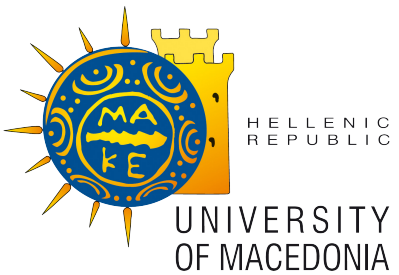
Data Acquisition, Price Discrimination and Endogenous Consumer Privacy (**Preliminary Results**)

Christos Constantatos [1] Apostolos Ioannis Martis [1]

[1]Department of Economics, University of Macedonia

24th Conference on Research on Economic Theory and Econometrics
Ios, July 11, 2026



Motivation – Use of Data

- Digital markets increasingly rely on consumer data for advertising, personalization, and competitive advantage.
  - Federal Trade Commission (2025) reports that data such as location, demographics and even mouse movements can be used by retailers.

Motivation – Privacy I

- Goldfarb and Tucker (2012): Increasing trend in privacy concerns over time, driven primarily by changes in privacy preferences.
- Privacy choices depend on when and how the trade-off is presented to consumers (Egelman et al. (2009); Tsai et al. (2011); Acquisti, John, and Loewenstein (2013)) leading in potential difference in the timing of privacy choice.

Purpose of the paper

Research Question

What is the effect of privacy on firms' pricing schedule, profits and welfare?

- Under which circumstances the timing of consumers' privacy decisions shape firms' pricing incentives, market segmentation, and welfare outcomes?

Closest Work

- Montes, Sand-Zantman, and Valletti (2019): Shared data maximizes consumer surplus; exclusive data acquisition maximizes profits but lowers total welfare.
  - Closest paper to ours. Differences:
    - *Unit demand*: Consumers buy from one firm or do not buy. We use continuous demand; consumers buy from both firms.
    - Because markets are covered, lower prices mainly redistribute surplus; privacy typically reduces welfare by adding a privacy cost without changing total quantity.
    - In our model, lower prices increase individual quantities and total surplus, so privacy may enhance welfare.

Demand Side I

Individual Utility Function:

$$U_i = k_i\,\alpha\,(q_{1i} + q_{2i}) - \frac{b}{2}(q_{1i}^2 + 2\theta\,q_{1i}q_{2i} + q_{2i}^2) + M, \qquad \alpha = b = 1.$$

- $k_i \sim U[\underline{k}, 1]$: Idiosyncratic intensity of willingness-to-pay, with $\underline{k} > 0$.
- $\theta \in [0, 1)$: Degree of product homogeneity low $\theta \Rightarrow$ differentiated products).

Individual demand:

$$D_{ji}(p_{ji}, p_{-ji}, k_i) = \frac{k_i(1-\theta) - p_{ji}(k_i) + \theta\,p_{-ji}(k_i)}{1-\theta^2}, \qquad j = 1, 2.$$

Total demand. Obtained by horizontal summation over the consumer types served at each price.

Old and New Consumers

Old consumers.

- Have purchased online before
- Their data are available, unless they decide to withhold them.
- They can pay personalized prices.

New consumers.

- New to the online market.
- No data about them are available to the firms.

Density. The population of each group is distributed uniformly on the same width $[\underline{k}, 1]$, however they differ in mass.

- We define as $h > 0$ the mass ratio of new to old consumers.

The Introduction of the Pool Market

- **Data concealment.** Old consumers may decide to withhold their personal information
  - They join the new market.
  - They are served at the uniform price
- **Pool Market.** The horizontal summation of new consumers and drop-outs from the personalized market

Introduction of cost asymmetry

- Anonymity can be obtained either through concealing online data or visiting a physical store.
  - In the last case, there is the possibility of cost-asymmetry on one or both markets (personalized/ pool markets).
- Let $(\gamma_1,\ \gamma_2)$ and $(c_1,\ c_2)$ be the firms' marginal costs in the personalized and pool market respectively.

	Personalized-market costs (γ_1, γ_2)	**Privacy-market costs** (c_1, c_2)
Fully Symmetric	$(c,\ c)$	$(c,\ c)$
Symmetric (Pool-Market) wedge	$(0,\ c)$	$(0,\ c)$
Symmetric (Person. Market) wedge	$(c,\ 0)$	$(c,\ 0)$
Pool-market asymmetry	$(0,\ 0)$	$(0,\ c)$
Personalized-market asymmetry	$(0,\ c)$	$(0,\ 0)$

Table: Types of cost asymmetry

Note: The cost is considered $c \geq 0$, for the symmetric cases, but $c > 0$ for the asymmetric ones

Benchmark cases: Full and No privacy I

Full Privacy. In this case, *no consumer data is available*, therefore, *no price discrimination is possible*. Equilibrium prices are based on aggregate demand.

- Equilibrium (uniform) prices:

$$p_1^N = \frac{(2+\theta)(1-\theta)(1+\underline{k})+2c\theta}{2(4-\theta^2)}, \qquad p_2^N = \frac{(1-\theta)(2+\theta)(1+\underline{k})+4c}{2(4-\theta^2)}$$

which, under symmetric, zero marginal cost collapses to:

$$p_j^N = p_N = \frac{(1-\theta)(1+\underline{k})}{2(2-\theta)}$$

Benchmark cases: Full and No privacy II

No Privacy. Under the absence of privacy, *arbitrage is prohibitively costly*, so *third-degree price discrimination is possible* for old consumers.

- Equilibrium (personalized) prices:

$$p_{1i}^D = \frac{(2+\theta)(1-\theta)k_i - c\theta}{4-\theta^2}, \qquad p_{2i}^D = \frac{(2+\theta)(1-\theta)k_i - 2c}{4-\theta^2}$$

which, under symmetric, zero marginal cost, becomes:

$$p_{ji}^D = p_i^D = \frac{1-\theta}{2-\theta}k_i, \; j = 1, \; 2.$$

Timing

- **Consumer privacy.** We consider two different endogenous-privacy models with respect to timing.
  - **Ex-ante privacy – Model A.** Individuals decide their privacy status before receiving price offers from firms.
  - **Ex-post privacy – Model B.** Firms initially make price offers to consumer and then individuals decide their privacy status

Ex-ante Privacy – Model A I

- Consumers can hide their data, by **excluding themselves from the dataset** acquired by the firms.
- To do so, consumers may:
  - Abstain from social media (which are usually big data collectors)
  - Always “disagree" upon giving consent to sharing data when prompted from a site.
  - Do not give access to their microphone and camera and avoid applications that require such access in order to function properly.

Ex-ante Privacy – Model A II

- Consumers have already made their privacy decision.
  - This defines a critical consumer with intensity of WTP, k_c, indifferent about hiding data or not.
    - ▷ Consumers in $[\underline{k},\ k_c]$ disclose their data
    - ▷ Consumers in $[k_c,\ 1]$ conceal their data because they know they will always receive unfavorable offers.
- The firm j's maximization problem in the second stage is:

$$\Pi_j^A = \underbrace{\int_{\underline{k}}^{k_c} p_{ji}\ D_{ji}(p_{ji}, p_{-ji}, k_i) f(k_i)\, dk_i}_{\text{Personalized Market}} + p_j \underbrace{\left(\int_{k_c}^{1} D_{ji}(p_j, p_{-j}, k_i) f(k_i)\, dk_i + \int_{\underline{k}}^{1} D_{ji}(p_j, p_{-j}, k_i)\ f(k_i)\ h\, dk_i \right)}_{\text{Pool Market}}$$

Model A – Symmetric (zero) marginal costs I

- **Equilibrium Prices**:
  1. *Personalized Market*: $p_{ji}^{A}(k_i;\ \theta) = \dfrac{1-\theta}{2-\theta}k_i$
  2. *Pool Market*: $p_A(h;\ \theta) = \dfrac{\left[3+2h-2\sqrt{h(h+1)}\right](1-\theta)}{3(2-\theta)}$
- **Price behavior.** Prices decrease in h. When the latter approaches infinity, the ex-ante privacy prices approaches the full-privacy one, i.e., $\lim_{h\to\infty} p_A(h;\theta) = p_N(\theta) = \dfrac{2(1-\theta)}{3(2-\theta)}$.

Model A – Symmetric (zero) marginal costs II

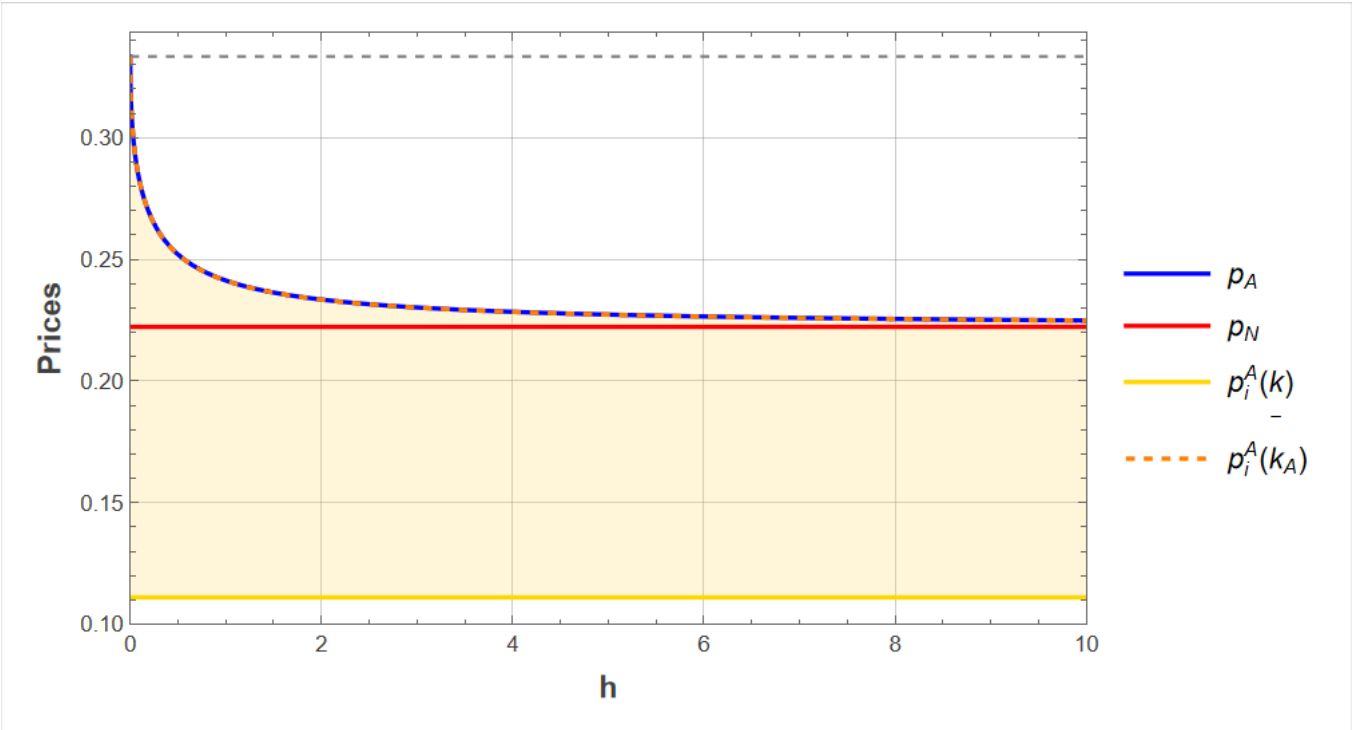


Figure: Prices in ex-ante privacy

Model A – Symmetric (zero) marginal costs III

- Let $V(p_{1i},\ p_{2i},\ k_i)$ be the indirect utility net of expenses, the critical type is determined by:

$$V\left(p_{1i}(k_c),\ p_{2i}(k_c),\ k_c\right) = V\left(p_1,\ p_2,\ k_c\right)$$

- **Critical Consumer.** $k_A = \frac{1}{3}\left(2h + 3 - 2\sqrt{h^2 + h}\right)$
- **Critical-consumer behavior.**
  - Privacy increases in h, i.e., $\frac{\partial k_c}{\partial h} < 0$.
  - Even when h approaches infinity, k_A remains above $\underline{k}$, $\lim_{h\to\infty} k_A(h) = \frac{2}{3}$.

Model A – Symmetric (zero) marginal costs IV

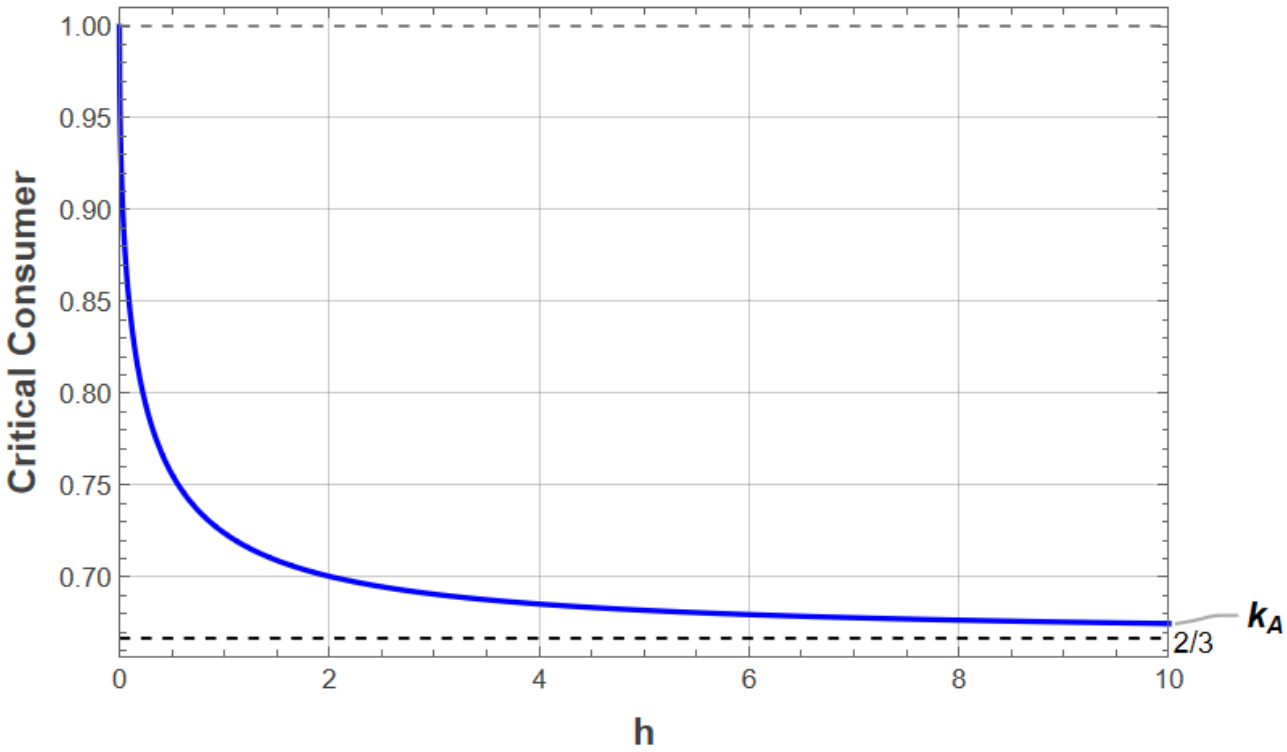


Figure: Critical Consumer in ex-ante privacy

Ex-post privacy – Model B I

- Alternatively, consumers may **hide their data from specific firms and transactions**
  - Using a VPN.
  - Purchasing at a physical store.
- This **requires consumers to know the prices** or **have an idea about them**.
- Therefore, **firms make their price offers and then consumers compare the personalized and pool prices**.

Ex-post privacy – Model B II

- Let, for presentational purposes, $p_2 \geq p_1$.
- **Market segments**. Three consumer segments arise; two critical consumers exist, k_{1c} and k_{2c}, who arise by this condition $p_{ji}(k_{jc}) = p_j,\ j \in \{1, 2\}$:
  1. **Personalized market**: Consumers in $[\underline{k},\ k_{1c}]$ purchase online from both firms.
  2. **Mixed market**: Consumers in $[k_{1c},\ k_{2c}]$ purchase from firm 1's physical store and online from firm 2
  3. **Old-consumer Private Market**: Consumers in $[k_{2c},\ 1]$ purchase from both firms' physical store.
- For tractability, assume $\theta = 1/2$.

Ex-post privacy – Model B III

- Firms' profits in Model B are:

$$\Pi_1^B = \underbrace{\int_{\underline{k}}^{k_{1c}(p_1,\ p_2)} (p_{1i} - \gamma_1)\ D_{1i}(p_{1i}, p_{2i}, k_i) f(k_i)\, dk_i}_{\text{Personalized Market}} + (p_1 - c_1) \underbrace{\int_{k_{1c}(p_1,\ p_2)}^{k_{2c}(p_1,\ p_2)} D_{1i}(p_1, p_{2i}, k_i) f(k_i)\, dk_i}_{\text{Mixed Market}}$$

$$+ (p_1 - c_1) \underbrace{\left(\int_{k_{2c}(p_1,\ p_2)}^{1} D_{1i}(p_1, p_2, k_i) f(k_i)\, dk_i + \int_{\underline{k}}^{1} D_{1i}(p_1, p_2, k_i)\ f(k_i)\ h\, dk_i \right)}_{\text{Pool Market}}$$

$$\Pi_2^B = \underbrace{\int_{\underline{k}}^{k_{1c}(p_1,\ p_2)} (p_{2i} - \gamma_2)\ D_{2i}(p_{1i}, p_{2i}, k_i) f(k_i)\, dk_i}_{\text{Personalized Market}} + \underbrace{\int_{k_{1c}(p_1,\ p_2)}^{k_{2c}(p_1,\ p_2)} (p_{2i} - \gamma_2)\ D_{1i}(p_1, p_{2i}, k_i) f(k_i)\, dk_i}_{\text{Mixed Market}}$$

$$+ (p_2 - c_2) \underbrace{\left(\int_{k_{2c}(p_1,\ p_2)}^{1} D_{2i}(p_1, p_2, k_i) f(k_i)\, dk_i + \int_{\underline{k}}^{1} D_{2i}(p_1, p_2, k_i)\ f(k_i)\ h\, dk_i \right)}_{\text{Pool Market}}$$

When the two models break I

- **Symmetric costs, symmetric results.**
  - Models A and B do not differ when:
    1. Costs are symmetric across markets and firms, i.e., $(\gamma_1,\ \gamma_2) = (c,\ c)$ and $(c_1,\ c_2) = (c,\ c)$,
    2. Costs are symmetrically asymmetric across markets, i.e., $(\gamma_1,\ \gamma_2) = (0,\ c)$ and $(c_1,\ c_2) = (0,\ c)$

When the two models break II

- **Asymmetric costs between markets**
  - **Model A (ex-ante).** Hard to justify cost differences between the personalized and pool markets: both are *online* markets, served through the same channel.
  - **Model B (ex-post).** Between-market cost differences arise naturally: the pool market includes *physical stores*, which consumers can visit instead of concealing data online.
  - **Consequences of the asymmetry:**
    - ▷ Firms either protect their online market or exploit a cost advantage in their physical-store market.
    - ▷ Consumers may hide from one firm while revealing to the other $\Rightarrow$ the **mixed market** arises.

Comparisons: What can be compared?

- **Models A and B are not directly comparable.** Ex-ante privacy cannot support cost asymmetries across markets, so each model is compared with the two polar benchmarks:
  - **D** – No privacy (full discrimination)
  - **N** – Full privacy (uniform pricing)

- In what follows, each table cell reports the **regime preferred** under the corresponding metric and cost level.

Model A vs. benchmarks (symmetric costs)

	Ranking	**Driving force**
Consumers	$CS_N > CS_A > CS_D$	Concealment stops at $k_A \geq 2/3 > \underline{k}$
Firms	$\Pi_D > \Pi_A > \Pi_N$	Personalization extracts surplus
Welfare	$W_N > W_A > W_D$	Consumer gains dominate profit losses

Takeaway

Ex-ante privacy lies strictly *between* the polar regimes: it shields consumers – but only partially (*suboptimal privacy*) – and raises welfare because lower prices expand individual quantities.

Model B vs. benchmarks: Pool-market asymmetry

vs. No privacy (D)	CS	Π	W
$c = 0$	**B**	D	**B**
Low c	D	**B**	D
Intermediate c	D	**B**	**B** (low h) / D (high h)
High c	D	**B**	**B**

- vs. Full privacy (N): $\Pi_B > \Pi_N$ and $CS_B < CS_N$, $\forall\, c, h > 0$.
- **Channel:** segmentation relaxes pool-market competition and supports higher uniform prices $\Rightarrow$ the gains from ex-post privacy accrue mainly to **firms**.

Model B vs. benchmarks: Personalized-market asymmetry

vs. D	CS	Π	W
$c = 0$	**B**	D	**B**
Low c	**B**	**B** (low h) / D (high h)	**B**
High c	**B**	**B**	**B**

vs. N	CS	Π	W
Low c	N	**B**	N
Intermediate c	**B**	N	**B**
High c	N	**B**	N

- At intermediate levels of c, endogenous (partial) privacy could benefit consumers more than full privacy.
- **Channel:** the personalized cost wedge disciplines discriminatory pricing $\Rightarrow$ the gains from ex-post privacy accrue mainly to **consumers**.

Two asymmetries, two channels

Pool-market asymmetry

Ex-post privacy gains route to **firm profits**: segmentation softens competition in the pool market. Consumers can be *hurt*.

Personalized-market asymmetry

Ex-post privacy gains route to **consumer surplus**: $CS_B > CS_D$ always, and welfare rises throughout ($W_B > W_D$).

- Both single-market cost asymmetries make ex-post privacy attractive relative to no privacy – but through **opposite channels**.

Conclusions I

What we examine.

- What is the effect of privacy to pricing and welfare?
- How the *timing* of consumers' privacy choice (ex-ante vs. ex-post) shapes pricing, segmentation, profits, and welfare in a data-driven market.

Market-specific cost asymmetry matters.

- **Symmetric costs.** A critical consumer $k_c(h)$ splits disclose from conceal; the mixed market vanishes, so *ex-ante and ex-post privacy coincide*. A personalized segment survives even as $h \to \infty$: *suboptimal privacy*.
- **Timing matters only under cross-market cost asymmetry**, where a mixed market arises and the two models diverge.

Conclusions II: When is privacy good – and for whom?

- **Consumers.** Privacy (in either form) generally protects consumers relative to full discrimination: $CS_A > CS_D$, and $CS_B > CS_D$ under personalized-market asymmetry.
  - *Exception:* pool-market asymmetry, where segmentation raises pool prices and $CS_B < CS_D$.
  - Privacy is never complete: concealment always stops short of the full-privacy outcome ($CS < CS_N$ in most configurations).
  - For intermediate costs, ex-post privacy can be even better than full privacy.

- **Firms.** Under symmetric costs, data is unambiguously profitable ($\Pi_D > \Pi_A > \Pi_N$). Under cost asymmetries, however, ex-post privacy can *raise* profits above full discrimination: privacy acts as a **competition softener** via segmentation.

Conclusions III: Welfare and policy implications

- **Welfare.** Because individual quantities respond to prices, privacy-induced price reductions *expand* output rather than merely redistribute surplus:
  - $W_A = W_B > W_D$ under symmetric costs;
  - $W_B > W_D$ *always* under personalized-market asymmetry;
  - non-monotone in (c, h) under pool-market asymmetry.
- **Policy.** Ex-ante instruments (consent regimes, data-collection opt-outs) and ex-post instruments (VPNs, offline purchase options) are **not equivalent**: which one dominates depends on the cost environment (c, h) and on which market bears the cost wedge.

THANK YOU FOR YOUR ATTENTION

Any questions?