

# Data, Privacy Laws, and Firm Production:

# Growing Importance of Data for Firms & Privacy Laws

- **Data**

## 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
  - A ffected 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

## 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

-



## 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 \_\_\_\_\_



## 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
  -

## Preview of Results

- 1.

## Preview of Results

- 1.

# 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**
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

Introduction

## 2 Institutional Setting

DiD Estimates of the Impact of GDPR

Production Function Framework

The Production Cost of GDPR

Conclusions

## 2 Institutional Setting

2.1 The General Data Protection Regulation (GDPR)

2.2 Cloud Computing and Data

# 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
  -

## 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:**

---

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

# 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      Variable

Costs      Costs

1. Operational changes: privacy notices, employee training (Art. 25)

# GDPR Affects a Wide Range of Industries and Firms



## 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**

## Our Data Source: Cloud Computing

- Cloud providers offer **on-demand** access to scalable IT resources through the Internet
- Firms request **storage**

### Some Examples of How Firms Use the Cloud



# 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
  - [Wid29.ibBe0ssOtailedginxt](#)

# 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











# Main Empirical Specification

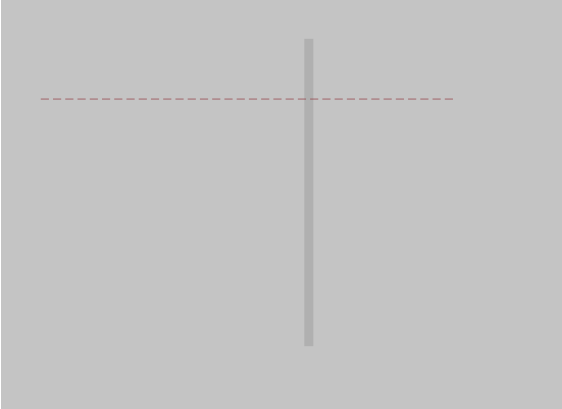
- **Challenge:** Lack of a natural control group due to regulation spillovers (Johnson, 2023)
- **Idea:** We observe the data centers firms use in addition to the country of origin
  - **Treated firms:** firms in the EU that store data only in EU data centers (domestic EU)
  - **Control firms:** firms in the US that store data only in US data centers (domestic US)  
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

## Decrease in Data Storage and Computation

# Decrease in Data Storage and Computation



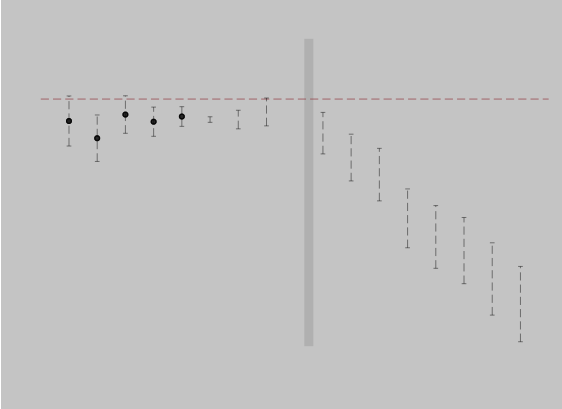
# Decrease in Data Storage and Computation



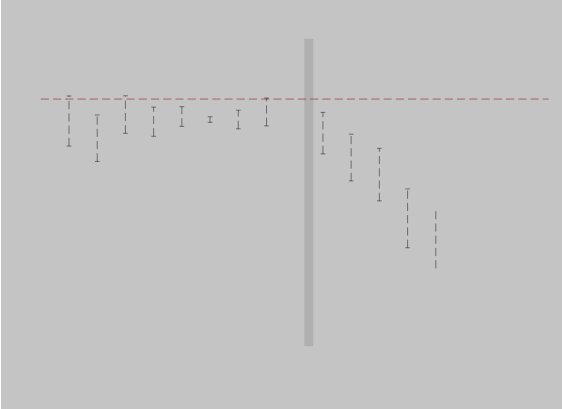
# Decrease in Data Storage and Computation



# Decrease in Data Storage and Computation



# Decrease in Data Storage and Computation





Estimated Long-run E

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

- Primary findings are the same across all industry groups
- Largest e

# 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

Introduction

Institutional Setting

DiD Estimates of the Impact of GDPR

## 4 Production Function Framework

The Production Cost of GDPR

Conclusions

# CES Information Production Function with Data and Compute

- Firms produce information ( ) by using data ( ) and computing ( ) w/ CES tech:

$$= ( ) + 1/$$

- : (unobserved) exogenous compute technology
- =  $1/(1 - )$ : the elasticity of substitution parameter, industry-specific

# CES Information Production Function with Data and Compute

Firms produce information ( ) by using data ( ) and computing ( ) w/ CES tech:

$$= ( ) + 1/$$

- : (unobserved) exogenous compute technology
- = 1/(1 - ): the elasticity of substitution parameter, industry-specific

# CES Information Production Function with Data and Compute

Firms produce information ( ) by using data ( ) and computing ( ) w/ CES tech:

$$= \left( \alpha \left( \beta D^{\rho} + (1-\beta) C^{\rho} \right)^{\frac{1}{\rho}} \right)^{\frac{1}{1-\sigma}}$$

# CES Information Production Function with Data and Compute

- Firms produce information ( $y$ ) by using data ( $x$ ) and computing ( $z$ ) w/ CES tech:

$$y = \left( \alpha x^\rho + (1-\alpha) z^\rho \right)^{1/\rho}$$

- $\alpha$ : (unobserved) exogenous compute technology
  - $\rho = 1/(1 - \sigma)$ : the elasticity of substitution parameter, industry-specific
- Agnostic about how firms use data in production function,  $\rho = (\rho_x, \rho_z)$



# CES Information Production Function with Data and Compute

- Firms produce information ( ) by using data ( ) and computing ( ) w/ CES tech:

$$= ( ) + 1/$$

- : (unobserved) exogenous compute technology
- = 1/(1 - ): the elasticity of substitution parameter, industry-specific

- Agnostic about how firms use data in production function, = ( ,

# Modeling The GDPR: A Wedge in the Production Function

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

**Pre-GDPR:**  $\tilde{c} =$                       **Post-GDPR:**  $\tilde{c} = (1 + \delta) \cdot$



# Modeling The GDPR: A Wedge in the Production Function

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

**Pre-GDPR:**  $\tilde{c} =$

# Modeling The GDPR: A Wedge in the Production Function

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

$$\text{Pre-GDPR: } \tilde{c} = c \quad \text{Post-GDPR: } \tilde{c} = (1 + \delta) \cdot c$$

- Cost-minimization FOCs w.r.t. data and compute post-GDPR for EU firms is:

(Doraszelski and Jaumandreu, 2018; Raval, 2019; Demirer, 2020)

$$\log \frac{w}{r} = \overbrace{\log \frac{w}{r}}^{\text{elasticity of substitution}} + \overbrace{\log(1 + \delta)}^{\text{GDPR cost shock}} + \overbrace{\log(\frac{1}{\alpha}) + \log(\frac{1}{1-\alpha}) + \log(\frac{1}{1-\alpha})}^{\text{compute-augmenting technology } (\log \frac{1}{\alpha})}$$

- Use equation to identify  $\alpha$  and  $\delta$  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  $(1 + \delta)$ :

Pre-GDPR:  $\tilde{c} = c$       Post-GDPR:  $\tilde{c} = (1 + \delta) \cdot c$

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

$$\log \frac{w}{r} = \underbrace{\frac{1}{\sigma}}_{\text{elasticity of substitution}} \log \frac{w}{r} + \underbrace{\log(1 + \delta)}_{\text{GDPR cost shock}} + \underbrace{\log(\epsilon) + \log(\gamma) + \log(\eta)}_{\text{compute-augmenting technology } (\log \epsilon)}$$

$\log(\epsilon)$   
firm-specific component
 $\log(\gamma)$   
industry time trend
 $\log(\eta)$   
time-varying shocks

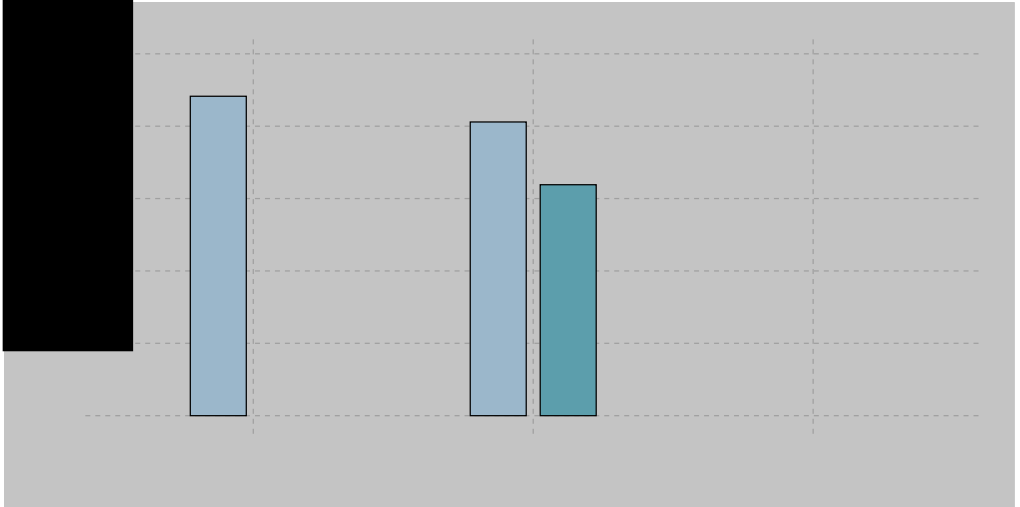
- Use equation to identify  $\sigma$  and  $\delta$  using US/EU, pre/post GDPR variation + shift-share design

# 4 Production Function Framework

## 4.1 Estimation Results

# Result: Elasticity of Substitution for EU Firms

Computations show strong complements; more so than "traditional inputs"





Introduction

Institutional Setting

DiD Estimates of the Impact of GDPR

Production Function Framework

## 5 The Production Cost of GDPR

Conclusions

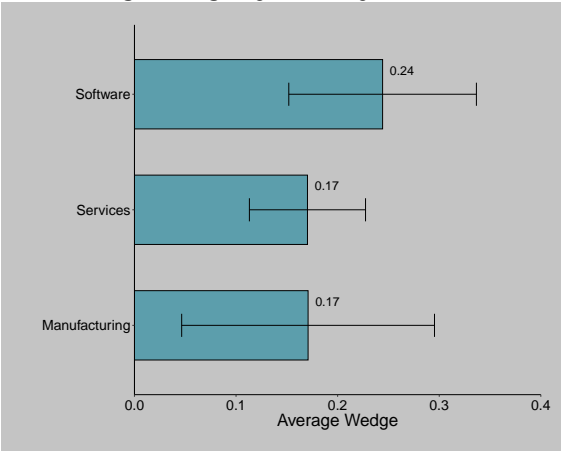
# 5 The Production Cost of GDPR

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

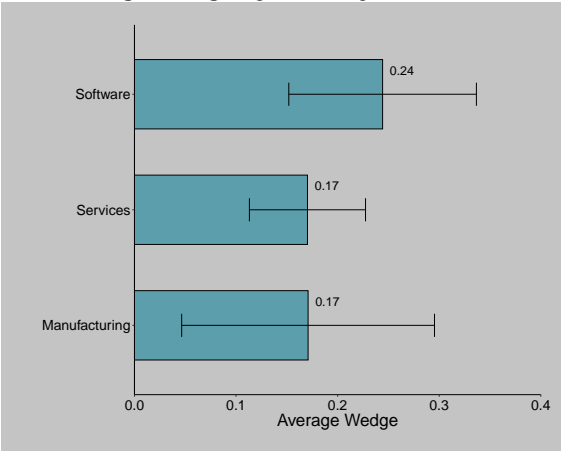
Average Wedge by Industry



GDPR 20% tax on price of storing data

# Average GDPR Wedge is 20% with Important Heterogeneity

Average Wedge by Industry



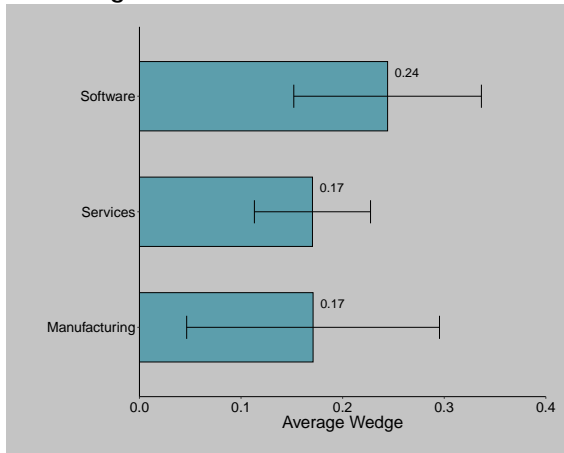
GDPR 20% tax on price of storing data

Firms where information is likely more important face larger costs:

- Software ( 24%) vs Manufacturing ( 17%)

## Average GDPR Wedge is 20% with Important Heterogeneity

Wedge Distribution



GDPR 20% tax on price of storing data

Firms where information is likely more important face larger costs:

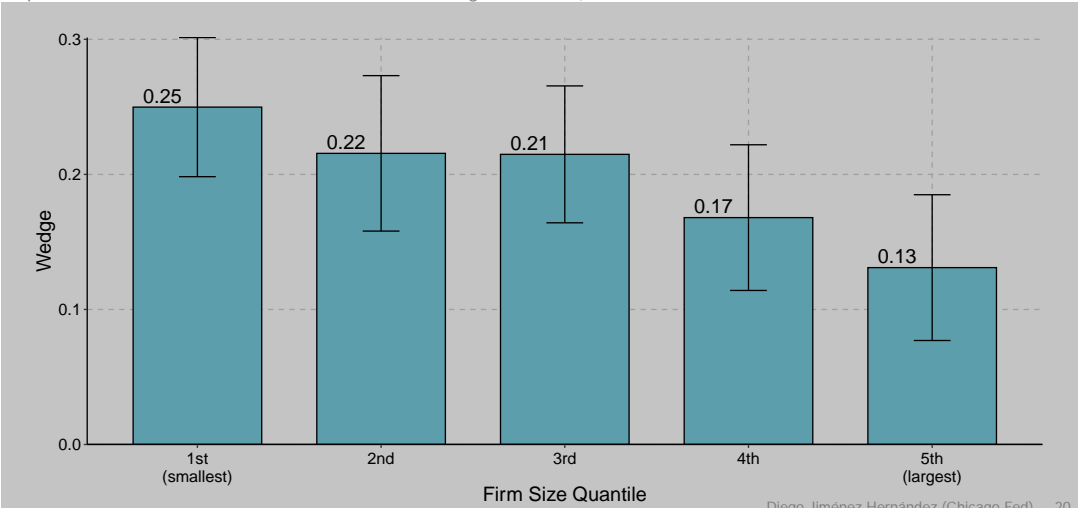
- Software ( 24%) vs Manufacturing ( 17%)

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):

$$(c, w, r) = (c) (w)^{1-\rho} + (1 +$$



## How Much Does GDPR Increase the Cost of Producing Information?

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

$$C(p, w, \tau) = \left( \frac{p}{w} \right)^{\frac{1}{1-\sigma}} + (1 + \tau)^{\frac{1}{1-\sigma}}$$

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

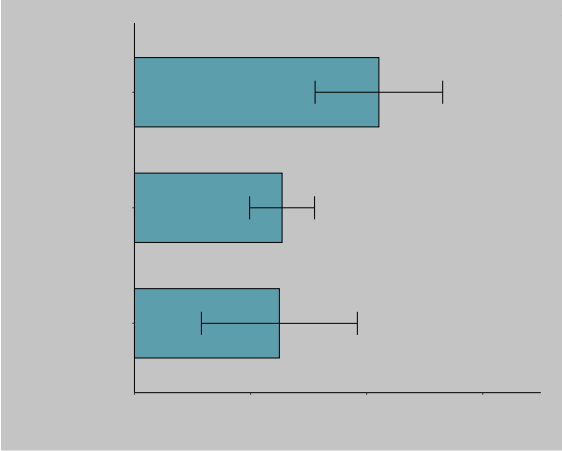
- Calculate counterfactual information cost without GDPR ( $\tau = 0$ )

**With GDPR:**  $C(p, w, \tau = \tau)$

**Without GDPR:**  $C(p, w, \tau = 0)$

# Average Increase in Information Cost is Only 3.7%

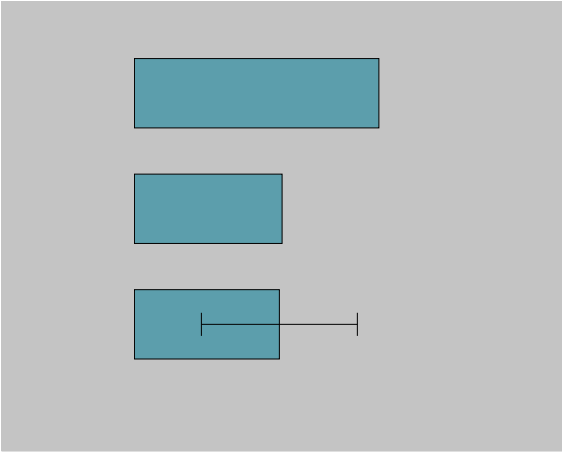
Avg. Increase in Information Cost by Industry



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

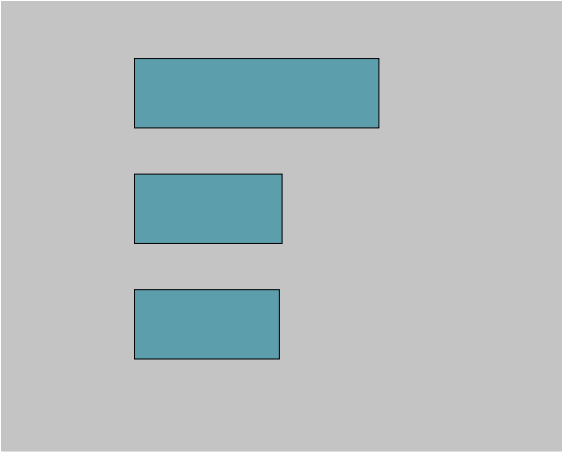
# Average Increase in Information Cost is Only 3.7%

Avg. Increase in Information Cost by Industry



# Average Increase in Information Cost is Only 3.7%

**Avg. Increase in Information Cost by Industry**



Introduction

Institutional Setting

DiD Estimates of the Impact of GDPR

Production Function Framework

The Production Cost of GDPR

## 6 Conclusions

# Conclusion

## What We Do:

- Use a production approach to study the effects on GDP 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 = 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 ( - 1%) since information expenditure shares tend to be low