The Value of Information in Competitive Markets: The Impact of Big Data on SMEs

Jose E. Galdon-Sanchez (Universidad Publica de Navarra)
Ricard Gil (Queen's University)
Guillermo Uriz-Uharte (University College London)

July 10, 2020

"The world´s most valuable resource is no longer oil, but data."

- The Economist, May 6th 2017

"From the dawn of civilization to 2003, we generated five exabytes of data. Now we generate that every two days and the pace is accelerating."

- Eric Schmidt, CEO Google, August 4th 2010

Introduction

- In a competitive market, information can be a source of competitive advantage: new business opportunities, better resource allocation, ...

- Potential high impact of Big Data IT on the economy, yet adoption is concentrated among large corporations:
  - Improvement on workplace organization (Brynjolfsson and Hitt, 2000; Bresnahan et al., 2002; Bartel et al., 2007) and management practices (Bloom et al., 2013; Bruhn et al., 2018; Giorcelli, 2019).
  - Improvement on market demand estimation (Bajari et al., 2019).
  - Enabling data-driven decision-making (DDDs) capabilities (Brynjolfsson et al., 2011; Brynjolfsson and McElheran, 2016a and 2016b; Agrawal et al, 2018; Kim, 2019).

Introduction (Cont.)

- Concentration of Big Data adoption among large corporations may increase market power concentration and reduce competition, potentially hindering innovation, R&D, etc.
- Important to understand why are small and medium enterprises (SMEs) not adopting? (Angle and Forman, 2018; Kim, 2019)
  - Do SMEs have lower returns to adoption? Do SMEs face higher adoption costs?
- A number of questions relevant for policy and managers arise:
  - Are private optimal decisions aligned with social optimal?
  - Should government facilitate access to Big Data for SMEs?

Our Research Project and Research Question

- We aim to answer two research questions:
  1. Do SMEs benefit from Big Data IT?
  2. How does Big Data IT Benefit SMEs (underlying mechanisms)?

- We evaluate the impact of an information-sharing program diffused by a large bank in Spain, designed to bring Big Data to SMEs owners of a credit-card point-of-sale (POS hereafter).
  - Upon adoption, the bank generated a monthly customer-specific report using credit-card transaction data.
  - The bank provided this program for free and adoption was voluntary.



Preview of Results

- Adoption of Big Data technology increases quarterly revenues of SMEs by 9%.

- Effect of adoption is largest for smallest SMEs, and does not differ according to sophistication or competition.

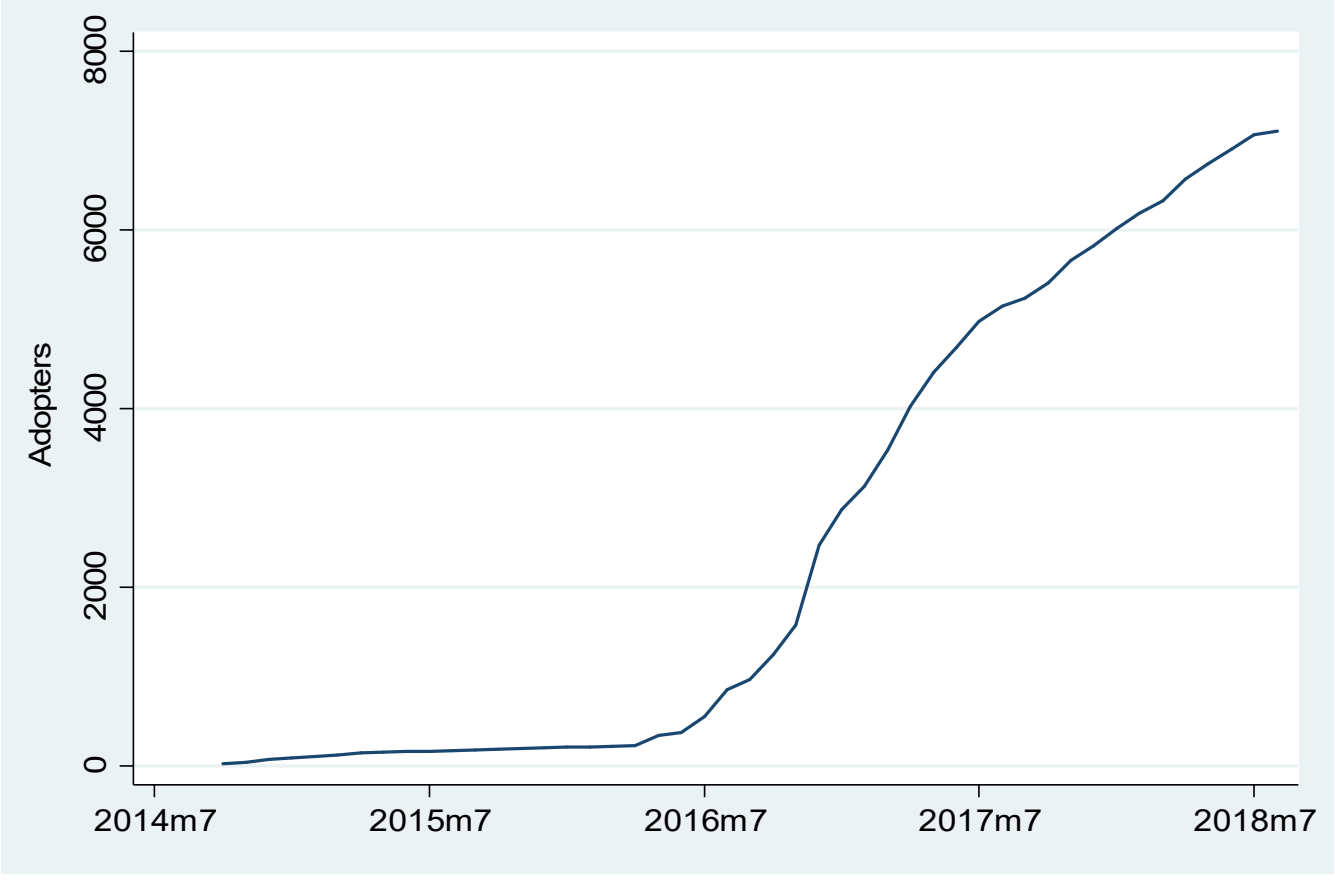
- Increase in quarterly revenues due to two distinct mechanisms:
  - "Demand-Side": discover new business opportunities, reaching to underserved demographics and diversifying their clientele portfolios.
  - "Supply-Side": reshuffling effort and workload from busy to quiet week times.

Context: The Program

- Program introduced by a large bank in July 2016.
  - Bank market share slightly above 15% in both credit cards and POS.
  - Pilot program in a region in mid 2014.
- Program generates monthly customer-specific report using information on credit-card transactions from customers of the bank (consumers).
- Free and voluntary adoption for all establishments with POS of the bank.
- Bank employees not compensated for diffusion of program (added value to customer)

Adopters 2014-2018

Figure 1: Number of adopters over time



Adopters per Sector

Adopters per Sector

Sector	Adopters
Restaurants	952
Clothing stores	726
Grocery stores	604
Hairdressing and beauty	411
Bars and cafeterias	404
Jewellery	333
Motor garages and car delearships	287
Shoe shops	270
Hotels	245
Pharmacies	236

Note: Ten sectors with the highest number of adopters up to July 2018.

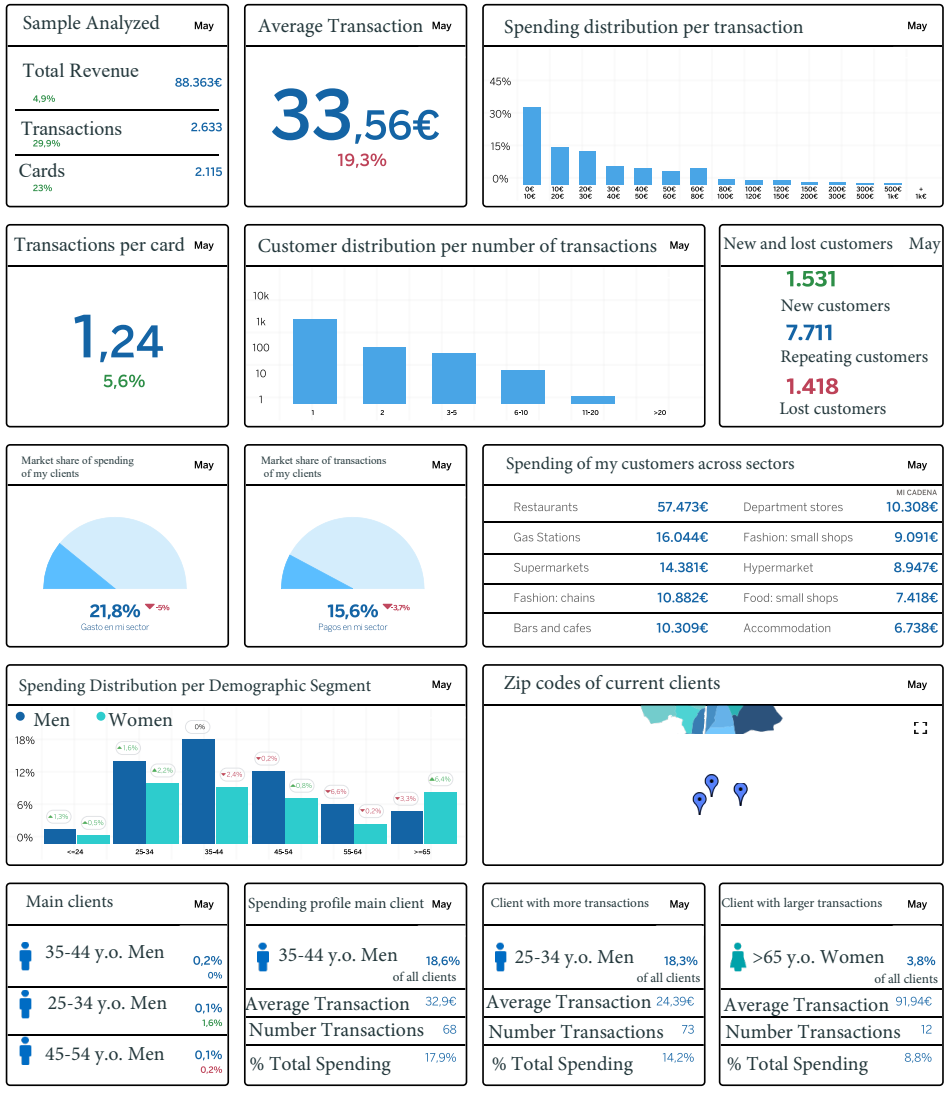
The Program (Cont.)

- Summary statistics on the number and value of credit-card transactions in the establishment during the previous month.
- Information disaggregated by client demographic group: age ($<$25, 25-34, 35-44, 45-54, 55-64, 65$+$), gender (male-female), zipcode, new or returning customers.
- Information also separated by time and day of week of purchases.
- Report compares focal establishment's performance to other establishments/competitors in the same sector and zipcode/city.

The Report - An Example

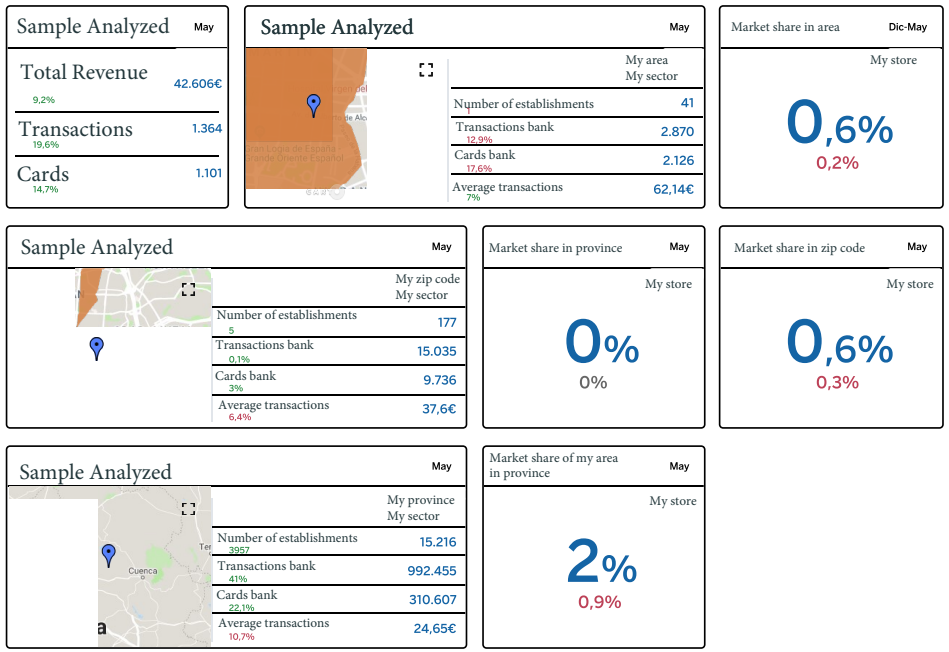
BANK

My Store



The Report - An Example (Cont.)

Bank
My Store



Our Data

- Credit-card transaction data from January 2014 to December 2018.
  - All transactions made with bank credit-cards in all POS: credit-card identifier, establishment identifier, time, value of transaction, information on the credit-card owner (age, gender, zipcode of residence, ...).
- Information for all retail establishments in the country with a POS: adress, sector, subsector.
  - Drop all establishments with less than 5 transactions per quarter.
  - Focus on establishments in sector-zipcode with 1+ adopter.
  - Despite data drop, we preserve meaningful variation to estimate impact of technology adoption.
- Final sample: quarterly aggregated data for 310,610 establishments, of which 7100 adopters.

Summary Statistics

Table 1: Descriptive Statistics

	Observ.	Mean	Std. Dev.	Min	Max
Full Sample					
Revenue	4,610,085	4,715	29,171	12	7,948,335
Transactions	4,610,085	120	710	5	227,139
Average Value of Transactions	4,610,085	64	101	2	15,000
Customers	4,610,085	74	338	2	134,725
Average Value per Customer	4,610,085	85	198	1	92,066
Adopters					
Revenue	63,639	6,248	18,730	15	537,791
Transactions	63,639	153	462	5	8,146
Average Value of Transactions	63,639	80	147	3	7,006
Customers	63,639	92	224	3	5,975
Average Value per Customer	63,639	102	200	1	10,500
Number of competitors	63,639	75	96	0	1,020
Sophistication	3,495	3.53	0.89	1.00	5.00

Notes: Statistics computed from a sample with quarterly level information at the establishment level.

Methodology

- Baseline model:

$$Y_{isjt} = \beta \; Adoption_{isjt} \; + \; \alpha_i \; + \; \delta_{sjt} \; + \; \epsilon_{isjt} \tag{1}$$

  - Y_{isjt} is the outcome of interest at the quarter level for establishment i, in sector s, in zipcode j and quarter t (revenues in baseline results).
  - $Adoption_{isjt}$ is a dummy equal to 1 if the establishment has adopted the technology before quarter t.
- We estimate (1) in first-differences,

$$\begin{aligned} Y_{isjt} - Y_{isjt-1} &= \beta \; (Adoption_{isjt} - Adoption_{isjt-1}) \; + \\ &+ \; (\delta_{sjt} - \delta_{sjt-1}) \; + \; (\epsilon_{isjt} - \epsilon_{isjt-1}) \end{aligned} \tag{2}$$

- Several endogeneity concerns: adoption may be more likely (i) in high performace SMEs, (ii) in fast-growing sectors and areas or SMEs. Specifications use SME FE, sector-area FE, SME-specific trends. To further address endogeneity (adoption may correlate with other investments), we use an IV.

Identification Strategy - IV Estimation

- We exploit differences in adoption propensity across establishments in the same sector-zipcode-quarter coming from a double source:
  1. Different establishments in the same sector-zipcode-quarter can have their accounts in different bank branches.
  2. Some bank branches put effort in advertising this technology and some others did not. They were not compensated for it.
- **Instrument the adoption decision of an establishment with the effort put by its bank branch in diffusing the technology.**
- We use the **number of other adopters in the bank branch in that period** to proxy for the effort of the bank branch in diffusing the technology.
- **Exclusion restriction**:

$$\mathbb{E}[\epsilon_{ijst}|\alpha_i, \delta_{sjt}, branch = A] = \mathbb{E}[\epsilon_{kjst}|\alpha_k, \delta_{sjt}, branch = B]$$

Main Results

Table 2: Baseline Results

Dependent variable: Δ Log revenue

	OLS (1)	OLS (2)	OLS (3)	1st-stg (4)	2nd-stg (5)
Δ Adoption t-1			0,00978		
			(0.0158)		
Δ Adoption	0.0455***	0.0458***			0.0902**
	(0.0157)	(0.0157)			(0.0386)
Δ Adoption t+1		-0,00263			
		(0.0148)			
Δ Adoption t+2		0,00395			
		(0.0161)			
Δ Adoption t+3		0,025			
		(0.0164)			
Peers IV				0.00446***	
				(0.00012)	
Sector-zipcd-quarter FE	Yes	Yes	Yes	Yes	Yes
Dummies first 4 quarters	Yes	Yes	Yes	Yes	Yes
Observations	4,610,085	4,610,085	4,610,085	4,610,085	4,610,085

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. . Standard errors are clustered at the establishment level and reported in parenthesis.

Heterogeneous Effects

Table 3: Heterogeneous Effects

Dependent variable: Δ Log revenue

	Sophistication		Size		Competition	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	OLS	IV	OLS	IV
Δ Adoption x High	0.0442*	0.0873**	0.0139	0.0725*	0.068***	0.109***
	(0.0232)	(0.0391)	(0.0171)	(0.0378)	(0.0209)	(0.0403)
Δ Adoption x Low	0.0463**	0.0976**	0.0796***	0.146***	0.0206	0.0629
	(0.021)	(0.0458)	(0.0267)	(0.0481)	(0.0216)	(0.0419)
Residual CF		-0.0541		-0.0702		-0.0471
		(0.0454)		(0.0443)		(0.0437)
Sector-zipcd-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Dummies first 4 quarters	Yes	Yes	Yes	Yes	Yes	Yes
p-value null equal returns	0.946	0.752	0.0374	0.0226	0.0986	0.106
Observations	4,610,085	4,610,085	4,610,085	4,610,085	4,610,085	4,610,085

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the establishment level and reported in parenthesis.

Mechanisms

- While exploring heterogeneity is interesting, understanding mechanisms is key for policy and managerial implications.
- We consider two different mechanisms:
  1. Demand-Side mechanism: Big Data enables establishments to discover new business opportunities, new customers types.
     - We show revenue increase is due to an increase in number of customers.
     - We show adoption changes clientele demographics, increase in share of customer type when SME lagging pre-adoption, decrease in concentration of customer types, and increase in customers from other zipcodes.
  2. Supply-Side mechanism: Big Data enables establishments to reshuffle effort more efficiently across peak and off-peak periods of time.
     - Identification problem: off-peak times may be where new business opportunities are.
     - Is there an increase in activity in off-peak hours when controlling for changes in customer types?

Demand-Side Mechanism

Table 4: Effects on Other Outcomes

Dep variable: Δ Log number of transactions, Δ log number of customers, Δ log average transaction value

	Transactions		Customers		Trans/Cust	
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Adoption	0.0436***	0.130***	0.0385***	0.119***	0.00514	0.0101
	(0.0120)	(0.0316)	(0.0113)	(0.0301)	(0.00325)	(0.00711)
Sector-zipcd-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Dummies first 4 quarters	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,610,085	4,610,085	4,610,085	4,610,085	4,610,085	4,610,085

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the establishment level and reported in parenthesis.

Demand-Side Mechanism

Table 6: Changes in Composition of Customers

Dependent variable: Δ Share in Prime Customer and Δ Log HHI of Cutomer Types

	Share Prime Customer				Concentration Customer Types			
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
Adoption	0.00168	-0.00578			-0.0249***	-0.0344*		
	(0.00315)	(0.0074)			(0.00715)	(0.0178)		
Adoption x High			-0.0197***	-0.0258***			-0.0576***	-0.0868***
			(0.00474)	(0.0081)			(0.0132)	(0.0232)
Adoption x Low			0.0236***	0.0174**			0.00477	-0.0207
			(0.00401)	(0.00796)			(0.00637)	(0.0174)
Residual CF				0.0069				0.0307
				(0.00836)				(0.0204)
Sector-zipcd-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dummies first 4 quarters	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-value null equal returns			0.00	0.00			0.00	0.00
Observations	4,610,085	4,610,085	4,610,085	4,610,085	4,610,085	4,610,085	4,610,085	4,610,085

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the establishment level and reported in parenthesis.

Supply-Side Mechanism

Table 8: Distribution of revenues in peak and off-peak time

Dependent variable: Δ Log revenue in peak and off-peak time of the week

	Peak time		Off-peak time			
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Adoption	0.0207	0.032	0.0815***	0.170***	0.0382**	0.0815**
	(0.0284)	(0.0651)	(0.0212)	(0.0543)	(0.0156)	(0.0387)
Sector-zipcd-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Dummies first 4 quarters	Yes	Yes	Yes	Yes	Yes	Yes
Demand Controls					Yes	Yes
Observations	4,610,085	4,610,085	4,610,085	4,610,085	4,610,085	4,610,085

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the establishment level and reported in parenthesis.

Net Welfare vs. Business Stealing

- What do these results mean for impact on net welfare?
  - Business stealing effect: what happens to non-adopters?
  - Are there any adoption impact at the aggregate zip code-sector level?

Dependent variable: Δ Log Revenue

	OLS (1)	OLS (2)	IV (3)	OLS (4)	OLS (5)	IV (6)	OLS (7)
Δ Adoption	0.0430***	0.0397**	0.095466**		0.0418***	0.0966**	
	(0.0158)	(0.0160)	(0.0398)		(0.0160)	(0.0398)	
Δ Adoption by competitor	-0.00423*	-0.0146***	-0.0117**	-0.0135***	-0.0114**	-0.0085	-0.0108**
	(0.002312)	(0.00497)	(0.00536)	(0.00502)	(0.00532)	(0.00571)	(0.00538)
Sector-quarter FE	Yes						
Sector-zipcd-quarter FE		Yes	Yes	Yes	Yes	Yes	Yes
Sector-zipcd Trends	Yes						
Subsector-zipcd Trends		Yes	Yes	Yes	Yes	Yes	Yes
Dummies first 4 quarters	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Drop out adopters				Yes			Yes
Effect only of first adopter					Yes	Yes	Yes
Observations	4,610,085	4,610,085	4,610,085	4,610,085	4,610,085	4,610,085	4,610,085

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the establishment level and reported in parenthesis.

Aggregate Effect

Table 11: Aggregate effect of adoption

Dependent variable: Δ Log Revenue

	OLS	OLS	1st-stage	2nd-stage
	(1)	(2)	(3)	(4)
Δ Adoption	0.0293***	0.0160**		-0.0160
	(0.00823)	(0.00715)		(0.186)
IV			0.0299***	
			(0.00389)	
Sector-quarter FE	Yes	Yes	Yes	Yes
Sector-zip code FE		Yes		
Observations	75,330	75,330	75,330	75,330

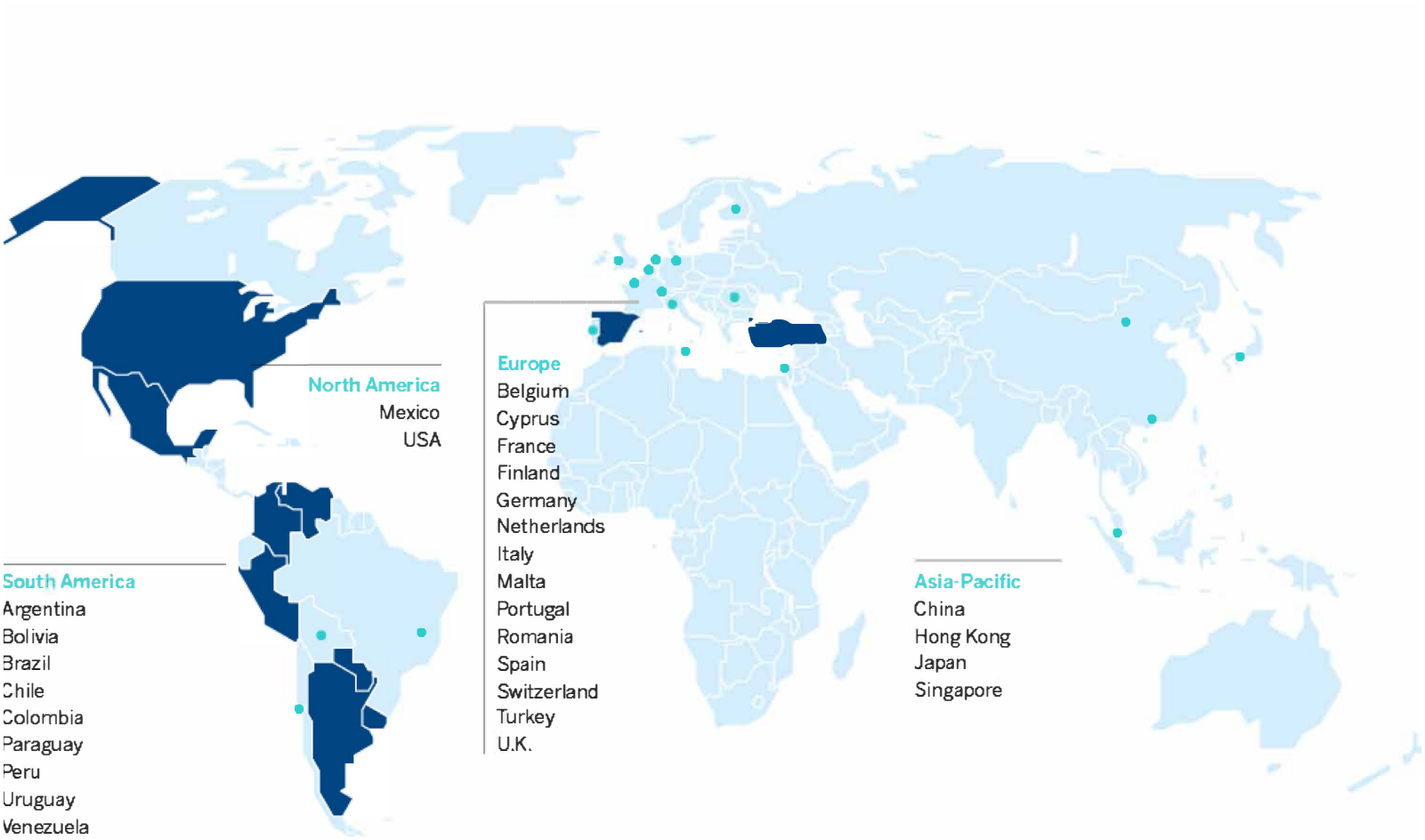
Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the sector-zipcode level and reported in parenthesis.

Conclusions

- We find an impact of adoption on establishment performance, quarterly revenues increase by 9%.
- Heterogeneous effects across establishment size, but not significant differences in sophistication or competition.
- Suggestive evidence that Big Data enables establishments to identify business opportunities and efficiently reshuffle their workload.
- Given positive returns to adoption, adoption costs seem to be driving factor for lack of adoption. Should government intervene and reduce barriers to adoption?

Thank you!

Bank Scale and Scope



€ 699 billion in total assets

78.1 million customers

30 countries

7,744 branches

32,658 ATMs

126,973 employees

Identification Strategy - IV Estimation

Figure 3: Instrumental variable identification

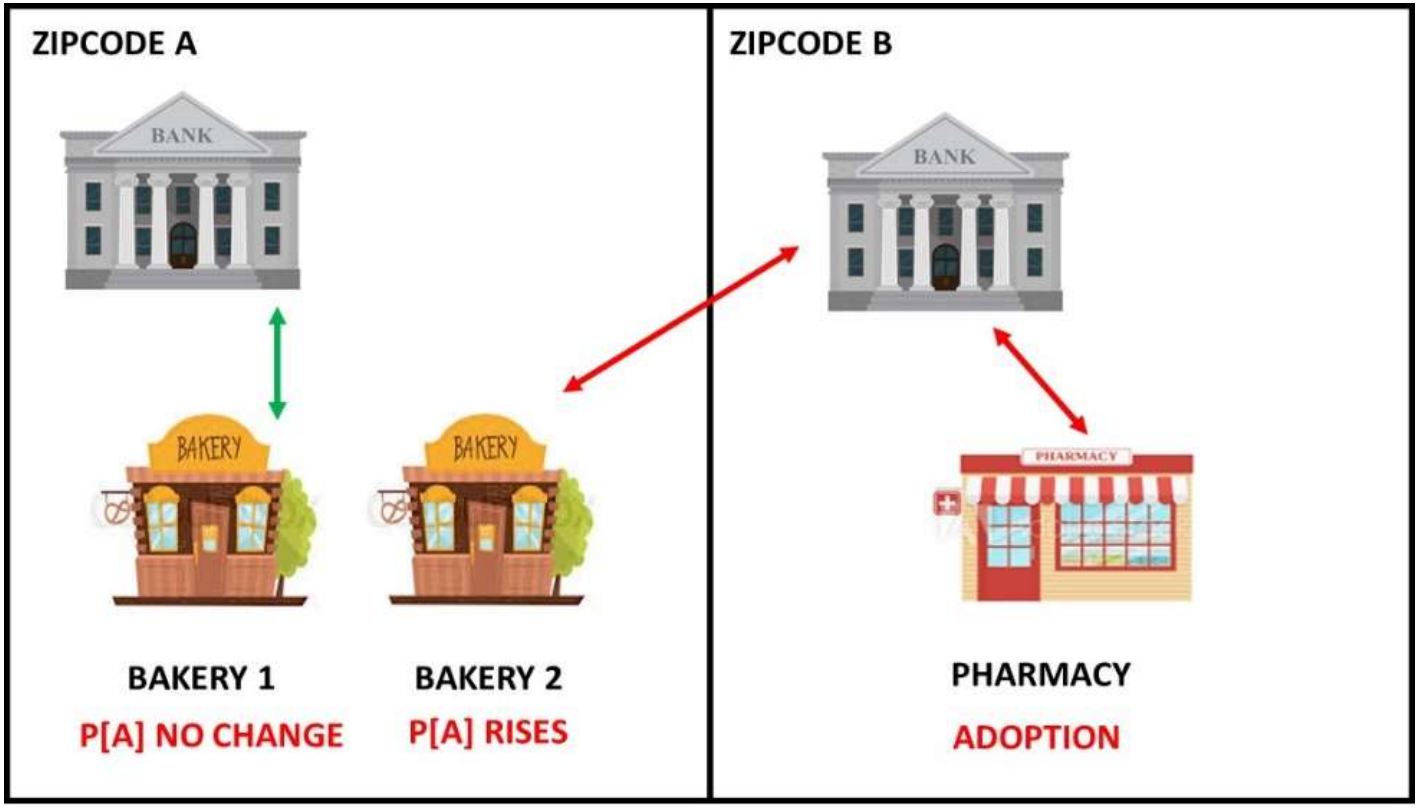


Table A1: Robustness Results

Dependent variable: Δ Log revenue

	OLS	1st-stg	2nd-stg	OLS	1st-stg	2nd-stg
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Adoption	0.0380***		0.106**	0.0577***		0.114**
	(0.0163)		(0.0421)	(0.0165)		(0.0447)
Peers IV		0.00450***			0.00432***	
		(0.00012)			(0.00012)	
Sector-zipcd-quarter FE	Yes	Yes	Yes			
Establishment time trend	Yes	Yes	Yes			
Subsector-zipcd-quarter FE				Yes	Yes	Yes
Dummies first 4 quarters	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,610,085	4,610,085	4,610,085	4,610,085	4,610,085	4,610,085

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the establishment level and reported in parenthesis.

Table A2: IV Robustness Results

Dependent variable: Δ Log revenue

	1st-stg (1)	2nd-stg (2)	OLS (3)	1st-stg (4)	2nd-stg (5)
Δ Adoption		0.0948**	0.0445***		0.0840**
		(0.0387)	(0.0159)		(0.0422)
Peers IV				0.00451***	
				(0.000121)	
Peers IV (no same sector)	0.00448***				
	(0.000119)				
Sector-zipcd-quarter FE	Yes	Yes	Yes	Yes	Yes
Bank-branch time trend			Yes	Yes	Yes
Dummies first 4 quarters	Yes	Yes	Yes	Yes	Yes
Observations	4,610,085	4,610,085	4,610,085	4,610,085	4,610,085

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the establishment level and reported in parenthesis.

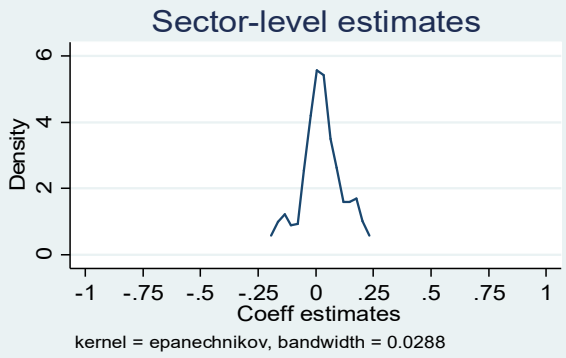
Heterogeneous Effects

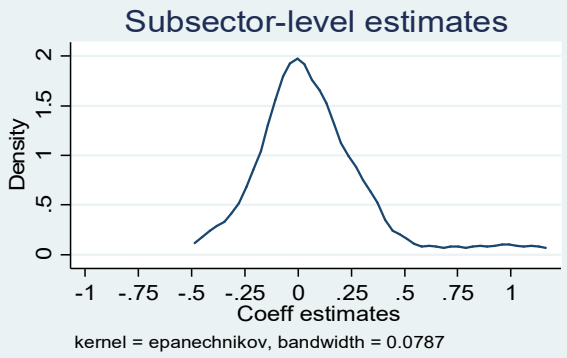
Explore heterogeneity behind main result:

- "Sophistication" of the adopting establishment. Questions
- Size: average quarterly credit card sales prior to adoption.
- Competition: number of other establishments in the same sector-zipcode.

Heterogeneous Effects (Cont.)

Figure 4: Treatment estimates across sectors, subsectors and regions





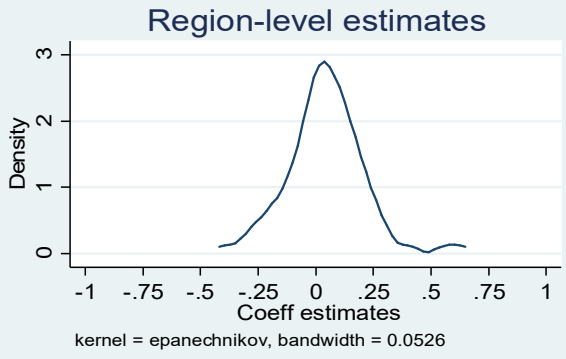


Table 5: Effects on Other Outcomes II

Dependent variable: Δ Log revenue per transaction and Δ log revenue per customers

	Rev/Trans		Rev/Cust	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Δ Adoption	0.00187	-0.0394**	0.00701	-0.0293
	(0.00906)	(0.0187)	(0.00961)	(0.0199)
Sector-zipcd-quarter FE	Yes	Yes	Yes	Yes
Dummies first 4 quarters	Yes	Yes	Yes	Yes
Observations	4,610,085	4,610,085	4,610,085	4,610,085

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the establishment level and reported in parenthesis.

Table 7: Attracting Customers from Other Areas

Dependent variable: Δ Share of revenue from customers from other zipcodes

	OLS (1)	IV (2)	OLS (3)	IV (4)
Δ Adoption	0.00570	0.00929		
	(0.00347)	(0.00583)		
Δ Adoption x Large Share			-0.00467	0.00114
			(0.00334)	(0.00624)
Δ Adoption x Small Share			0.0154***	0.0216***
			(0.00592)	(0.00734)
Residual CF				-0.00676
				(0.00694)
Sector-zipcd-quarter FE	Yes	Yes	Yes	Yes
Dummies first 4 quarters	Yes	Yes	Yes	Yes
p-value null equal returns			0.003	0.002
Observations	4,610,085	4,610,085	4,610,085	4,610,085

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the establishment level and reported in parenthesis.

Table 9: Concentration of Revenues over the week

Dependent variable: Δ Log HHI of revenues over the week

	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Δ Adoption	-0.0152***	-0.0440***	-0.00915*	-0.0312***
	(0.00555)	(0.0127)	(0.00526)	(0.0119)
Sector-zipcd-quarter FE	Yes	Yes	Yes	Yes
Dummies first 4 quarters	Yes	Yes	Yes	Yes
Demand Controls			Yes	Yes
Observations	4,610,085	4,610,085	4,610,085	4,610,085

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the establishment level and reported in parenthesis.

Heterogeneous Effects - Sophistication

Adopters may answer three different questions regarding their analytical, marketing and digital capabilities when registering onto the online platform that will grant them access to the monthly reports:

1. How digital are you? (1) I do not use computers or internet in my daily file; (3) I have personal social media. I use internet daily. I use internet to communicate with my customers/providers; (5) I make internet-based marketing campaigns and analyze the traffic in my webpage.
2. Do you use data for management? (1) I only use intuiti on-driven management practices. I think measuring and analyzing data has no value for my business; (3) I analyze my sales periodically. I read news articles with information about my sector, and think how to apply this to my business; (5) I have a database /CRM with detailed information about my customers. I buy market studies to plan my activity.
3. What is your relation with marketing? (1) I never do marketing campaigns; (3) I make promotions, 2x1, gifts, etc. Sometimes I have made mail campaigns or bought advertising space;(5) I advertise my business in the media (physical advertising, press, or the internet).

back