Information, Beliefs and Resource Allocation in Competitive Markets: The Impact of IT and Big Data on SMEs[1]

Jose Enrique Galdon-Sanchez, UPNA

Ricard Gil, Queen's University

Guillermo Uriz-Uharte, UCL

January 2020

PRELIMINARY AND INCOMPLETE

DO NOT QUOTE WITHOUT PERMISSION

Abstract

Technology may improve communication and coordination of resources within firms, but it may also provide information about the firms' competitive environment. While the existing literature has focused on the former, using large corporations as empirical context, the study of the latter is scant. We fill this gap in the literature by evaluating the impact of a "big data" information technology diffused by a large Spanish bank among its small and medium-size business customers. Using proprietary "big data" on credit card transaction information, we show that technology adoption increases establishment revenue by 9%. These gains in revenue come from the information technology prompting establishments to target existing, yet unexploited, business opportunities. Consistent with this mechanism, we find that adopting establishments increase their sales to underserved customer segments. Not only they increase their number of customers, their new customers also come from underrepresented geographic areas and gender-age groups in their customer portfolio prior to adoption. We cannot reject that adopting establishments do not improve their resource allocation upon technology adoption.

JEL Codes: G20, L20, L80, M15, O32, O33

Keywords: ICT, big data, adoption, retailing, small and medium-size establishments.

[1] We thank Antonio Cabrales, Imran Rasul, James MacKinnon, Riccardo D'Adamo, and seminar participants at UCL, Amazon, EARIE, European University Institute, and APIOC - Tokyo for their comments. We thank Fundacion BBVA for funding this research project. The usual disclaimer applies.

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

The Economist May 6th 2017

1. Introduction

While neoclassical economics implicitly assumes that [perfect] information is widely available to firms and decision makers, the crude reality is that imperfect and asymmetric information is ubiquitous in markets and organizations. In fact, economists have showed that information plays a central role in understanding the development and functioning of a wide variety of contexts such as monetary policy and financial markets (Hayek, 1945; Fama, 1970; Lucas, 1972; Grossman and Stiglitz, 1980), labor and education markets (Stigler, 1962; Spence, 1973), healthcare and insurance markets (Rothschild and Stiglitz, 1976), or product markets where quality and reputation are key determinants of competitive advantage (Akerloff, 1970).

A key mechanism through which information affects the economy is decision-making. Not only consumers make purchasing decisions based on information available to them through advertising and consumer reports, but also information is a key input for firms in their day-to-day production and marketing strategies. More and better information may increase a firm´s productivity in a variety of ways both lowering costs (more efficient resource allocation and improving production processes) and better understanding of business opportunities (product customization and forecasting demand). The rise of information technology in the last few decades has lowered the marginal cost of collecting, processing and using information for decision-making (Brynjolfsson et al., 2011; Brynjolfsson and McElheran, 2016a and 2016b; Agrawal et al, 2018), originating the eruption of the “big data” revolution and data-driven decision making (DDD hereafter) over traditional decision making based on intuition.

However, the access and adoption of big data IT has concentrated in large corporations and has been anecdotal among medium and small size firms.[2] If DDD is the current “best

[2] Brynjolfsson and McHeleran (2016b) show that data-driven decision-making is concentrated in plants with three key advantages: size, high levels of potential complements such as information technology and educated workers, and “awareness.”

practice," a puzzle arises of why don't all firms adopt? A combination of low returns to adoption and high fixed cost of adoption may be deterring small and medium-size firms from adopting and gaining access to "big data". Understanding the determinants of big data IT adoption is important because information access determines firms' competitive advantage. The sparse adoption patterns may widen the performance differences between firms with "intuition-driven" and "data-driven" decision-making practices, and increase further market concentration with all its consequences on market outcomes such as prices, quality, and innovation.

Not surprisingly then, the growing literature on IT and "big data" adoption has mainly focused on large firms because these are more likely to adopt. These firms benefit from these technologies from improving their internal processes (Brynjolfsson and Hitt, 2000; Bresnahan et al., 2002; Bartel et al., 2007) and gaining better access to markets in conjunction with other management practices (McElheran, 2014 and 2015). For this reason, the existing literature presents a gap in understanding both the impact of this technology and the "data-driven" decision-making in small and medium size enterprises (SMEs hereafter). On the one hand, the observed low adoption rates in SMEs may be inefficient if SMEs are overestimating the costs of adoption or are not internalizing potential gains to social welfare from adoption. On the other hand, the scarce adoption rates may be privately efficient if SMEs derive low returns or face high costs of adoption. Then, we may consider whether private adoption decisions are socially optimal. If returns to adoption are low, private and socially optimal technology allocation may not differ much. In contrast, if low adoption rates are mainly caused by high adoption costs (Forman and Goldfarb 2005; Tambe and Hitt 2012; McElheran, 2015 & 2019), private decisions may be socially inefficient opening the door for government intervention to lower such costs of adoption.

To determine efficiency in the allocation of technology to firms in an economy, it is necessary to disentangle adoption decisions from their returns and adoption costs. Only then we may be able to assess whether direct government intervention or intervention through regulation in sharing policies may be welfare improving. To disentangle the two channels, adoption returns vs costs, we would need (1) information about technology adoption in SMEs, and (2) an exogenous shifter of adoption costs. In this paper, we aim to disentangle both channels and estimate the distribution of returns to adoption of "big data" information technology that facilitates the implementation of data-driven decision-making practices.

To do so, we use information from a large bank in Spain about the deployment of a big data information-sharing program among SMEs customers of the bank. Upon voluntarily and freely signing up to the program, SMEs receive a report on their own sales profiles relative to other neighboring establishments in their same sector. Despite an earlier pilot release in 2014, the program was officially launched in the spring of 2016 for the whole country, targeting all establishments with a bank point-of-sale (hereafter POS). We used information on credit and debit card transactions for nearly all POS in the country of study between 2014 and 2018. Our final working data contains quarterly information for 310,610 establishments, out of which 7,110 adopted the technology across all provinces in the country, 17 sectors and 70 subsectors.

Our empirical methodology uses OLS regressions of first-differences of quarterly revenue on first-differences of adoption with sector-zipcode-quarter fixed effects and establishment-specific time trends as baseline specification. We complement this analysis with an instrumental variable approach where we take advantage of the fact that different establishments within the same sector-zipcode dyad are affiliated to different bank branches. Our instrumental variable is then the number of adopters across all sectors, other than the focal establishment, in the establishment's bank branch. The rationale for the instrument comes from detailed conversations with bank managers in that the bank did not compensate its employees for the diffusion of the program, and therefore differences in program diffusion across branches were explained by idiosyncratic preferences and affinity of branch employees with the program.

We find that adoption is associated with 4.5% increase in revenue from credit and debit card transactions, and our instrumental variable strategy shows that adoption causally increases establishment revenue by 9%. Our evidence also points out that the increase in revenue comes from an increase in the number of customers. We investigate these findings further through two potential mechanisms. On the one hand, adoption may prompt establishments to target existing, yet unexploited, business opportunities. On the other hand, adoption may help establishments improving the efficiency of their internal resource allocation. While we cannot reject the latter supply-driven mechanism, we find support for the former demand-driven mechanism. Our evidence shows that adopting establishments increase their sales to underserved customer segments. Not only they increase their number of customers, their new customers also come from underrepresented geographic and gender-age groups in their customer portfolio prior to adoption.

When describing our findings, it is important to highlight the fact that adopters not only change their portfolio of customers when they discover new business opportunities, they choose to broaden their customer base into a more diverse portfolio of customers. While some theories may predict that more information may drive establishments to become more specialized, we find the opposite, that is, establishments with more information start serving more customer types. This finding is important because it has direct consequences for the impact of information on the degree of competition and ultimately consumer surplus and total welfare. If establishments specialized because of receiving more information, the degree of competition would go down, prices would increase and welfare would decrease. Instead, our findings suggest a positive association between more information in a market, the degree of competition and total welfare in a market.

The closest paper to ours is Brynjolfsson and McElheran (2016b) who estimate the impact on productivity in manufacturing establishments due to a switch towards data-driven decision-making (DDD hereafter) through IT investments. Ours differs from their contribution in that *(i)* our retail establishments; and *(ii)* our subjects are SMEs. Because we are able to estimate the causal effect on productivity of big data and DDDs adoption on SMEs, we can contribute to the question of why we observe such scant patterns of ICT adoption in SMEs relative to large firms and corporations. Our paper also contributes to a small literature on the study of the impact of ICTs on SMEs (Angle and Forman, 2018).

While managerial implications of our findings are clear for managers of SMEs, policy implications are ever more relevant. In our setting (an average OECD economy), large firms (more than 50 employees) account only for less than 1% of all firms in the country and 48% of employment whereas SMEs account for more than 50% of employment and almost 99% of firms. These patterns in the size distribution of firms and employment are representative from all industrialized and OECD countries. To the extent that our results provides estimates of the private returns of "big data" IT adoption for SMEs, intervention and government policy aiming to correct for socially efficient adoption is desirable.

The structure of the paper is as follows. Section 2 reviews the literature on IT adoption while providing a theoretical framework for our study. We describe the empirical setting and our data in section 3. Section 4 lays out the methodology and discusses identification. In section 5, we describe our main results and explore mechanisms. Finally, section 6 concludes.

2. Literature Review and Theoretical Framework

While this paper is empirical in nature, our interests and findings contribute to several different streams of literature with solid theoretical foundations. For this reason, we use our literature review in this section to provide a theoretical framework for our empirical strategy and research question.

First and foremost, our paper contributes to the literature that studies persistent performance differences (PPDs hereafter) among otherwise-equal firms within an industry, and its overall consequences for income inequality, growth and competition. As a stepping stone, our analysis has implications for the study of the distribution of establishment-level revenues across retail industries, and the impact of "big data" technology on the distribution of earnings. While traditional explanations for the dispersion in productivity have pointed out competition (Syverson, 2004 and 2011; Hsieh and Klenow, 2009) or search costs (Hortacsu and Syverson, 2004), Gibbons and Henderson (2013) highlight the importance of management practices to explain the observed distribution of PPDs in an economy. Others such as Bloom and Van Reenen (2007) have provided consistent evidence that certain managerial practices are more likely to be associated with high productivity levels and that some of these managerial practices are enabled by the adoption of information and communication technologies (Sadun and Van Reenen, 2005; Bloom et al., 2012). Our paper examines how adoption of "big data" ICTs triggers changes in behavior by small and medium-size establishments, which translates into changes in performance.

Our paper is also relevant for the literature that studies whether more information makes markets more competitive (Rothschild and Stiglitz, 1978; Stiglitz, 1985). While more transparency on the consumer side is associated with more competition, more transparency on the producer side is thought to have opposite effects as it facilitates tacit collusion (Kuhn and Vives, 1995; Stigler, 1964; Tirole, 1988). In a series of papers, Schultz (2004 & 2009) study the impact of transparency on product variety and product differentiation with their obvious implications for competition. Brown and Goolsbee (2002) show that the rise of the internet made insurance markets more competitive. Granados et al. (2006) study how IT adoption affect market information and transparency. Gavazza and Lizzeri (2009) explicitly consider how transparency in political markets shapes economic policy. Other applications of the effect of information and transparency are grade cards in restaurants (Jin and Leslie, 2003) or credit registries (Liberti et al., 2019). A number of papers study the role of information as determinant of the degree of competition between incumbent firms (Pettengill, 1979; Choi et al., 1990; Bertoletti and Poletti, 1997; Huck et al., 2000; Carlin et al., 2012). Our paper studies how establishments change their behavior when they receive information about their own performance relative to their competitors. Although our study does not directly contribute to this literature, our findings directly speak about how markets with more adopters change relative to markets with fewer adopters, and have implications for the understanding of how market dynamics may change when competitors become better informed.

Our project also borrows from the literature that studies how organizations collect, process and communicate information within their boundaries. In our setting, establishments adopt a technology that provides monthly reports about their own sales and the profile of their clientele relative to their direct competitors (other establishments in the same sector and same district). Therefore, a change in behavior post adoption must be triggered not by new information but information that was previously ignored. We link our paper to the literature dealing with inattention in organizations. Cyert and March (1963) highlight the impact of organizational slack, while Cohen et al. (1972) introduce the model of organizations as garbage cans. Most recently, Dessein et al. (2016) and Dessein and Prat (2016) study the impact of rational inattention on organizational focus. Our work relates to theoretical analysis in Matějka (2016) regarding inattentive sellers and price rigidity, as well as Anderson and De Palma (2012) in their study of retailing outlets competing for consumer attention. Our paper findings are consistent with inattention from the retailers' side in their assessment of business opportunities. In the same way, our results are also consistent with a well-established literature in management that studies absorptive capacity (Cohen and Levinthal, 1990). In summary, this literature claims that differences in learning across organizations may be determined by prior investments in their learning capacity.

While there is a large literature on technology adoption, our focus is on information and communication technologies (ICTs hereafter). An early part of this literature has focused on the estimation of the effects of the internet as ICT (Forman et al., 2012). Others have focused on the impact of ICTs in firm organization (Brynjolfsson and Hitt, 2000; Bresnahan et al. 2002; Bloom et al, 2013), R&D investment and innovation (Mohnen et al., 2018; Uriz-Uharte, 2019), and productivity (Hauswald and Marquez, 2003; Sadun and Van Reenen, 2005; Dong et al., 2009; Bloom et al., 2012). Most recently, a number of papers have studied the role of IT in enabling DDDs (Brynjolfsson and McElheran, 2016a and 2016b; Brynjolfsson et al., 2011).

While most of the aforementioned literature has focused on large firms, there is a scant literature studying the impact of ICTs in SMEs (Angle and Forman, 2018; Forth and Bryson, 2018; Viollaz, 2017). Our paper contributes to the ICT literature by focusing in the gains of "big data" ICTs in SMEs. The relevance of our findings hinges in the fact that SMEs in most developed economies account for close to half of employment and around 95% of firms. Therefore, our work furthers the understanding of how "big data" can improve firm productivity in firms where gains from internal organization are negligible.

Finally, as firms and organizations leverage ICTs to take advantage of "big data," there is a recent and growing empirical literature that focus on the study of the impact of ICTs and big data in retail. McAfee and Brynjolfsson (2012) argue that big data enables data-driven decision-making and therefore allows managers to evaluate and measure precisely the impact of their decisions. Einav et al. (2017) assess gains from e-commerce, Farboodi et al. (2019) present data as valuable intangible asset driving the skewness of firm size and productivity distribution, and Bajari et al. (2019) show that "big data" allows firms for lower forecasting errors and therefore better decision making. Goldfarb and Tucker (2019) survey the literature on digital economics and ICTs.

3. Institutional Detail and Data Description

3.1. The Bank

Our empirical setting is the market for SMEs in Spain, and our data come from one of the largest banks in the country. Hereafter, we refer to the data provider as "the bank". The bank is a major player in the credit card market both as credit (and debit) card issuer and credit card POS provider.

Amidst its prevalence and salience in the marketplace, the bank launched a pilot program for its POS clients in one region of the country in the fall of 2014 and went national in the spring of 2016. The program aimed to bring "big data" technology to SMEs using the bank's credit card POS.[3] The bank provided this program free, and adoption was voluntary. It is also important to note that the bank did not compensate its employees for the diffusion of this program. If anything, bank employees would offer the adoption of the program as a source of value added to an already existing business relationship with the client.

To join the program, a POS client would follow a two-step process. First, the client would physically visit a bank branch and meet with a branch employee that would facilitate signing up for the program. Once the client had signed up, the bank would send her an email with setting up information for accessing the incoming monthly reports. Second and last, the client would need to follow the indications in the email received. These instructions would prompt the client to answer a few questions regarding her analytical and marketing savviness. At this point in the process, the newly signed up customer becomes familiar with the online platform that the bank uses to deliver its monthly report. This platform contains different tools and orientation videos to familiarize the client with the report information and therefore maximize the understanding, accessibility and customer experience from this service. Finally, note that when clients sign up for this service online, they must acknowledge a waiver on their liability with the program. Regardless of when a customer signs up for the program, the signee would receive its first report during the first week of the following calendar month.

[3] A Bank manager supervising the program went on public record to describe the program as "This program brings data technology available only for big firms to SMEs. Through this tool, retailers can get to know better their sector and customers. This allows them to improve their decision making."

Upon opting in for this service, the bank generates for each adopter a monthly report, which becomes available through the program's online platform. This report contains summary statistics regarding the number and value of credit card transactions in the previous month. The report disaggregated this information on credit card transactions by client demographic groups such as age, gender and zipcode as well as other classifications such as new vs. returning customers or the time and day of transactions. The report also contained the same set of aggregated information for business competitors in the same zipcode. This set of information on each store's direct competitors provided a reference point and allowed program participants discover differences between their own performance and client portfolio and those of their closest competitors.

To understand further the program, we need to describe the nature of the information used to generate the monthly reports. The reports originated from credit and debit card transactions made by bank-issued cards in all POS in the country (both POS from the bank and from other financial institutions). Because the bank of our study holds a substantial market share in the credit card market in the country, the report information issued by the program and received by the adopters is representative of the population of credit card transactions in the market for both the adopter and her competitors.

3.2. Data Description

Our data is the universe of all transactions from credit cards issued by a large bank (referred as "the bank" throughout the paper) from January 2014 to December 2018. The data is unique in that it details for each transaction establishment-specific and card-specific identifiers. On the one hand, it is important to note that we observe any establishment in the country as long as this establishment has an active POS. The data set also contains information on the establishment location, sector and subsector. On the other hand, the data contains cardholder information at the card level such as age, gender and residence zipcode. A zipcode in our context is equivalent to a 5-digit zipcode in the US.

Overall, the raw data contains transaction-level information for nearly 2.5 million establishments distributed across all provinces, 17 sectors and 70 subsectors. Because of our confidentiality agreement with the bank, we aggregate transaction information at the establishment-quarter level. Additionally, we make two other changes to our initial data set. First, we drop all establishments with less than 5 transactions on average per quarter. Second, we focus our analysis on all establishments in sector-zipcode pairs where we observe, at least, one adopter during our sample period. These changes decrease computational burden while preserving all the within-zipcode-sector variation in technology adoption from the original data. This variation is precisely what will allow us to achieve our goal of estimating the impact of technology adoption at the establishment level.

Our final working data set contains information from a total of 310,610 establishments, including all 7,100 technology adopters in the universe. Figure 1 shows the evolution of the number of adopters from July 2014 to end of 2018. While the bank first launched the technology as a pilot program in a few locations, its official launching took place in mid-2016 where the number of adopters increased rapidly to a level right around 7,100 in late 2018. Table 1 shows that our data set accounts for a total of 4,610,085 establishment-quarter observations. In our sample, the average establishment collects 4,715 Euros per quarter spread across 120 transactions. These distributions are clearly skewed, as the average transaction value is 64 euros. Finally, it is important to note that the average store sells to 74 customers in a quarter and the average value per customer is 85 Euros.

The bottom half of Table 1 describes these variables and other characteristics that we used to explore impact heterogeneity for the subsample of 7,100 adopters. The average adopter collects 6200 Euros per quarter in 153 transactions with an average transaction of 80 Euros. Each adopter serves 92 customers per quarter, each of which spends 102 Euros on average. Finally, adopters have on average of 75 competitors of the same sector in their same zipcode.

Finally, we use the fact that 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.[4] Each one of these questions provide Likert

[4] The three questions and potential answers are as follows. First question: How digital are you? (1) I do not use computers often or internet in my daily file; (2) I have an email account. I use internet to see the news, search for information, etc.; (3) I have personal social media. I use internet daily. I use internet to communicate with my customers/providers; (4) I have social media and business webpage. I have hired a product online at least once. I use internet daily to communicate with my customers/providers; (5) I make internet-based marketing campaigns and analyze the traffic in my webpage. I use online tools for management.
Second question: Do you use data for management? (1) I only use intuition-driven management practices. I think measuring and analyzing data has no value for my business; (2) I think there is a value in data, but I do

scales from 1 to 5. We create a measure of analytical savviness by averaging all three answers of all adopters who answer all three questions. Not all adopters respond to this questionnaire. In fact, only 3,495 adopters out of the total 7,100 responded (49.2%). The average sophistication score following this measure is 3.53, with a median of 3.67 and a standard deviation of 0.89. Once we have described our data, we proceed to present our empirical methodology in the following section.

4. Empirical Methodology and Identification

4.1. Baseline Regressions

Our baseline specification is such that,

$$Y_{isjt} = \mu + \beta Adoption_{isjt} + \gamma X_{isjt} + \alpha_i + \theta_{sjt} + u_{isjt} \qquad (1)$$

where Y_{isjt} is the log of the outcome variable such as number of transactions, revenues, or number of new customers for establishment i in sector s located in zipcode j and quarter t. Our main variable of interest is $Adoption_{isjt}$, which is a dummy variable that takes value 1 if establishment i has adopted the technology before quarter t. This variable varies within establishment over time for adopters, and remains at 0 for non-adopters. See Figure 2 for a representation of the timeline between the time an establishment signs up, the delivery of its first report and our variable $Adoption_{isjt}$ taking value 1. In this example, the establishment signs up in the middle of the second quarter (month 4) and only starts receiving a report on May 1st. Our adoption variable takes value 1 in the quarter following adoption and all quarters after that.

---

not know where to find data or what I could use it for; (3) I analyze my sales periodically. I read news articles with information about my sector, and think how to apply this to my business; (4) I use measure my sales and analyze the data in order to improve. I have a database with my customers' contact. I search on the internet information about my sector; (5) I have a database /CRM with detailed information about my customers, and I use this to make promotions. I analyze my sales margins by product. I buy market studies to plan my activity. Third question: What is your relation with marketing? (1) I never do marketing campaigns; (2) I take care of my shop window and my service to attract and increase customer loyalty, but I never do marketing campaigns out of my establishment; (3) I make promotions, 2x1, gifts, etc. Sometimes I have made mail campaigns or bought advertising space; (4) I frequently make marketing campaigns, advertising and discounts. I use email and social media to cultivate customer relations; (5) I have a marketing plan in which I design campaigns and events. I inform my clients about customer-specific promotions. I count with a loyalty program. I advertise my business in the media (physical advertising, press, or the internet).

Our regressions specification also includes time-varying controls X_{isjt} such as dummies for the first four quarters an establishment enters our sample as well as establishment fixed effects α_i and sector-zipcode-quarter-specific fixed effects θ_{sjt}. Finally, u_{isjt} is our residual.

Our working specification will take first differences from specification (1) above,

$$\Delta Y_{isjt} = \beta \Delta Adoption_{isjt} + \gamma \Delta X_{isjt} + \theta_{sjt} + \Delta u_{isjt} \qquad (2)$$

where ΔY_{isjt} is first differences in our dependent variable, and $\Delta Adoption_{isjt}$ is first differences in technology adoption. It is important to note that $\Delta Adoption_{isjt}$ takes value 1 in the quarter right after adoption and value 0 in all other quarters. This specification in first-differences also contains controls X_{isjt} such as dummies for the first four quarters an establishment enters our sample, sector-zipcode-quarter-specific fixed effects θ_{sjt}, and a residual Δu_{isjt}.

Before coping with endogeneity concerns in the next subsection, we argue here that estimating our parameter of interest β with first-differences partially addresses issues of autocorrelation in the error term. Moreover, this regression specification relaxes the requirement of strict exogeneity in the regressors only requiring weak exogeneity for the consistency of estimates.

4.2. Instrumental Variables and Identification

A pervasive concern in the technology adoption literature, and elsewhere in the empirical economics, is the endogeneity and self-selection of establishments into adoption of a technology. In our context, this concern is problematic if the establishment-specific idiosyncratic error terms are correlated with adoption. Examples of such instances would include differences in trends across adopters and non-adopters within a sector-zipcode pair, and sporadic episodes of positive or negative growth that coincide with the timing of adoption. In these cases, the first-differences regression specification (2) with OLS will erroneously attribute changes in productivity to technology adoption.

To address this problem, we look for changes in an establishment's environment that may exogenously change the probability of adoption across establishment within sector-zipcode-quarter triads while being orthogonal to establishment-specific productivity and demand shocks. With this goal in mind, we derive an instrumental variable strategy that exploits the fact that different establishments in the same sector-zipcode dyad may hold their corporate bank account in different bank branches located in different zipcodes. Hereafter, we call the bank branch where an establishment has its corporate bank account the establishment branch.

Our instrument is the number of adopters per quarter (across sectors and zipcodes) other than the focal establishment in the establishment branch. Figure 3 sheds light on the rationale behind our instrumental variable. Assume two zipcodes, A and B. Each zipcode has a bank branch. There are two bakeries in zipcode A (bakery 1 and bakery 2) and one pharmacy in zipcode B. Our instrument highlights the variation in establishment branch for each of the establishments' location. While bakery 1 located in zipcode A uses the bank branch in zipcode A, bakery 2 also located in zipcode A uses the bank branch in zipcode B. The pharmacy in zipcode B uses the bank branch in zipcode B.

Our identification strategy posits that the number of adopters at the establishment branch (as opposed to the branch in the same zipcode of the focal establishment) increases the probability of adoption. In our example of Figure 3, the pharmacy adopts the technology and that increases the probability of adoption of bakery 2 because they share the same establishment branch. In contrast, the probability of adoption of bakery 1 does not change due to the pharmacy's adoption despite being in the same sector and zipcode as bakery 2 because bakery 1 does not share establishment branch with the pharmacy. Therefore, our instrument provides variation in the probability of adoption across establishment in the same sector-zipcode dyad.

Our conversations with bank managers provide a strong foundation for our instrumental variable strategy. As explained in our institutional detail section, the bank did not compensate its employees for the diffusion and adoption of this technology. If anything, HQ paid for brochures and advertising boards and distributed them equally among bank branches. The variation in adoption across branches was rooted in the affinity of their employees with the program. The larger the affinity of an employee, the higher the level of her promotional effort despite not being compensated for it.

In other words, and through the lens of our example in Figure 3, the increase in the probability of adoption of bakery 2 may come from two different channels. On the one hand, branch employees in zipcode B may exert larger promotional effort on the diffusion of the program, and therefore increase the probability of adoption of bakery 2. On the other hand, the pharmacy's adoption also increases the probability of adoption of bakery 2 through peer effects at the establishment branch level. In our empirical application, we do not observe promotional effort of the program at the branch employee level. Therefore, our instrument relies on variation across bank branches in the number of adopters over time.

Reached this point, our identification strategy needs to address the validity of our exclusion restriction. Our strategy exploits differences in probability of adoption across establishments within the same sector-business-quarter triad, which in fact takes into account all sector-zipcode-quarter level productivity and demand shocks. Then, our exclusion restriction assumption would fail if a correlation exists between establishment-specific shocks and promotional effort of the establishment branch within a quarter. Equivalently, heterogeneous trends in performance within sector-zipcode across different establishments affiliated to different establishment branches would also violate our exclusion restriction. Note that we include bank-branch time trends to control for this possible concern in a robustness specification in Appendix Table A2.

Moreover, our identification strategy does not rest on the assumption that different establishments within a sector-zipcode dyad with different establishment branches are alike. Even if there is self-selection of establishments into different branches of different characteristics (perhaps located in different zipcodes), our identification strategy exploits differences in promotional effort of the program over time within branch and mostly relies on the timing of promotional effort being orthogonal to the timing of program introduction to market.

Note that even if there exists peer-effects between establishments of a same sector-zipcode dyad that do not share establishment branch, this alternative mechanism would work against the variation provided by our instrument and utilized by our identification strategy. Nevertheless, note that (1) the introduction of sector-zipcode-quarter perfectly captures this type of peer effects between establishments in the same sector and zipcode, and (2) this second order effects should not be a concern for our exclusion restriction.

A final and necessary exclusion restriction for the plausibility of our instrumental variable is that sharing the same establishment branch only affects the probability of adoption, but it does not directly affect performance. We produce evidence in Table A2 in the Appendix as robustness check where the instrumental variable (IV hereafter) does not include peers in the same sector. Finally, it is paramount to emphasize the fact that the bank did not introduce any other program with [partially or fully] overlapping characteristics during our sample period.

5. Results and Mechanisms

We describe the results of our empirical analysis in three different steps. First, we show our main results of running regression specification (2) and follow up with exploring heterogeneity in the impact of adoption of the technology. Second, we continue our analysis by investigating mechanisms behind the main results. Third, we conclude this section with a discussion of the results while linking back to existing literature.

5.1. Main Results and Heterogeneity

Table 2 shows the results of running our baseline specification where the dependent variable is the log of quarterly credit card revenue. Columns 1-3 run OLS regressions in first-differences under alternative deviations of the baseline specification. Column 1 shows that adoption is associated with an increase of 4.6% in revenue. Columns 2 and 3 are the result of running leads and lags dummies of the adoption quarter. On the one hand, column 2 shows that the increase in revenues is concentrated in the quarter after adoption and we do not observe further increases in subsequent quarters. On the other hand, column 3 runs a placebo test where we observe that there are no increases in revenue preceding the quarter of adoption.

The last two columns of Table 2 implement our IV strategy. Column 4 shows estimates of the first stage results of column 5. Column 5 shows the results of running instrumental variables on the baseline specification of Column 1. We find that the effect jumps from 4.6%

to 9.0%. We carry out the Hausman test to check for endogeneity, and we cannot reject the null hypothesis that adoption is exogenous (p-value=0.24).[5]

Following Forman et al. (2012) and Uriz-Uharte (2019), we believe the estimate magnitude is larger than in the baseline regressions due to the existence of heterogeneous returns to technology adoption. If bank branches with customer establishments with higher potential returns of technology adoption made more promotional effort, then we are likely to observe a jump in their estimate of returns from adoption when applying our instrumental variable strategy. In other words, the local average treatment effect may be larger than the average treatment effect. This implies that although the instrument affects revenue only through its impact on technology adoption, the returns to technology adoption are larger for those establishments whose adoption decisions is most strongly affected by our instrument.

Once we have determined that technology adoption causally increases establishment revenue by 9%, we investigate the presence of heterogeneity in this effect. We explore heterogeneous effects in two different ways. First, we investigate heterogeneous effects across sectors, subsectors and geographical regions. We plot the distribution of effects across these three dimensions in Figure 4. Note that all three distributions of effects are centered around zero, and that the heterogeneity across sectors shows the lowest variance with range between -0.25 and +0.25. The distribution with largest variance is across subsectors ranging from -0.5 to 1, and the distribution across regions is in between those as it reflects different distributions of sector and subsectors across regions.[6]

Second, we investigate heterogeneous effects across different establishment characteristics. For this purpose, we split our sample of adopters into three different dimensions: analytical savviness of the adopter, establishment size prior to adoption, and degree of local market competition. We report our heterogeneity results for both OLS and IV control function in

[5] While Table 2 presents our baseline results, Table A1 includes regression specifications with establishment-specific time trends and subsector-zipcode-date FE. All our findings are robust to changes in the specification.
[6] Retail sectors benefitting more from adoption are technologies, home wellness and beauty, or accommodation. Retial sectors benefitting less are sports and toys, or supermarkets. A closer look into subsectors shows positive returns of adoption (other than the above mentioned sectors) for tobacco stores, car rental shops, musical instruments, photography, fast-food restaurants, gardening and floristry. Subsectors with negative returns are pubs and discos, press, optician shops and gas stations. So far as geographical regions are concerned, those with a higher number of inhabitants (and adopters) make up for most of the centered distributions of returns around 5-8%, while positive and negative outliers correspond to small regions.

Table 3. Columns 1 and 2 investigate how analytical savviness drives the impact of technology adoption. For this matter, we use our measure of sophistication coming from the adopters answering 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. Hereafter and for simplicity, we call this variable level of sophistication and we create dummies for adopters above and below the median level of sophistication. Column 1 runs OLS first-difference regressions and shows that adoption is associated with increases of 4.4% and 4.6% in revenue for adopters above and below the median level of sophistication, respectively. Column 2 applies our IV control function approach and shows that the returns are now 8.7% and 9.7% for adopters above and below the median level of sophistication, respectively. Note that Table 3 reports the p-value of the test for equal returns for both firm types. According to the reported p-values of 0.95 and 0.75, we cannot reject that these rates of return are statistically the same. This suggests the role of cognitive capacity does not seem to drive returns to technology adoption.[7]

Next, we explore how establishment size correlates with the impact of technology adoption. We measure size by the average quarterly revenue of an establishment in all observed quarters prior to adoption. We then create a dummy variable "*Large*" that gives value 1 to an establishment if its size is above the median size of adopters in the same sector. We also create a dummy variable "*Small*" that gives value 1 to an establishment if its size is below the median size of adopters in the same sector. Column 3 shows that the impact of technology adoption in large establishments is not statistically different from zero, and it is 7.96% in small establishments. When applying our instrumental variable strategy, the estimate for large establishments becomes statistically significant at 7.3% and the estimate for small establishments increases to 14.6%. These findings point out that the returns to access to this technology vary greatly with establishment size. We are able to reject that these returns are the same, and so safely conclude that smaller establishments benefit more from technology adoption.

[7] This is consistent with the fact that the program provides, processes and analyzes data for the adopter. The information is easy to understand and that may be consistent with the fact that we observe an impact even for less sophisticated adopters. This finding is important when considering policy implications regarding access to big data IT technology of less sophisticated and smaller establishments.

Finally, Columns 5 and 6 explore the heterogeneity of the results along the dimension of degree of local market competition. For this purpose, we calculate the average number of competitors in the same sector and zipcode for each adopting establishment over the sample period. The number of competitors averages 74 with a median of 45 and a standard deviation of 90 (highly skewed distribution ranging from three to 967)**.** Once again, we create dummies that divide the adopters into those above and below the median number of competitors. Results in column 5 show that the association between adoption and revenue increases is statistically significant for establishments in highly competitive markets, and it is not in less competitive environments. Column 6 reports our IV results and shows that adoption increases revenue by 11% in more competitive markets. These results across more and less competitive sector-zipcode dyads are statistically different from each other at the 11% level.

In summary, our heterogeneity results are insightful in depicting scenarios where technology adoption derives in higher returns. Our findings in Table 3 show that those establishments of smaller size and those operating in more competitive markets derive higher returns from adoption. Sophistication and digital experience do not seem to matter for the returns to technology adoption in our context.

5.2. Mechanisms

Our findings in the previous section establish that technology adoption increases establishment revenue by 9%. Moreover, we also find that this effect is heterogeneous. In fact, smaller establishments and in more competitive markets seem to benefit more from adoption. In this subsection, we aim to understand the mechanisms behind our findings.

In our empirical setting, establishments adopt a technology that provides information on their performance relative to others in their local market. This new information may have two types of direct effects that we define as two distinct mechanisms. On the one hand, the report received may highlight business opportunities that the establishment was not aware of or did not paid much attention to in the past. The receipt and processing of this information may drive an adopter to serve different customer profiles, that is, different age-gender groups, nearby zipcodes, or customers that purchase their goods and services during different times of the week. On the other hand, the information provided by the report may trigger adopters to reallocate their resources more efficiently towards customer groups and times during the week where their marginal returns to effort are higher. While the former mechanism requires exploiting new business opportunities, the latter implies reallocating existing levels of effort and resources. Hereafter and for simplicity, we call the latter "demand-driven" mechanisms and the former "supply-driven" mechanisms.

5.2.1. Demand-Driven Mechanisms

We start our analysis of mechanisms by investigating how the increase in revenue relates to the number of transactions and customers. Table 4 shows results using three different dependent variables. While Column 1 shows that the adoption of technology is associated with an increase of 4.4% in the number of transactions, Column 2 uses our IV strategy and reports that the causal effect of technology adoption in the number of transactions is a 13% increase. Parallelly, Columns 3 and 4 investigate whether the increase in the number of transactions comes paired with an increase in the number of customers and we find that adoption technology increases the number of customers by 12%. Consistently with these two results, the number of transactions per customer experiences a marginal increase of 1% (not statistically significant) as shown in Column 6.

Table 5 turns to the study of the impact of technology adoption on the average transaction value and the average revenue per customer. Columns 1 to 4 use OLS first-differences and IV regressions to estimate the effect of technology adoption on these variables. We find that, consistently with our results on the impact on revenue and number of transactions and customers, technology adoption decreases the average transaction value by 3.9% and has no statistically significant effect on the revenue per customer. These findings suggest that the new customers spend less money (and in smaller transactions) than the average customer prior to adoption.

Once we have established that technology adoption facilitates the discovery of new business opportunities through increases in the number of customers and the number of transactions, we continue our analysis by examining whether the average demographic profile of the customers of an establishment changes upon adoption. Because we show in Tables 4 and 5 that there are differences between the average customer pre-adoption and average customer post adoption, we now examine changes in the customer profile by age, gender and zipcode.

Importantly for us, the report identifies customers according to two gender groups (male/female), six age groups (<25, 25-34, 35-44, 45-54, 55-64, 65>), and customer dwelling zipcode. Moreover, the report highlights the most important customer profile of the store and the sector-zipcode independently. This allows establishments to identify discrepancies between their main customer type and the main customer type of their closest competitors, other establishments in the same sector and in the same local market.

For this purpose, we identify the main customer type (one of the 12 gender-age groups described above) for each sector-province dyad and calculate the share of revenues from each establishment's main customer type according to their sector-province dyad. Then we create a dummy variable "*Large Share*" that equals 1 if the share of revenues from the main customer type is above the median among all adopters, and 0 otherwise. The dummy variable "*Small Share*" gives value 1 to an adopter if the share of revenues from the main customer type is below the median among all adopters, and 0 otherwise. Columns 1 and 2 in Table 6 show that technology adoption does not significantly change the share of revenues from the main customer type. Yet, in Columns 3 and 4 we investigate whether the no-effect is a true no-effect or suffers from compositional issues. Indeed, findings in Columns 3 and 4 show that those establishments with larger shares of main customer type pre-adoption are likely to decrease the share of revenues from main customer type upon adoption. Conversely, establishments with smaller shares increase their share of revenues from the main customer type.

In Appendix Table A3, we investigate whether our findings on changes in the share of revenue from the main customer type are driven by the numerator (more or less sales to this customer type) or the denominator (more or less sales to other customer types and in total). Our results show that findings in Columns 1 to 4 of Table 6 are driven by: (1) establishments with small share of sales to the prime customer group increasing their sales to this group; and (2) establishments with a high share of sales to the prime customer group decreasing the share of their sales to this group as a result of selling more to other groups but not reducing their sales to prime group.

Columns 5 to 8 examine how the diversity of the customer profile per establishment changes with adoption. Our dependent variable uses information of the shares of revenue per each of the 12 age-gender groups in each establishment and computes a Herfindahl-Hirschman Index (HHI hereafter) of customer diversity. Our HHI measure would take value 1 if an establishment sold 100% of their goods and services to only one of the 12 groups, and would take value 0.083 if it sold equally to all 12 age-gender groups. Columns 5 and 6 show that technology adoption decreases the concentration of sales by 3.4%.

Columns 7 and 8 explore whether the decrease in concentration comes from establishments with high or low degrees of concentration pre adoption. For this purpose, we compute the HHI of customer type concentration for each adopter pre adoption. Then we create a dummy variable "*High Concentration*" that equals 1 if the HHI of the establishment is above the median among all adopters, and 0 otherwise. The dummy variable "*Low Concentration*" gives value 1 to an adopter if its HHI is below the median among all adopters, and 0 otherwise. Results in Columns 7 and 8 show that the decrease in concentration in Columns 5 and 6 is entirely coming from establishments with high concentration rates. Those establishments in the upper half of the concentration distribution decrease concentration by 8.7% upon technology adoption.

Finally, Table 7 provides evidence of whether adopters change the spatial composition of their customer base. For simplicity, we compute for each establishment the share of revenue from customers from other zipcodes. Columns 1 and 2 show that technology adoption does not seem to have an effect on the average share of revenue from customers in other zipcodes. Building from this finding, Columns 3 and 4 decompose the main effect into adopters with large and small shares of revenue from customers in other zipcodes pre adoption. We create a dummy variable "*Large Share*" that equals 1 if the revenue share coming from customers in other zipcodes is above the median among all adopters, and 0 otherwise. The dummy variable "*Small Share*" gives value 1 to an adopter if its revenue share coming from customers in other zipcodes is below the median among all adopters, and 0 otherwise. Our results provide evidence that the increase in revenue from customers in other sectors is concentrated in establishments with low share of such type of customer pre adoption.

As far as the demand-driven mechanism is concerned, evidence in Tables 6 and 7 is consistent with a mechanism where establishments discover new business opportunities and implement new marketing strategies to take advantage of the new (to them) information. Our findings show that the increase in revenue comes as a direct consequence of establishments expanding their customer portfolio in a variety of ways. Adopters do not just increase their number of customers, but they target new age-gender profiles and look for customers beyond their zipcode.

5.2.2. Supply-Driven Mechanisms

Alternatively, we consider a second mechanism for the impact of the newly revealed information contained in the report received by adopters. For simplicity, we call it "supply-driven" mechanisms. In addition to the discovery of new business opportunities, establishments may also learn that competing establishments organize their sales in different days of the week and times of the day. In some cases, a reorganization of their time schedule during the week may help establishment managers to improve the logistical efficiency of their allocation of resources such as personnel, time and effort. Upon receiving information from the monthly report, an establishment may reallocate clients and sales to different parts of the week (days or hours), improving the distribution of workload during the week.

Following this logic, we study changes in the distribution of revenue across different days and hours upon technology adoption. To do so, we divide the week in 4 time slots, namely, weekday morning (until 3 pm), weekend morning, weekday evening (after 3 pm) and weekend evening. We identify the peak shopping time for each sector-province dyad and calculate the share of revenues from each establishment's peak shopping time according to their sector-province dyad. Table 8 regresses the log of revenues at the peak time of the week and off-peak (the other three time slots) on adoption. Columns 1 to 4 show no change in the sales at peak time and an increase in the sales off-peak time. This shift of business hours could be driven by *(i)* a supply-side gain in efficiency, or *(ii)* shifting business to serve new demographics with different shopping schedules. To distinguish between these two explanations, Columns 5 and 6 control for changes in demographics of the clientele. Once we control for changes in demand demographics (log of sales for each of the 12 age-gender customer categories and log of sales for out-of-zipcode customers), the magnitude of the effect goes down from 17% to 8%. While not shown here, this finding is robust to controlling for changes in the HHI of customer types, the share of out-of-zipcode sales, or total sales. These findings suggest that technology adoption triggers changes in business hours not explained by changes in demographics and therefore may be due to improvements in supply-side efficiency.

An alternative way to investigate this same issue is parallel to our analysis in the customer portfolio above and implies the use of the concentration measure HHI for the distribution of revenues among all four time slots. Our HHI measure would take value 1 if an establishment sold 100% of their goods and services during only one of the four time slots, and would take value 0.25 if it sold equally in all four time slots. Columns 1 and 2 of Table 9 show that technology adoption decreases concentration by 4.4%. When controlling for changes in demand demographics in Columns 3 and 4, the magnitude decreases slightly to 3.1%. Note that these results are consistent with our findings in Table 8 above. Technology adoption both discovers business opportunities and improves logistical efficiency in adopting establishments.

Before concluding this section, we want to note that it is empirically challenging to separate reshuffling resources across time slots during the week from the discovery of new business opportunities as these may come hand-in-hand. We attempt to disentangle these two channels with a different set of empirical evidence that aims to estimate whether establishments reshuffle resources across different time slots while holding constant their customer demographic portfolio.

In summary, this section investigates the role of demand-driven and supply-driven mechanisms. On the one hand, we find compelling evidence consistent with the existence of a demand-driven mechanism, that is, technology adopters are able to identify new business opportunities and tilt their customer portfolio in response to the monthly information received. On the other hand, we also find evidence that increases in sales due to technology adoption are also coming from improving processes and workload distribution. We discuss our results in the following section.

5.3. Discussion

We qualify our findings in two ways. First, we place our findings within the existing ICT adoption literature. Because of lack of adoption in SMEs (or lack of comprehensive data), the previous literature has focused on large corporations and emphasized the role of ICTs in improving coordination among employees, departments, and divisions within their organizations (Brynjolfsson and Hitt, 2000; Bresnahan et al., 2002; Brynjolfsson et al., 2011; Bloom et al., 2014). Our paper differs from the existing literature in that diverges attention to SMEs and estimates the returns to "big data" ICT to shed light on the puzzle of why are SMEs not investing in these technologies. Our findings show large heterogeneity in the effect of adoption and an average increase in revenues of 9%. Most importantly, our study of mechanisms shows that the increase in revenues is driven by both an improvement of marketing strategies and a more efficient internal organization. Our findings are indicative that if SMEs with high returns of adoption are not adopting due to high adoption costs, intervention may be justified by either providing the technology from government sources or allowing businesses to share information.

Second, we must wonder if the increase in revenues due to technology adoption comes from business stealing (potentially from non-adopters) or is net value generated from better service. For this purpose, Table 10 investigates the effect of adopters on non-adopters' revenues. We first define non-adopter competitors as the rest of the establishments in the same zipcode-sector dyad. Column 1 estimates the impact of adoption on adopters and non-adopters including sector-quarter fixed effects (we cannot control for zipcode-sector-quarter fixed effects given our definition of non-adopter). Column 1 also include sector-zipcode trend and trend squared. In Column 2, we define competitors as establishments in the same subsector-zipcode so that we can introduce zipcode-sector-quarter fixed effects but also subsector-zipcode specific trend and trend squared. We note that adoption is associated with decreases in revenues of non-adopters and that the impact is stronger in closer competitors – those in your sector (1.5% decrease when subsector competitor adopts the technology). When we instrument for adoption, the impact on non-adopters remains qualitatively unchanged. Column 4 runs the same specification of Column 2 dropping all adopters and so comparing performance of non-adopters before and after adoption in their sector-quarter FE and finding the same exact finding of a drop of revenue of 1.4% upon adoption of a competitor. Finally, Columns 5 to 7 account for a different definition of adoption where it only has an impact on non-adopters the first time it occurs within a sector and zipcode. Our OLS findings are robust to this definition change, while our result when implementing IV becomes statistically non-significant (although still negative and close to 1%).

These findings seem to suggest that some of the gains in revenues following adoption are coming from business stealing effects from competitors. Therefore, we cannot reject that adoption has no effect in total welfare. Even though we cannot use the findings in Table 10 to conclude whether this technology is welfare improving, we must consider the fact that consumers switching from one establishment to another must be enjoying net increases in their utility by revealed preference. If so, switching behavior should be an indication of welfare-improving technology.[8] Additionally, our findings also suggest that adopters become more efficient which translates into lower costs for the same level of revenue and surplus generated. If so, welfare gains from widespread technology adoption may come from efficiency gains and not so much sales and consumer surplus.

6. Conclusions

This paper evaluates the impact of the adoption of ICTs and "big data" in SMEs. In our empirical context, small and medium-size establishments were invited by their credit card POS provider to register free of charge in a program that would deliver monthly reports about their performance and their competitors' performance as well as demographic and geographic characteristics of their customers and those of the customers of their competitors. Using first-difference regressions and an IV strategy, we estimate a causal impact of adoption on revenues of 9%. We explore whether these effects are heterogeneous and find that smaller adopters and adopters in more competitive markets benefit more from technology adoption. We find no differences across the level of sophistication and digital experience.

We also investigate mechanisms through which new or better structured information delivered by the monthly reports may have triggered the observed increase in revenues. We

[8] Switching behavior would be associated with welfare decreases in extraordinary cases such as (1) firms with lower marginal costs are losing market share (probably unusual in retail); or (2) there is firm exit combined with an increase in competition where small establishments are gaining more than mid and big–size establishments (rather implausible).

find that adopting establishments increase their revenues from both targeting underserved market segments and reshuffling resources and effort to off-peak times that were underutilized (prior to adoption).

Our findings have managerial and policy implications for the understanding of adoption and economic impact of new technologies. Departing from the existence of PPDs coupled with increases in market power of large firms (de Loecker and Eckout, 2018) and decreases in business dynamism (Akcigit, 2019), it is important to understand how the arrival of the new Big Data IT revolution can affect these trends. The adoption of first-generation IT was mainly concentrated among large firms contributing to increase the gap between large and small firms. However, these patterns of adoption could be expected as these technologies were mainly intended to improve internal coordination and these gains are lower in small firms. By contrast, second-generation IT not only focuses on offering firms opportunities for better internal organization, but also offers them information about their competitive environment (consumers' preferences and/or competitors' actions). Thus, there is a large scope for small firms to benefit from this new generation of Big Data IT. However, if high adoption costs prevent the adoption of Big Data IT by small firms it is likely the case that the distance between large and small firms will grow even larger. As a result, private adoption decisions may be socially inefficient opening the door for government intervention to mitigate adoption costs.

While our evidence suggests a sizable average return on adoption and heterogeneity across establishments of different sizes, a cautious interpretation of our findings calls for an estimate of the lower bound of the cost of adoption. Future research should investigate the nature of these costs. It is important to understand whether the scarce adoption patterns observed in SMEs can be addressed with mere awareness campaigns, provision of other technologies that exhibit complementarities with the current "big data" technological wave or lack of skilled human capital able to operate such technology and process its information to be used as valid input in decision-making.

References

Akcigit, Ufuk, and Sina T. Ates (2019). "What Happened to U.S. Business Dynamism?" NBER Working Paper #25756.

Akerlof, George (1970). "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84: 488-500.

Anderson, Simon P., and André De Palma. "Competition for Attention in the Information (Overload) Age." *The RAND Journal of Economics*, vol. 43, no. 1, 2012, pp. 1–25.

Angle, Patricia, and Chris Forman (2018). "Does IT Level the Playing Field for Small Establishments? Evidence from Manufacturing." SSRN Electronic Journal.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb (2018). Prediction Machines: The Simple Economics of Artificial Intelligence. Harvard Business School Press.

Bajari, Patrick, Victor Chernozhukov, Ali Hortaçsu and Junichi Suzuki (2019). "The Impact of Big Data on Firm Performance: An Empirical Investigation," *AEA Papers and Proceedings*, vol 109, pp. 33-37.

Bartel, Ann, Casey Ichniowski, and Kathryn Shaw (2007). "How Does Information Technology Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement, and Worker Skills." *Quarterly Journal of Economics* 122 (4): 1721–58.

Bertoletti, Paolo, and Clara Poletti (1997). "X-Inefficiency, Competition and Market Information." *The Journal of Industrial Economics*, 45(4): 359–375.

Bloom, Nicholas, Luis Garicano, Raffaella Sadun, and John Van Reenen. (2014). "The distinct effects of Information Technology and Communication Technology on firm organization," *Management Science*, vol. 60, no. 12, Pages 2859-2885.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. (2012). "Americans do I.T. Better: US Multinationals and the Productivity Miracle," *American Economic Review*, 102 (1), 167-201.

Bloom, Nicholas, and John Van Reenen (2007). "Measuring and Explaining Management Practices across Firms and Countries," *Quarterly Journal of Economics* 122(4): 1351-1408.

Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt (2002). “Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence.” *Quarterly Journal of Economics* 117 (1): 339–76.

Brynjolfsson, Erik, and Lorin M. Hitt (2000). “Beyond Computation: Information Technology, Organizational Transformation and Business Performance.” *Journal of Economic Perspectives* 14 (4): 23–48.

Brynjolfsson, Erik, Lorin M. Hitt, and Heekyung Hellen Kim. (2011). “Strength in Numbers: How Does Data-Driven Decision Making Affect Firm Performance?” Unpublished.

Brynjolfsson, Erik, and Kristina McElheran (2016a). “The Rapid Rise of Data-Driven Decision Making,” *American Economic Review Papers and Proceedings*, 106(5), pp. 133-139.

Brynjolfsson, Erik, and Kristina McElheran (2016b). “Data in Action: Data-Driven Decision Making in U.S. Manufacturing.” Center for Economic Studies Working Paper 16–06.

Brown, Jeffrey R., and Austan Goolsbee (2002). Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry. *Journal of Political Economy*, 110:3, 481-507.

Carlin, Bruce I., Shaun W. Davies, and Andrew Iannaccone (2012). “Competition, Comparative Performance, and Market Transparency.” *American Economic Journal: Microeconomics*, vol. 4, no. 4, pp. 202–237.

Choi, S. Chan, Wayne S. Desarbo, and Patrick T. Harker (1990) “Product Positioning under Price Competition.” *Management Science*, vol. 36, no. 2, pp. 175–199.

Cohen, Wesley, and Daniel Levinthal (1990), "Absorptive capacity: A new perspective on learning and innovation", *Administrative Science Quarterly*, Volume 35, Issue 1 pg. 128-152

Cohen, M.D.; March, J.G.; Olsen, J.P. (1972). "A garbage can model of organizational choice". *Administrative Science Quarterly*, **17** (1), 1–25.

Cyert, Richard, and James March. (1963). “A Behavioral Theory of the Firm.” 2nd ed. Prentice Hall, Englewood Cliffs, NJ.

De Loecker, Jan, and Jan Eeckhout (2018). “Some Thoughts on the Debate about (Aggregate) Markup Measurement,” Working Paper KU Leuven.

Dessein, Wouter, Andrea Galeotti, and Tano Santos (2016). "Rational Inattention and Organizational Focus" *American Economic Review*, Vol. 106:1522-1536.

Dessein, Wouter, and Andrea Prat (2016). "Attention in Organizations." *The Oxford Handbook of the Economics of Networks*, Edited by Yann Bramoullé, Andrea Galeotti, and Brian Rogers.

Dong, Shutao, Sean Xin Xu, Kevin Xiaoguo Zhu (2009). Research Note-Information Technology in Supply Chains: The Value of IT-Enabled Resources Under Competition. *Information Systems Research*, vol. 20, No. 1.

Einav, Liran, Peter Klenow, Benjamin Klopack, Jonathan Levin, Larry Levin and Wayne Best (2017). “Assessing the Gains from E-Commerce,” mimeograph.

Fama, Eugene (1970). “Efficient Capital Markets: A Review of Theory and Empirical Work," Journal of Finance, 25: 383-417.

Farboudi, Maryam, Roxana Mihet, Thomas Philippon, and Laura Veldkamp (2019). “Big Data and Firm Dynamics,” *American Economic Review Papers and Proceedings* 109: 38-42.

Forman, Chris, Avi Goldfarb, Shane Greenstein. 2012. “The internet and local wages: A puzzle,” *American Economic Review* 102 (1), 556-575.

Forth, John, and Alex Bryson (2019). “Management practices and SME performance,” forthcoming at *Scottish Journal of Political Economy*.

Gavazza, Alessandro, and Alessandro Lizzeri (2009). “Transparency and Economic Policy.” *The Review of Economic Studies*, vol. 76, no. 3, pp. 1023–1048.

Gibbons, Robert and Rebecca Henderson (2013). "What Do Managers Do? Exploring Persistent Performance Differences among Seemingly Similar Enterprises." Chap. 17 in *The Handbook of Organizational Economics*, edited by R. Gibbons and J. Roberts, 680–731. Princeton University Press.

Goldfarb, Avi, and Catherine Tucker (2019). Digital Economics. *Journal of Economic Literature*, vol 57(1), pages 3-43.

Granados, Nelson F.; Gupta, Alok; and Kauffman, Robert J. (2006) "The Impact of IT on Market Information and Transparency: A Unified Theoretical Framework," *Journal of the Association for Information Systems*, vol. 7(3), Article 7.

Grossman, Sanford, and Joseph Stiglitz (1980). “On the Impossibility of Informationally Efficient Markets,” *American Economic Review*, 70: 393-408.

Hayek, Friedrich (1945). “The Use of Knowledge in Society,” *American Economic Review*, 35: 519-30.

Hauswald, Robert, and Robert Marquez (2003). Information Technology and Financial Services Competition. *The Review of Financial Studies*, vol. 16, Issue 3, Pages 921–948.

Hsieh, Chang-Tai, and Peter Klenow (2009). “Misallocation and Manufacturing TFP in China and India,” *The Quarterly Journal of Economics*, 124 (4): 1403–1448.

Hortaçsu, Ali, and Chad Syverson (2004). Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds. *The Quarterly Journal of Economics*, vol. 119, Issue 2, May 2004, Pages 403–456.

Huck, Steffen, Hans-Theo Normann, and Jörg Oechssler (2000). Does information about competitors’ actions increase or decrease competition in experimental oligopoly markets? *International Journal of Industrial Organization*, vol. 18, Issue 1, Pages 39-57.

Jin, Ginger, and Phillip Leslie (2003). “The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards,” *The Quarterly Journal of Economics*, vol. 118, Issue 2, Pages 409–451.

Kühn, Kai-Uwe, and Xavier Vives (1995). “Information exchanges among firms and their impact on competition.” Editorial: Luxembourg. Office for Official Publications of the European Communities; Lanham, MD.

Liberti, Jose, Amit Seru and Vikrant Vig (2019). “Information, Credit and Organization,” Stanford GSB Working Paper.

Lucas, Robert Jr. (1972). “Expectations and the Neutrality of Money,” Journal of Economic Theory, 4: 103-24.

Matějka, Filip (2016). “Rationally Inattentive Seller: Sales and Discrete Pricing.” *Review of Economic Studies* 83 (3): 1156-1188.

McAfee, Andrew, and Erik Brynjolfsson (2012). “Big Data: The Management Revolution.” Harvard Business Review 26 September.

McElheran, Kristina (2014). “Delegation in Multi-Establishment Firms: Adaptation vs. Coordination in I.T. Purchasing Authority,” *Journal of Economics and Management Strategy*, 23(2), pp. 225-257.

McElheran, Kristina (2015). “Do Market Leaders Lead in Business Process Innovation? The Case(s) of E-Business Adoption,” *Management Science*, 61(6), pp. 1197-1216.

Mohnen, Pierre, Michael Polder, and George van Leeuwen (2018). “ICT, R&D and Organizational Innovation: Exploring Complementarities in Investment and Production,” NBER Working Paper 25044. http://www.nber.org/papers/w25044.

Pettengill, John S. (1979). “Monopolistic Competition and Optimum Product Diversity: Comment.” *The American Economic Review*, vol. 69, no. 5, pp. 957–960.

Rothschild, Michael, and Joseph Stiglitz (1976). “Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information,” *Quarterly Journal of Economics* 90: 629-649.

Rothschild, Michael, and Joseph Stiglitz (1978). “Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information.” *Uncertainty in Economics, Readings and Exercises*, pages 257, 259-280.

Sadun, Raffaella and John Van Reenen (2005). “Information technology and productivity: it ain't what you do it's the way that you do IT.” Centre for Economic Performance, London School of Economics and Political Science, London, UK.

Schultz, Christian (2004). Market transparency and product differentiation. *Economics Letters* Volume 83, Issue 2, Pages 173-178.

Schultz, Christian (2009). Transparency and product variety. *Economics Letters*. Volume 102, Issue 3, Pages 165-168.

Spence, Michael (1973). "Job Market Signaling," *Quarterly Journal of Economics*, 87 (3): 355–374.

Stigler, George (1962). "Information in the Labor Market," *Journal of Political Economy* 70(5): 94-105.

Stigler, George (1964). "A Theory of Oligopoly," *Journal of Political Economy*, 72 (1): 44-61.

Stiglitz, J. (1985). Information and Economic Analysis: A Perspective. *The Economic Journal, 95*, 21-41.

Syverson, Chad (2004). "Market Structure and Productivity: A Concrete Example." *Journal of Political Economy* 112(6): 1181-1222.

Syverson, Chad (2011). "What Determines Productivity?" *Journal of Economic Literature* 49(2): 326–65.

Tirole, Jean (1988). "The Theory of Industrial Organization," MIT PRESS. Cambridge, MA.

Uriz-Uharte, Guillermo (2019). "The Value of Firm Communication for Innovation: Evidence from the Adoption of Communication Technologies," UCL Working Paper.

Viollaz, Mariana. (2017). "ICT Adoption in Micro and Small Firms: Can Internet Access Improve Labor Productivity?" CesIFO Working Paper Series 6839.

Wu, Lynn, Bowen Lou, and Lorin Hitt (2019). "Data Analytics Supports Decentralized Innovation Communities," Forthcoming at Management Science.

Figure 1: Number of adopters over time

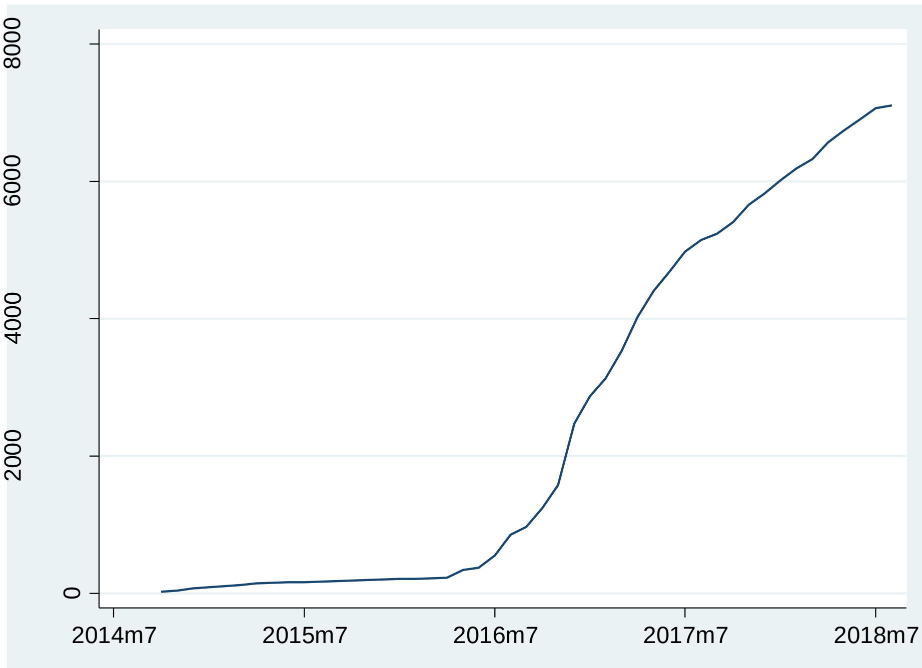

Figure 2: Timeline of Adoption

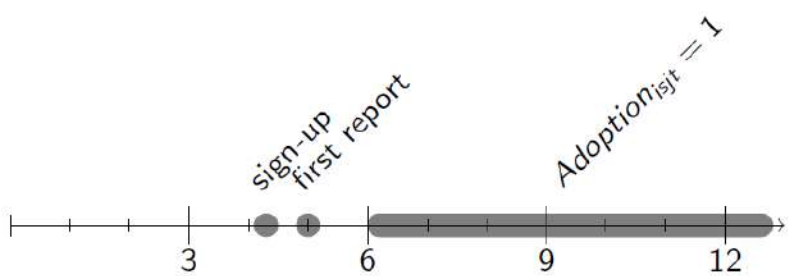

Figure 3: Instrumental variable identification

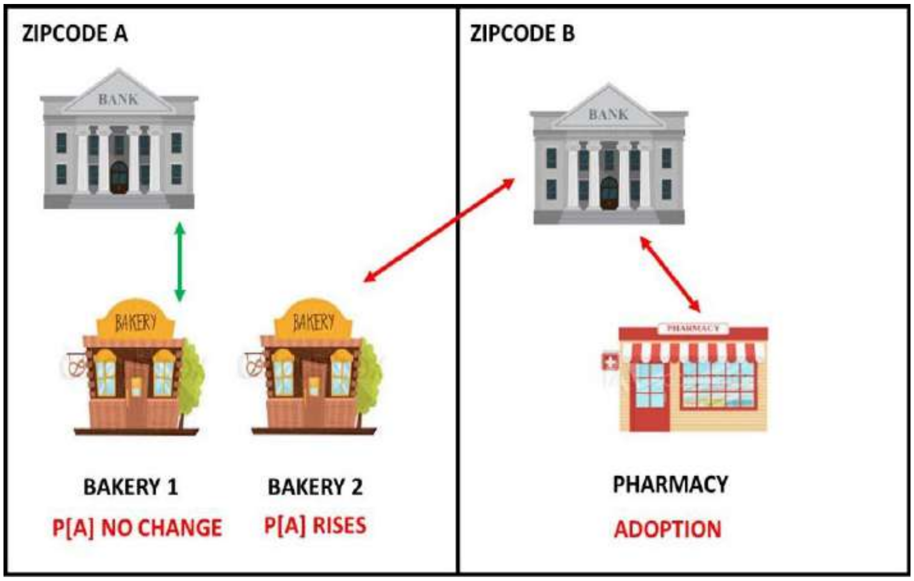

Figure 4: Treatment estimates across sectors, subsectors and regions

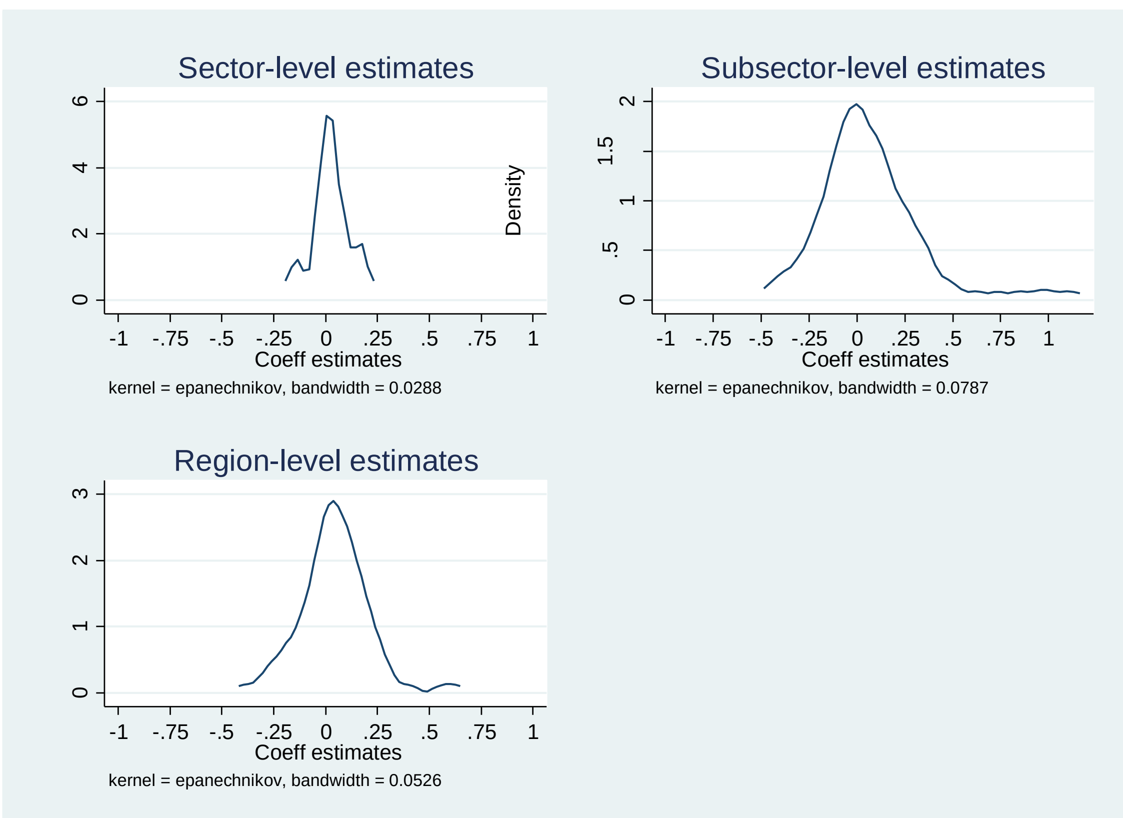

Table 1: Descriptive Statistics

	Observ.	Mean	Std. Dev.	Min	Max
<u>**Full Sample**</u>					
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
<u>**Adopters**</u>					
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.

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.

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.

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.

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 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)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
	OLS	**IV**	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**
Δ Adoption	0,00168	-0,00578			-0.0249***	-0.0344*		
	(0.00315)	(0.0074)			(0.00715)	(0.0178)		
Δ Adoption x High			-0.0197***	-0.0258***			-0.0576***	-0.0868***
			(0.00474)	(0.0081)			(0.0132)	(0.0232)
Δ 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.

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 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.

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.

Table 10: Impact on close competitors

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.

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	2nd-stg	2nd-stg	1st-stg	2nd-stg
	(1)	(2)	(3)	(4)	(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.

Table A3: Changes in Composition of Customers

Dependent variable: Δ Log revenue from sales to prime customer

	(1) OLS	(2) IV	(3) OLS	(4) IV
Adoption	0.0514	0.132*		
	(0.0335)	(0.0754)		
Adoption x High share			-0.112***	-0.0216
			(0.0403)	(0.0798)
Adoption x Low share			0.219***	0.310***
			(0.0534)	(0.0855)
Residual CF				-0.103
				(0.0866)
Sector-zipcd-quarter FE	Yes	Yes	Yes	Yes
Dummies first 4 quarters	Yes	Yes	Yes	Yes
p-value null equal returns			0.00	0.00
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.