Data, Privacy Laws, and Firm Production: Evidence from GDPR

Mert Demirer[1] Diego Jiménez Hernández[2] Dean Li[1] Sida Peng[3]

[1]Massachusetts Institute of Technology
[2]Federal Reserve Bank of Chicago
[3]Microsoft

FTC Micro Conference
November 2nd, 2023

*Opinions expressed herein are solely the authors' and do not reflect the opinions and beliefs of the Federal Reserve Bank of Chicago or the Federal Reserve System.

Growing Importance of Data for Firms & Privacy Laws

- **Data** plays an important and growing role in firm production
- New **privacy regulations** have emerged to govern data collection, storage, and analysis

Growing Importance of Data for Firms & Privacy Laws

- **Data** plays an important and growing role in firm production
- New **privacy regulations** have emerged to govern data collection, storage, and analysis
- **EU's General Data Protection Regulation**: comprehensive and consequential privacy law
  - Affected over **20M firms** across many countries that target EU residents (no min. size threshold)

Growing Importance of Data for Firms & Privacy Laws

- **Data** plays an important and growing role in firm production
- New **privacy regulations** have emerged to govern data collection, storage, and analysis
- **EU's General Data Protection Regulation**: comprehensive and consequential privacy law
  - Affected over **20M firms** across many countries that target EU residents (no min. size threshold)
- Firms need to take costly measures to comply with GDPR
  - Data security, customer delete requests, record-keeping, large penalties for breaches, etc.
  - Important **compliance costs** ∼€1.7M for SME; >€70M for large organizations[1] (fixed and variable)

[1] Accenture, IAPP-EY Annual Privacy Governance Report

Production Approach: GDPR as Increased Cost of Managing Data

- These regulations increase the **cost of data** to firms and **affect their input choices**
  - Generate a wedge between the marginal product of data and its price (Hsieh and Klenow, 2009)
  - Affect firms' data and computation choices

Production Approach: GDPR as Increased Cost of Managing Data

- These regulations increase the **cost of data** to firms and **affect their input choices**
  - Generate a wedge between the marginal product of data and its price (Hsieh and Klenow, 2009)
  - Affect firms' data and computation choices

- Literature focused on firm outcomes, little evidence on firms' margins of adjustment / choices
- Requires a framework to analyze how firms use and process data

Production Approach: GDPR as Increased Cost of Managing Data

- These regulations increase the **cost of data** to firms and **affect their input choices**
  - Generate a wedge between the marginal product of data and its price (Hsieh and Klenow, 2009)
  - Affect firms' data and computation choices

- Literature focused on firm outcomes, little evidence on firms' margins of adjustment / choices
- Requires a framework to analyze how firms use and process data

This paper:

1. How do firms combine data and computation in production?
2. What is the cost of the GDPR for firms, and how do they adjust their data/computation inputs?

Data and Methods

- Confidential data from one of the largest **cloud computing providers**, 2016-2021
  - Monthly data on data storage and computation for 100,000+ firms worldwide
  - Spans many industries (software, manufacturing, retail, finance)

Data and Methods

- Confidential data from one of the largest **cloud computing providers**, 2016-2021
  - Monthly data on data storage and computation for 100,000+ firms worldwide
  - Spans many industries (software, manufacturing, retail, finance)

- **Event study:** Compare data and computation of EU firms (treated) relative to US (control)
  - Data directly targeted by regulation; computation affected through firm re-optimization/substitution

Data and Methods

- Confidential data from one of the largest **cloud computing providers**, 2016-2021
  - Monthly data on data storage and computation for 100,000+ firms worldwide
  - Spans many industries (software, manufacturing, retail, finance)

- **Event study:** Compare data and computation of EU firms (treated) relative to US (control)
  - Data directly targeted by regulation; computation affected through firm re-optimization/substitution

- **Production function:** CES tech. to combine data and computation in information production
  - GDPR is wedge between data's cost & marginal product

Preview of Results

1. **GPDR significantly changes data-compute input mix:** firms become less “data intensive”
   - Stored data ↓ 26%; computation ↓ 15% (both in EU relative to US)

Preview of Results

1. **GPDR significantly changes data-compute input mix:** firms become less “data intensive”
   - Stored data ↓ 26%; computation ↓ 15% (both in EU relative to US)

2. **Data and computation are strong complements:** elasticity of substitution is 0.3-0.4
   - Data storage and computation are “stronger complements” than traditional inputs

Preview of Results

1. **GPDR significantly changes data-compute input mix:** firms become less “data intensive”
   - Stored data ↓ 26%; computation ↓ 15% (both in EU relative to US)

2. **Data and computation are strong complements:** elasticity of substitution is 0.3-0.4
   - Data storage and computation are “stronger complements” than traditional inputs

3. **Average GDPR wedge is equivalent to 20% tax** that increases the cost of storing data
   - Wedges are nearly twice as large for smaller firms (25% vs 13%)
   - Producing information became ∼4% more costly after the GDPR

Preview of Results

1. **GPDR significantly changes data-compute input mix:** firms become less "data intensive"
   - Stored data ↓ 26%; computation ↓ 15% (both in EU relative to US)

2. **Data and computation are strong complements:** elasticity of substitution is 0.3-0.4
   - Data storage and computation are "stronger complements" than traditional inputs

3. **Average GDPR wedge is equivalent to 20% tax** that increases the cost of storing data
   - Wedges are nearly twice as large for smaller firms (25% vs 13%)
   - Producing information became ∼4% more costly after the GDPR

- **Caveat:** This is **not** a full analysis of the welfare effects of the GDPR or privacy laws

Contribution to the Literature

1. **The impact of the GDPR on firms**

   online tracking (Aridor et al., 2022; Lefrere et al., 2022; Lukic et al., 2023); business ventures (Jia et al., 2021); app development (Kircher and Foerderer, 2020; Janßen et al., 2021; Kircher and Foerderer, 2023); third-party ads (Johnson et al., 2022; Peukert et al., 2022); e-commerce revenue (Goldberg et al., 2023); effectiveness of targeted ads (Aridor et al., 2022; Matos and Adjerid, 2022); profits, and sales (Koski and Valmari, 2020; Chen et al., 2022); internet interconnectivity (Zhuo et al., 2021); + many others

   → Study the key margin targeted by privacy laws: data

   → Study firms' choices rather than outcomes using a production approach

Contribution to the Literature

1. **The impact of the GDPR on firms**

2. **Data as an input to the production of goods and services**

   (e.g., Jones and Tonetti, 2020; Cong et al., 2021; Farboodi and Veldkamp, 2022)

→ Empirical analysis of how firms use data and computation in a production approach

→ The first paper to incorporate and estimate both data and computation in firm production

Contribution to the Literature

1. **The impact of the GDPR on firms**

2. **Data as an input to the production of goods and services**

3. **Economics of privacy**

(Goldfarb and Tucker, 2011; Goldfarb and Tucker, 2012; Acquisti et al., 2016; Athey et al., 2017; Choi et al., 2019; Montes et al., 2019; Ichihashi, 2020; Loertscher and Marx, 2020; Chen et al., 2021; Krähmer and Strausz, 2023, + many others)

→ Evaluate the effects of the largest privacy regulation on firms

Contribution to the Literature

1. **The impact of the GDPR on firms**

2. **Data as an input to the production of goods and services**

3. **Economics of privacy**

4. **Literature on misallocation**

   (Hsieh and Klenow, 2009; Restuccia and Rogerson, 2017)

→ Model privacy regulation costs as a wedge to study GDPR compliance costs

Presentation Outline

1. Introduction
2. Institutional Setting
3. DiD Estimates of the Impact of GDPR
4. Production Function Framework
5. The Production Cost of GDPR
6. Conclusions

2 Institutional Setting

What is the General Data Protection Regulation (GDPR)?

- **Description:** Passed in April 2016 and went into effect in May 2018
  - Replaced and harmonized Data Protection Directive from 1995
- **Scope:** GDPR applies to firms located in EU or collecting “personal data” from EU residents
  - Protections apply to employee and customer data (e.g., IP addresses, location, shift schedules)
- **Enforcement:** Supervisory authorities in EU states enforce the regulation
  - Upon request, firms must be able to demonstrate their compliance
- **Compliance:** Heterogeneity in cost and timing of compliance
  - Survey evidence that only 10% of firms were compliant at the time of the implementation

GDPR is a “Data Protection” Law (General **Data Protection** Regulation)

- Imposes a set of company obligations to protect data in addition to individual rights

GDPR is a “Data Protection” Law (General **Data Protection** Regulation)

- Imposes a set of company obligations to protect data in addition to individual rights

Firm Responsibilities under GDPR:

1. Operational changes: privacy notices, employee training (Art. 25)
2. Designation / hiring of data protection officers (Art. 37)
3. Handling customer delete/transfer requests expeditiously (Art. 14)
4. Records of processing activities, impact assessment and analysis (Art. 37)
5. Data security: increase security requirements, breach notification (Art. 32)
6. Increased liabilities with penalties up to 4% of *global* revenue (Art. 83)

GDPR is a “Data Protection” Law (General **Data Protection** Regulation)

- Imposes a set of company obligations to protect data in addition to individual rights

Firm Responsibilities under GDPR:	Fixed Costs	Variable Costs
1. Operational changes: privacy notices, employee training (Art. 25)	✓	
2. Designation / hiring of data protection officers (Art. 37)	✓	✓
3. Handling customer delete/transfer requests expeditiously (Art. 14)	✓	✓
4. Records of processing activities, impact assessment and analysis (Art. 37)		✓
5. Data security: increase security requirements, breach notification (Art. 32)		✓
6. Increased liabilities with penalties up to 4% of *global* revenue (Art. 83)		✓

GDPR is a “Data Protection” Law (General **Data Protection** Regulation)

- Imposes a set of company obligations to protect data in addition to individual rights

Firm Responsibilities under GDPR:	Fixed Costs	Variable Costs
1. Operational changes: privacy notices, employee training (Art. 25)	✓	
2. Designation / hiring of data protection officers (Art. 37)	✓	✓
3. Handling customer delete/transfer requests expeditiously (Art. 14		
4. Records of processing activities, impact assessment and analysis (Art. 37)		✓
5. Data security: increase security requirements, breach notification (Art. 32)		✓
6. Increased liabilities with penalties up to 4% of *global* revenue (Art. 83)		✓

18% firms hire multiple DPOs
(Hughes and Saverice-Rohan, 2017)

GDPR is a “Data Protection” Law (General **Data Protection** Regulation)

- Imposes a set of company obligations to protect data in addition to individual rights

Firm Responsibilities under GDPR:	Fixed Costs	Variable Costs
1. Operational changes: privacy notices, employee training (Art. 25)	✓	
2. Designation / hiring of data protection officers (Art. 37)	✓	✓
3. Handling customer delete/transfer requests expeditiously (Art. 14)	✓	✓
4. Records of processing activities, impact assessment and analy **1% fully automate, 64% manually** (DataGrail, 2020)		
5. Data security: increase security requirements, breach notification (Art. 32)		✓
6. Increased liabilities with penalties up to 4% of *global* revenue (Art. 83)		✓

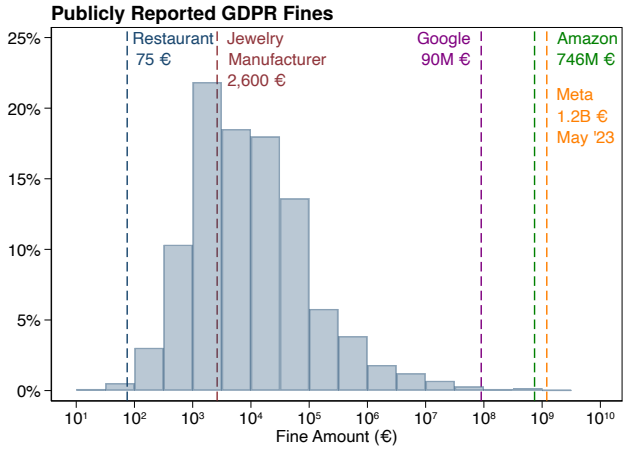
GDPR is a “Data Protection” Law (General **Data Protection** Regulation)

- Imposes a set of company obligations to protect data in addition to individual rights

Firm Responsibilities under GDPR:	Fixed Costs	Variable Costs
1. Operational changes: privacy notices, employee training (Art. 25)	✓	
2. Designation / hiring of data protection officers (Art. 37)	✓	✓
3. Handling customer delete/transfer requests expeditiously (Art. 14)	✓	✓
4. Records of processing activities, impact assessment and analysis (Art. 37)		✓
5. Data security: increase security requirements, breach notification (Art. 32)		✓
6. Increased liabilities with penalties up to 4% of *global* revenue (Art. 83)		✓

31% firms buy liability insurance
(LLP, 2018)

GDPR Affects a Wide Range of Industries and Firms



Notes: Figure presents the distribution of 1,730 publicly available GDPR fines from `enforcementtracker.com`, noting that not all GDPR fines are made public. Fines are presented in undeflated nominal terms (€). The restaurant fine was due to "Insufficient technical and organizational measures to ensure information security." The jewelry manufacturer was fined due to "Insufficient fulfillment of information obligations."

▸ largest fines ▸ fine types

2 Institutional Setting

2.1 The General Data Protection Regulation (GDPR)

2.2 Cloud Computing and Data

Our Data Source: Cloud Computing

- Cloud providers offer **on-demand** access to scalable IT resources through the Internet
- Firms request **storage** (hard drives), **computing** (virtual machines) and other IT services
- Widely adopted technology, over 90% of large organizations use cloud computing[1]

[1] Forbes (2022)

Our Data Source: Cloud Computing

- Cloud providers offer **on-demand** access to scalable IT resources through the Internet
- Firms request **storage** (hard drives), **computing** (virtual machines) and other IT services
- Widely adopted technology, over 90% of large organizations use cloud computing[1]

Some Examples of How Firms Use the Cloud

	Netflix	**Carrefour**
Data storage	Storage of video files, user info	Inventory & sales, online orders, payroll
Compute	Process requests, monitor quality, analytics	POS systems, supply chain

[1] Forbes (2022)

Data: Cloud Computing Usage from 2016 - 2021

1. **Cloud data**: detailed cloud usage from one of the largest service providers
   - Types: **storage** (gigabytes) and **computation** (number of cores $\times$ number of hours)
   - Unit of observation: firm−service−server location−month (e.g., MIT, Compute, East Coast, May/18)
   - Observe list prices and paid prices

Data: Cloud Computing Usage from 2016 - 2021

1. **Cloud data**: detailed cloud usage from one of the largest service providers
   - Types: **storage** (gigabytes) and **computation** (number of cores $\times$ number of hours)
   - Unit of observation: firm−service−server location−month (e.g., MIT, Compute, East Coast, May/18)
   - Observe list prices and paid prices

   Limitations:
   - We have limited knowledge on how firms use data stored in the cloud
   - May not capture all data and computation: multi-cloud and traditional IT

Data: Cloud Computing Usage from 2016 - 2021

1. **Cloud data**: detailed cloud usage from one of the largest service providers

2. **Aberdeen/Harte-Hanks**: establishment level technology adoption (including cloud)
   - Observe ~2.5 million US and ~2 million EU establishments
   - Provides information on the extensive margin of cloud adoption and multi-homing
   - Widely used in literature to measure IT adoption

     (Brynjolfsson and Hitt, 2003; Bresnahan et al., 2002; Bloom et al., 2012)

Data: Cloud Computing Usage from 2016 - 2021

1. **Cloud data**: detailed cloud usage from one of the largest service providers

2. **Aberdeen/Harte-Hanks**: establishment level technology adoption (including cloud)

3. **Duns & Bradstreet and Orbis**: information on industry classification and employment
   - Unit of observation: firm
   - Employment information available only for EU firms (fuzzy matching algorithm)

Summary Statistics: Top 8 Industries in Sample

Industry	Firms (%)	in EU (%)	Computation (%)	Storage (%)
Services	42.6	40.9	36.3	31.9
Software	25.4	59.8	17.6	20.8
Manufacturing	8.3	54.4	10.5	11.6
Retail Trade	5.8	46.9	5.2	5.4
Finance & Insurance	5.5	44.9	11.4	10.8
Wholesale Trade	5.2	52.3	3.7	4.5
Transportation	3.4	41.7	6.5	6.4
Construction	1.8	46.9	1.9	1.9
Total	98.0	48.1	93.1	93.3

Summary Statistics: Top 8 Industries in Sample

Industry	Firms (%)	in EU (%)	Computation (%)	Storage (%)
Services	42.6	40.9	36.3	31.9
Software	25.4	59.8	17.6	20.8
Manufacturing	8.3	54.4	10.5	11.6
Retail Trade	5.8	46.9	5.2	5.4
Finance & Insurance	5.5	44.9	11.4	10.8
Wholesale Trade	5.2	52.3	3.7	4.5
Transportation	3.4	41.7	6.5	6.4
Construction	1.8	46.9	1.9	1.9
Total	98.0	48.1	93.1	93.3

Summary Statistics: Top 8 Industries in Sample

Industry	Firms (%)	in EU (%)	Computation (%)	Storage (%)
Services	42.6	40.9	36.3	31.9
Software	25.4	59.8	17.6	20.8
Manufacturing	8.3	54.4	10.5	11.6
Retail Trade	5.8	46.9	5.2	5.4
Finance & Insurance	5.5	44.9	11.4	10.8
Wholesale Trade	5.2	52.3	3.7	4.5
Transportation	3.4	41.7	6.5	6.4
Construction	1.8	46.9	1.9	1.9
Total	98.0	48.1	93.1	93.3

Summary Statistics: Top 8 Industries in Sample

Industry	Firms (%)	in EU (%)	Computation (%)	Storage (%)
Services	42.6	40.9	36.3	31.9
Software	25.4	59.8	17.6	20.8
Manufacturing	8.3	54.4	10.5	11.6
Retail Trade	5.8	46.9	5.2	5.4
Finance & Insurance	5.5	44.9	11.4	10.8
Wholesale Trade	5.2	52.3	3.7	4.5
Transportation	3.4	41.7	6.5	6.4
Construction	1.8	46.9	1.9	1.9
Total	98.0	48.1	93.1	93.3

3 DiD Estimates of the Impact of GDPR

Main Empirical Specification

- **Challenge:** Lack of a natural control group due to regulation spillovers (Johnson, 2023)

- **Idea:** We observe the data centers firms use in addition to the country of origin
  - **Treated firms**: firms in the EU that store data only in EU data centers (domestic EU)
  - **Control firms**: firms in the US that store data only in US data centers (domestic US)

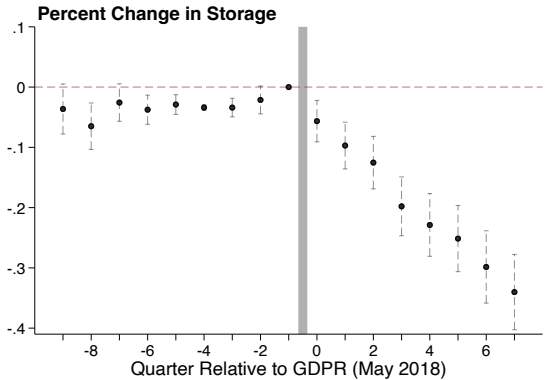
  $\longrightarrow$ Eliminates multi-national firms

- **Sample:** EU and US firms who continuously use the cloud 24-13 months before the GDPR

- Use difference-in-differences with flexible trends by industry and pre-GDPR usage deciles

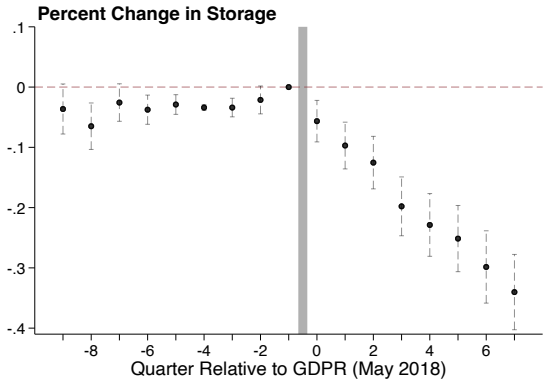
▸ control definition details

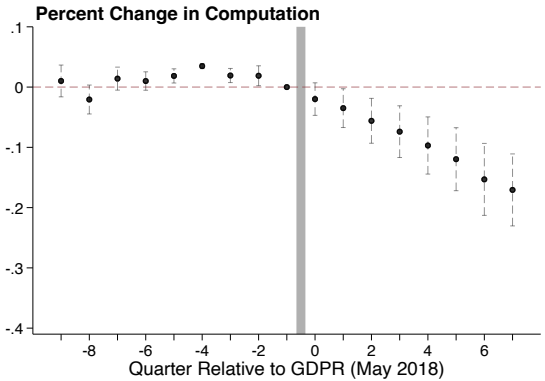
Decrease in Data Storage and Computation



- Sharp decrease in data storage

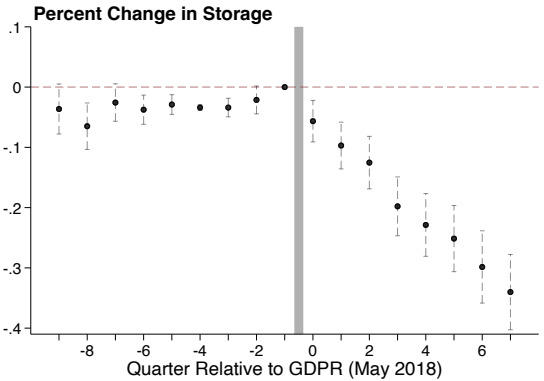
Decrease in Data Storage and Computation

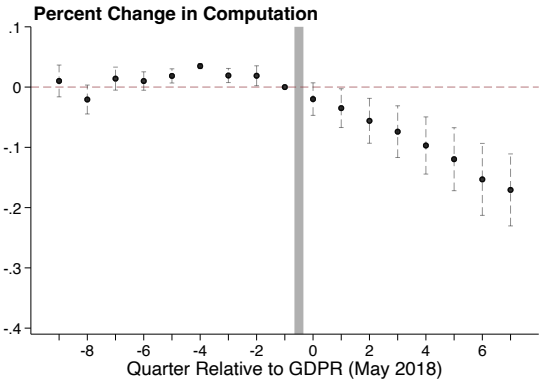




- Sharp decrease in both storage and compute, consistently larger drop in storage

Decrease in Data Storage and Computation

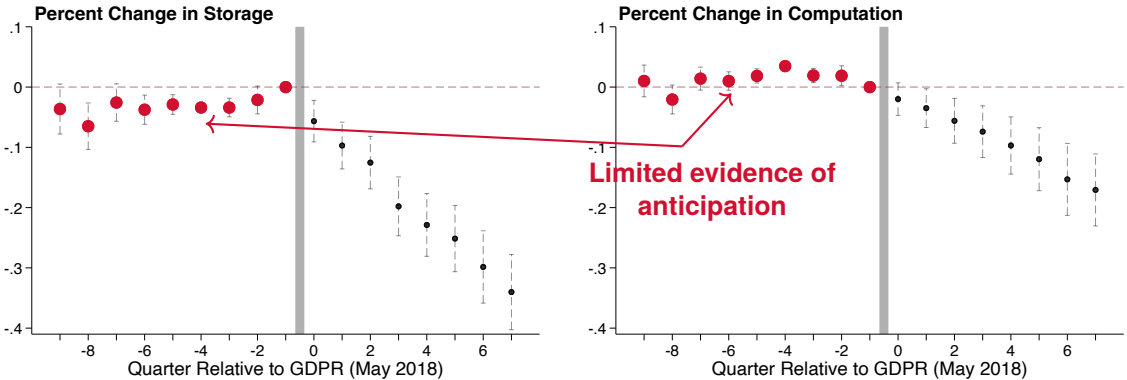




- Sharp decrease in both storage and compute, consistently larger drop in storage

→ Firms become less data intensive after the GDPR (suggests complementarity)

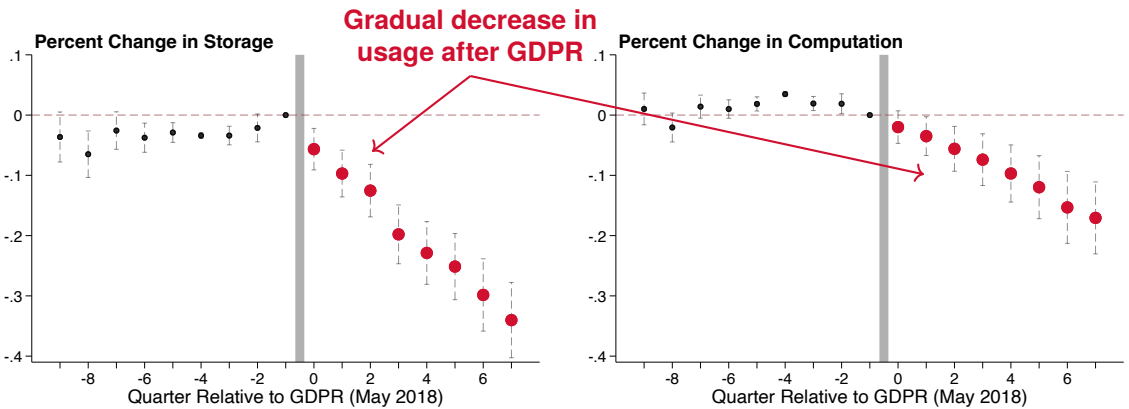
Decrease in Data Storage and Computation



- Sharp decrease in both storage and compute, consistently larger drop in storage

→ Firms become less data intensive after the GDPR (suggests complementarity)

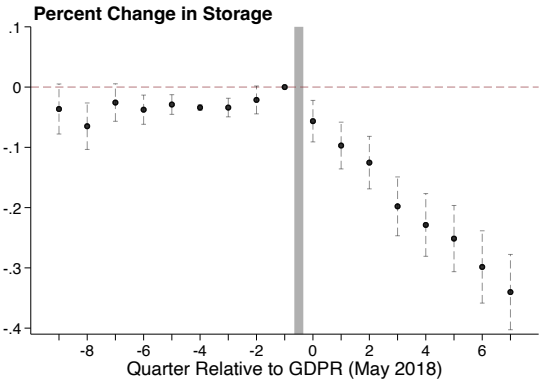
Decrease in Data Storage and Computation

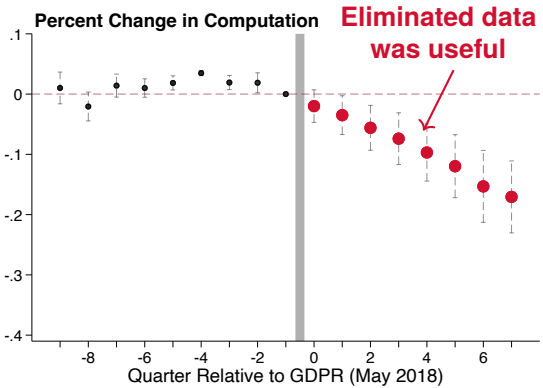


- Sharp decrease in both storage and compute, consistently larger drop in storage

→ Firms become less data intensive after the GDPR (suggests complementarity)

Decrease in Data Storage and Computation



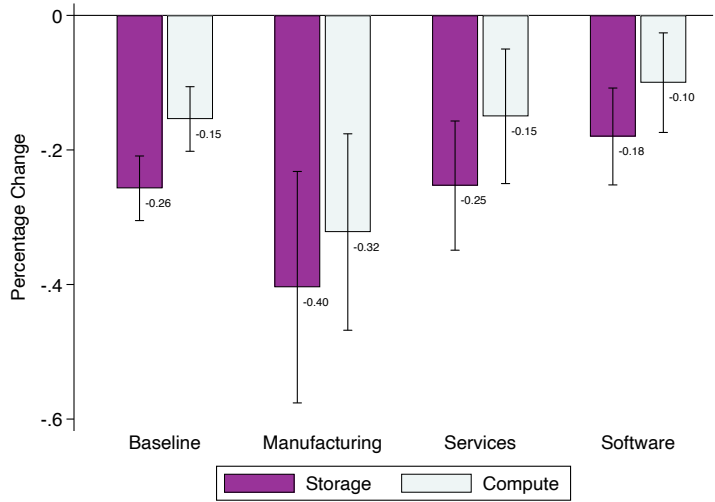


- Sharp decrease in both storage and compute, consistently larger drop in storage

→ Firms become less data intensive after the GDPR (suggests complementarity)

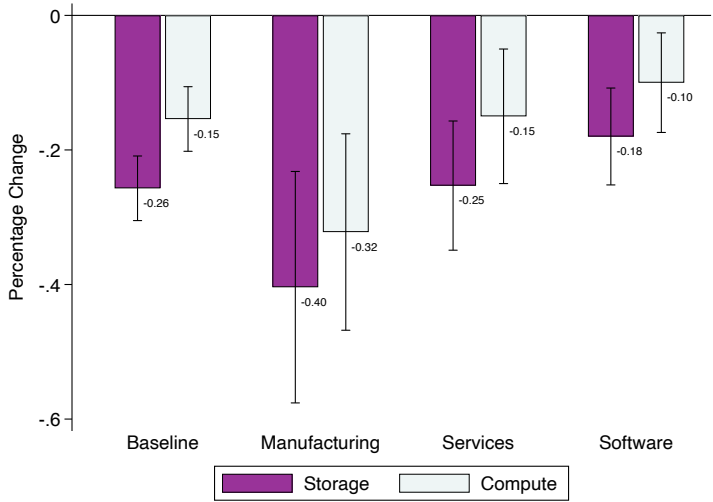
Estimated Long-run Effects by Industry (Two Years After GDPR)

- Primary findings are the same across all industry groups



Estimated Long-run Effects by Industry (Two Years After GDPR)

- Primary findings are the same across all industry groups
- Largest effect in manufacturing
- Suggestive that data and computation are less essential for manufacturing firms



Summary: GDPR Changes Firms' Data and Computation Input Choices

- **Additional Analyses:**
  1. Results not driven by differences in prices between US and EU
  2. Substitution (to other cloud providers, or in-house IT) unlikely to explain results
  3. Larger effect sizes, but not statistically significant wrt. country's enforcement strictness
     (Goldberg et al., 2023; Johnson, 2022)

- **Key Takeaways:**
  1. GDPR changed firms' data and computation input choices
  2. Results suggestive of a **wedge** between marginal product of storing data and its price

- **Next**: Model firms' input decisions using **production framework** to quantify the GDPR cost

4 | Production Function Framework

CES Information Production Function with Data and Compute

- Firms produce information (I_{it}) by using data (D_{it}) and computing (C_{it}) w/ CES tech:

$$I_{it} = \left(\omega_{it}^{c}(C_{it})^{\rho} + \alpha D_{it}^{\rho}\right)^{1/\rho}$$

  - ω_{it}^{c}: (unobserved) exogenous compute technology
  - $\sigma = 1/(1-\rho)$: the elasticity of substitution parameter, industry-specific

CES Information Production Function with Data and Compute

- Firms produce information (I_{it}) by using data (D_{it}) and computing (C_{it}) w/ CES tech:

$$I_{it} = \left(\omega_{it}^{c}(C_{it})^{\rho} + \alpha D_{it}^{\rho}\right)^{1/\rho}$$

  - ω_{it}^{c}: (unobserved) exogenous compute technology
  - $\sigma = 1/(1-\rho)$: the elasticity of substitution parameter, industry-specific

CES Information Production Function with Data and Compute

- Firms produce information (I_{it}) by using data (D_{it}) and computing (C_{it}) w/ CES tech:

$$I_{it} = \left(\omega_{it}^{c}(C_{it})^{\rho} + \alpha D_{it}^{\rho}\right)^{1/\rho}$$

  - ω_{it}^{c}: (unobserved) exogenous compute technology
  - $\sigma = 1/(1-\rho)$: the elasticity of substitution parameter, industry-specific

CES Information Production Function with Data and Compute

- Firms produce information (I_{it}) by using data (D_{it}) and computing (C_{it}) w/ CES tech:

$$I_{it} = \left(\omega_{it}^{c}(C_{it})^{\rho} + \alpha D_{it}^{\rho}\right)^{1/\rho}$$

  - ω_{it}^{c}: (unobserved) exogenous compute technology
  - $\sigma = 1/(1-\rho)$: the elasticity of substitution parameter, industry-specific

- Agnostic about how firms use data in production function, $Y = f(K, L, I)$

CES Information Production Function with Data and Compute

- Firms produce information (I_{it}) by using data (D_{it}) and computing (C_{it}) w/ CES tech:

$$I_{it} = \left(\omega_{it}^{c}(C_{it})^{\rho} + \alpha D_{it}^{\rho}\right)^{1/\rho}$$

  - ω_{it}^{c}: (unobserved) exogenous compute technology
  - $\sigma = 1/(1-\rho)$: the elasticity of substitution parameter, industry-specific

- Agnostic about how firms use data in production function, $Y = f(K, L, I)$

- **Cost minimization:** data storage and computing can be adjusted flexibly in the cloud
  - Firms take compute and data prices, p_{it}^{c} and p_{it}^{d} (cloud computing prices) as given
  - They then choose the optimal C_{it} and D_{it} to minimize information cost every period

Modeling The GDPR: A Wedge in the Production Function

- We model GDPR as an increase in the marginal cost of data storage by $(1 + \lambda_i)$:

$$\textbf{Pre-GDPR:} \quad \tilde{p}^d_{it} = p^d_{it} \qquad\qquad \textbf{Post-GDPR:} \quad \tilde{p}^d_{it} = (1 + \lambda_i) \cdot p^d_{it}$$

- Cost-minimization FOCs w.r.t. data and compute post-GDPR for EU firms is:
  (Doraszelski and Jaumandreu, 2018; Raval, 2019; Demirer, 2020)

$$\log\left(\frac{C_{it}}{D_{it}}\right) = \overbrace{\sigma}^{\text{elasticity of substitution}} \log\left(\frac{p^d_{it}}{p^c_{it}}\right) + \overbrace{\log(1 + \lambda_i)}^{\text{GDPR cost shock}} + \overbrace{\log(\omega^c_i) + \log(\phi^c_t) + \log(\varepsilon_{it})}^{\text{compute-augmenting technology } (\log \omega^c_{it})}$$

- Use equation to identify λ_i and σ using US/EU, pre/post GDPR variation + shift-share design

Modeling The GDPR: A Wedge in the Production Function

- We model GDPR as an increase in the marginal cost of data storage by $(1 + \lambda_i)$:

$$\textbf{Pre-GDPR:} \quad \tilde{p}^d_{it} = p^d_{it} \qquad \textbf{Post-GDPR:} \quad \tilde{p}^d_{it} = (1 + \lambda_i) \cdot p^d_{it}$$

- Cost-minimization FOCs w.r.t. data and compute post-GDPR for EU firms is:
  (Doraszelski and Jaumandreu, 2018; Raval, 2019; Demirer, 2020)

$$\log\left(\frac{C_{it}}{D_{it}}\right) = \overbrace{\sigma}^{\text{elasticity of substitution}} \log\left(\frac{p^d_{it}}{p^c_{it}}\right) + \overbrace{\log(1+\lambda_i)}^{\text{GDPR cost shock}} + \overbrace{\log(\omega^c_i) + \log(\phi^c_t) + \log(\varepsilon_{it})}^{\text{compute-augmenting technology } (\log \omega^c_{it})}$$

- Use equation to identify λ_i and σ using US/EU, pre/post GDPR variation + shift-share design

Modeling The GDPR: A Wedge in the Production Function

- We model GDPR as an increase in the marginal cost of data storage by $(1 + \lambda_i)$:

$$\textbf{Pre-GDPR:} \quad \tilde{p}^d_{it} = p^d_{it} \qquad \textbf{Post-GDPR:} \quad \tilde{p}^d_{it} = (1 + \lambda_i) \cdot p^d_{it}$$

- Cost-minimization FOCs w.r.t. data and compute post-GDPR for EU firms is:
  (Doraszelski and Jaumandreu, 2018; Raval, 2019; Demirer, 2020)

$$\log\left(\frac{C_{it}}{D_{it}}\right) = \overbrace{\sigma}^{\text{elasticity of substitution}} \log\left(\frac{p^d_{it}}{p^c_{it}}\right) + \overbrace{\log(1+\lambda_i)}^{\text{GDPR cost shock}} + \overbrace{\log(\omega^c_i) + \log(\phi^c_t) + \log(\varepsilon_{it})}^{\text{compute-augmenting technology } (\log \omega^c_{it})}$$

- Use equation to identify λ_i and σ using US/EU, pre/post GDPR variation + shift-share design

Modeling The GDPR: A Wedge in the Production Function

- We model GDPR as an increase in the marginal cost of data storage by $(1 + \lambda_i)$:

$$\textbf{Pre-GDPR:} \quad \tilde{p}^d_{it} = p^d_{it} \qquad \textbf{Post-GDPR:} \quad \tilde{p}^d_{it} = (1 + \lambda_i) \cdot p^d_{it}$$

- Cost-minimization FOCs w.r.t. data and compute post-GDPR for EU firms is:
  (Doraszelski and Jaumandreu, 2018; Raval, 2019; Demirer, 2020)

$$\log\left(\frac{C_{it}}{D_{it}}\right) = \overbrace{\sigma}^{\text{elasticity of substitution}} \log\left(\frac{p^d_{it}}{p^c_{it}}\right) + \overbrace{\log(1+\lambda_i)}^{\text{GDPR cost shock}} + \overbrace{\log(\omega^c_i) + \log(\phi^c_t) + \log(\varepsilon_{it})}^{\text{compute-augmenting technology } (\log \omega^c_{it})}$$

- Use equation to identify λ_i and σ using US/EU, pre/post GDPR variation + shift-share design

Modeling The GDPR: A Wedge in the Production Function

- We model GDPR as an increase in the marginal cost of data storage by $(1 + \lambda_i)$:

$$\textbf{Pre-GDPR:} \quad \tilde{p}^d_{it} = p^d_{it} \qquad \textbf{Post-GDPR:} \quad \tilde{p}^d_{it} = (1 + \lambda_i) \cdot p^d_{it}$$

- Cost-minimization FOCs w.r.t. data and compute post-GDPR for EU firms is:
  (Doraszelski and Jaumandreu, 2018; Raval, 2019; Demirer, 2020)

$$\log\left(\frac{C_{it}}{D_{it}}\right) = \overbrace{\sigma}^{\text{elasticity of substitution}} \log\left(\frac{p^d_{it}}{p^c_{it}}\right) + \overbrace{\log(1+\lambda_i)}^{\text{GDPR cost shock}} + \overbrace{\underbrace{\log(\omega^c_i)}_{\text{firm-specific component}} + \underbrace{\log(\phi^c_t)}_{\text{industry time trend}} + \underbrace{\log(\varepsilon_{it})}_{\text{time-varying shocks}}}^{\text{compute-augmenting technology } (\log \omega^c_{it})}$$

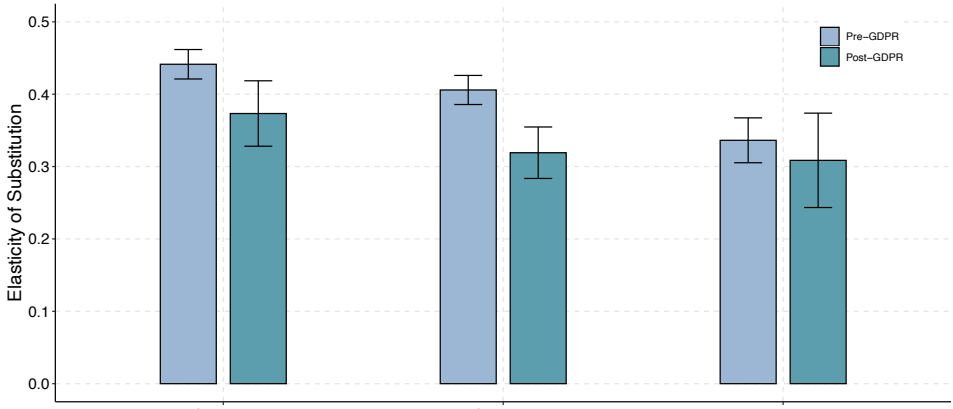
- Use equation to identify λ_i and σ using US/EU, pre/post GDPR variation + shift-share design

4 Production Function Framework

4.1 Estimation Results

Results on Elasticity of Substitution for EU Firms

Compute and data are strong complements; more so than “traditional inputs”



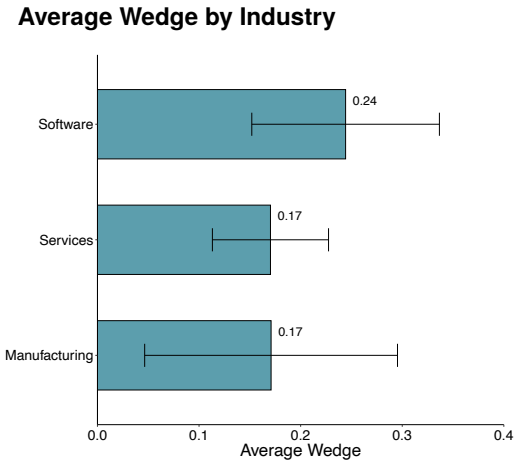
Notes: This figure presents our estimation results of the elasticity of substitution between storage and computing (σ) across industries. We present separate estimates for the pre- and post-GDPR (σ_1 and σ_2, respectively). Standard errors are calculated using 100 bootstrap repetitions.

5 The Production Cost of GDPR

5.1 Changes in the Cost of Data Storage

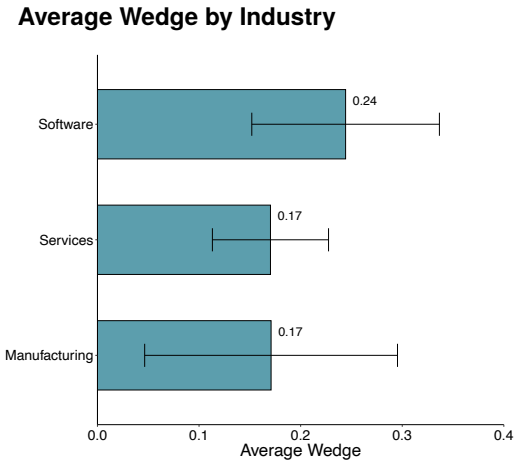
5.2 Changes in the Cost of Information Production

Average GDPR Wedge is 20% with Important Heterogeneity



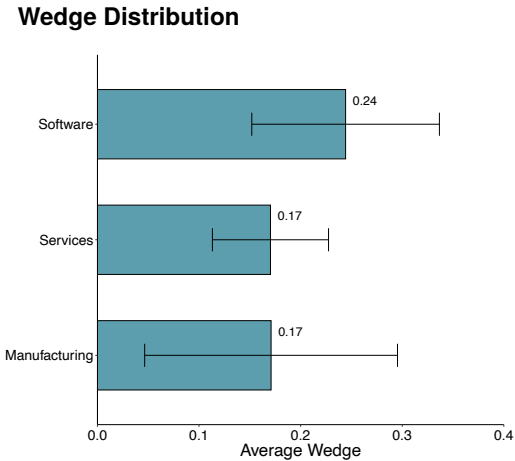
- GDPR $\simeq$ 20% tax on price of storing data

Average GDPR Wedge is 20% with Important Heterogeneity



- GDPR ≃ 20% tax on price of storing data
- Firms where information is likely more important face larger costs:
  - Software (≃ 24%) vs Manufacturing (≃ 17%)

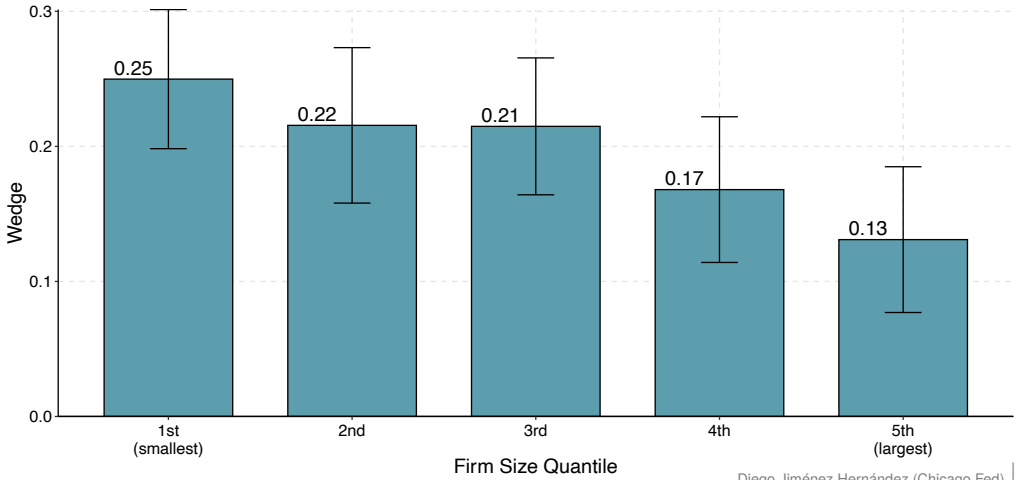
Average GDPR Wedge is 20% with Important Heterogeneity



- GDPR ≃ 20% tax on price of storing data
- Firms where information is likely more important face larger costs:
  - Software (≃ 24%) vs Manufacturing (≃ 17%)
- What explains the large cost heterogeneity?

Wedges Negatively Correlated with Firm Size (Employment)

- Larger firms face lower wedges, consistent with the literature
  (Campbell et al., 2015; Koski and Valmari, 2020; Goldberg et al., 2023)



5 The Production Cost of GDPR

5.1 Changes in the Cost of Data Storage

5.2 Changes in the Cost of Information Production

How Much Does GDPR Increase the Cost of Producing Information?

- From CES production function, the cost of producing a unit of information (without subscripts):

$$CI^*(p, \omega^c, \lambda) = \left[(\omega^c)^\sigma \left(p^c\right)^{1-\sigma} + \alpha^\sigma \left((1+\lambda) p^d \right)^{1-\sigma} \right]^{1/(\sigma-1)}$$

heterogeneity depends on prices, compute productivity, elasticity of substitution, and wedges

How Much Does GDPR Increase the Cost of Producing Information?

- From CES production function, the cost of producing a unit of information (without subscripts):

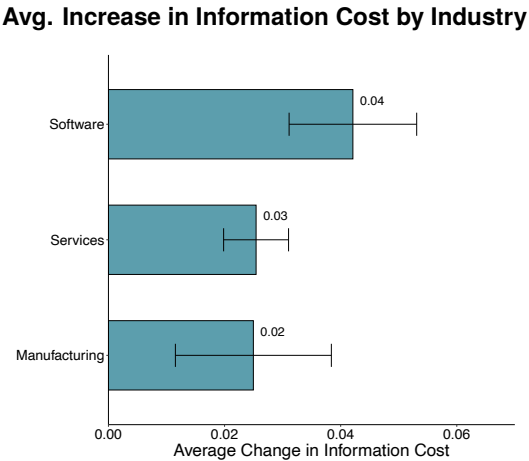
$$CI^*(p, \omega^c, \lambda) = \left[(\omega^c)^\sigma \left(p^c\right)^{1-\sigma} + \alpha^\sigma \left((1+\lambda)p^d\right)^{1-\sigma} \right]^{1/(\sigma-1)}$$

  heterogeneity depends on prices, compute productivity, elasticity of substitution, and wedges

- Calculate counterfactual information cost without GDPR ($\lambda_i = 0$)

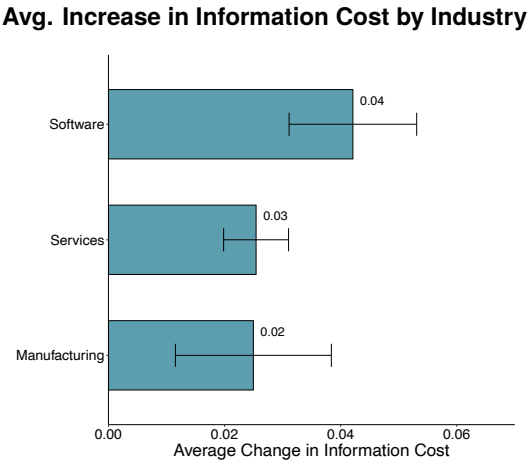
With GDPR: $CI^*(p, \omega^c, \lambda = \lambda_i)$ **Without GDPR:** $CI^*(p, \omega^c, \lambda = 0)$

Average Increase in Information Cost is Only 3.7%



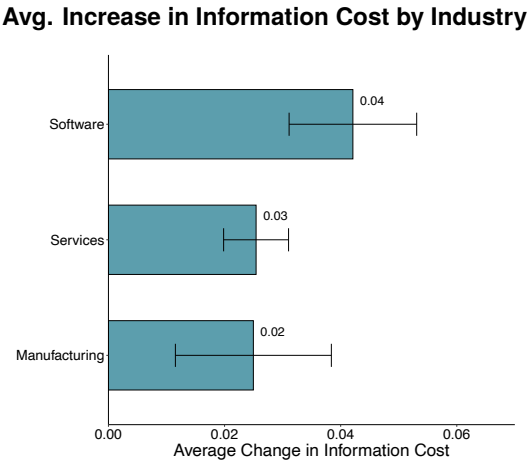
- Information cost increases only by 3.7% on average, with important heterogeneity

Average Increase in Information Cost is Only 3.7%



- Information cost increases only by 3.7% on average, with important heterogeneity
- How to reconcile 3.7%↑ in cost of information with 20%↑ in wedges? Smaller data expenditure shares ($\sim 20\%$)

Average Increase in Information Cost is Only 3.7%



- Information cost increases only by 3.7% on average, with important heterogeneity
- How to reconcile 3.7%↑ in cost of information with 20%↑ in wedges? Smaller data expenditure shares ($\sim 20\%$)
- Back-of-the-envelope under Cobb-Douglas: production costs ↑ 0.34%-0.66% for software; 0.05%-0.15% for less-data intensive industries

6 | Conclusions

Conclusion

What We Do:

- Use a production approach to study the effects on GDPR on data and computation

Results:

- DiD estimates suggest that GDPR reduced firm demand for data and computation:
  - Firm storage declined by 26%; computing declined by 15%
- Data and computation are strong complements in production function
- Production function framework estimates GDPR $\simeq$ 20% tax on data storage:
  - This leads to only 4% increase in the cost of information because it targets cheaper input
  - Total production costs are lower ($\leq 1\%$) since information expenditure shares tend to be low