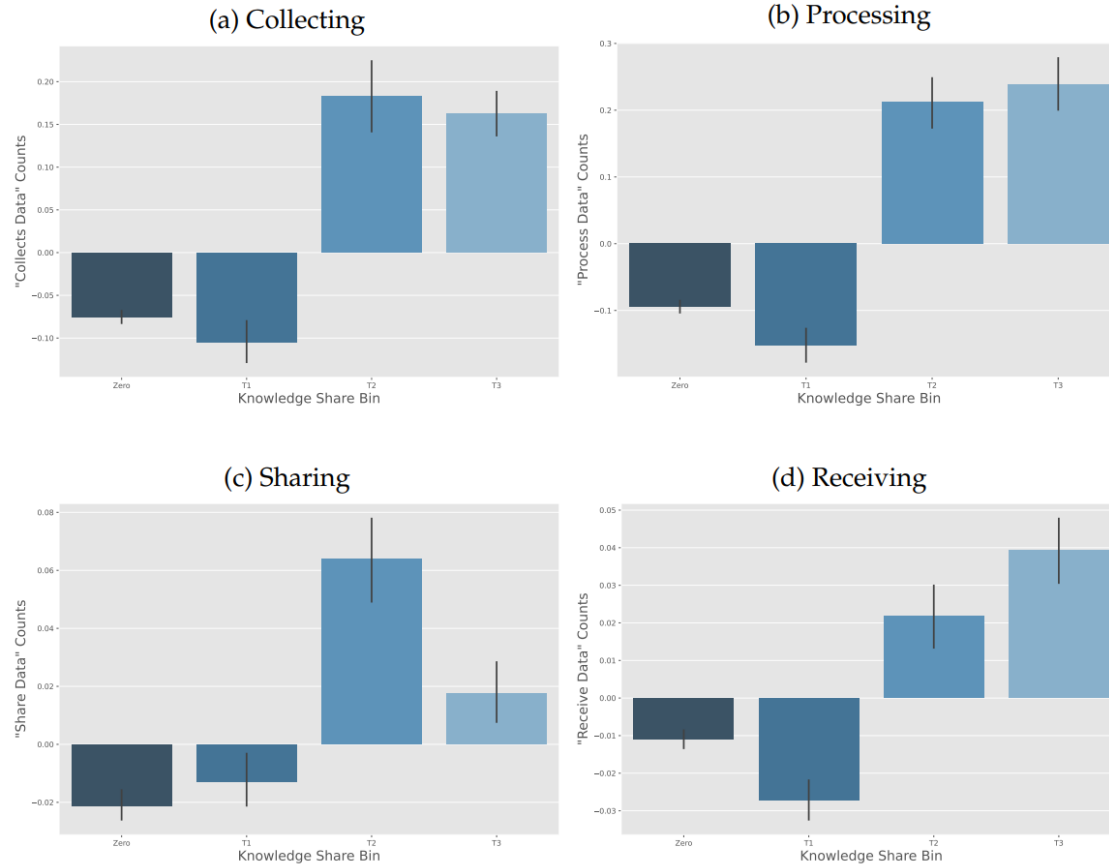
- **What is the evidence for this Two-tier model?**
 - Mostly **collection** Tier1 vs mostly **processing** Tier2 → language in 10k risk section vs data collected in policies.
 - **How much more do they extract?**

Data collected

(b) Data Points



Vs. Text analysis in risk section of 10K.



#3: The Two-tier data model is very appealing

- **What is the evidence for this Two-tier model?**
 - Mostly collection Tier1 vs mostly processing Tier2 → language in 10k risk section vs data collected in policies.
 - How much more do they extract?
- **Strengthen the understanding of each Tier.** Data collector vs. large Platforms in Tier1? What type of firms are Tier2? “less sophisticated privacy policies, Intermediate levels of risk”
- Tier2 with high knowledge share → very R&D focused and less consumer facing. Still, highly knowledge economy.
- Two-tiers w/o mediation of knowledge/share?

#4 How stylized facts relates to market of Information Intermediaries?

- Information intermediaries **collect, process** and distribute information to businesses. Tier1?
- Intermediaries' concentration and high profits (monopoly outcomes).

Data Brokers	Acxiom, Experian, Equifax	Aggregate & sell consumer data
Search Engines	Google, Bing	Provide targeted advertising
Social Media Platforms	Facebook, Instagram, LinkedIn	Offer ad targeting based on user activity
Retail & E-commerce Platforms	Amazon, Alibaba, Walmart	Use data to personalize recommendations
Financial Information Providers	Bloomberg, FICO	Sell credit & financial risk data

#5: The Two-tier consumer data model, larger questions

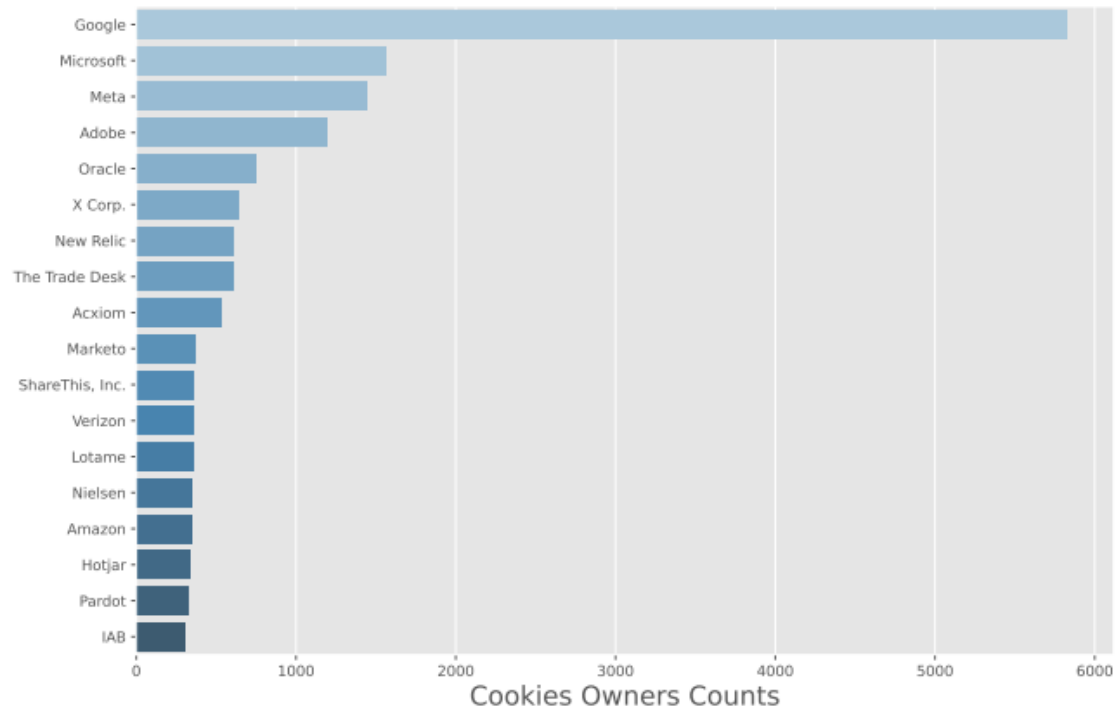
What are some first order implications of this market structure for competition and firm strategies?

- **Do firms have an incentive to integrate vertically with data collectors?**
- **Or will the collectors buy the processors?**

- **Nowadays, What is NOT consumer data? You might be buying digital companies for their client database, future clients. As cost of acquisition is going up.**
 - JPM buying Student loans platform Frank... it went awry but the motivation was to acquire data & new digital channel.

#5. A Concentration story...

(e) Cookies Ownership



Novel data & New insights on how firms interact with consumer data and manage privacy & cyber risk.

1. Hedging legal risk vs. consumer protection.
2. What is behind the link between data extraction and cyber risk?
3. Two tier model further exploration. With other measures of IT sophistication rather than R&D.
4. Where are information intermediaries?
5. Future exploration on scale and market concentration.



Thank you!