REPORT

Data as a Driver of Economic Efficiency

Jian Jia and Liad Wagman
Illinois Institute of Technology, Chicago, IL

DATA Catalyst

October 2018

Data has become a key input in technology-driven innovation and production, spanning industry sectors from pharmaceuticals and healthcare, to automative, smart infrastructure, and broader decision making. This report presents economic analyses of the consequences of data regulation for investment in new technology ventures, for consumer prices, and for economic welfare. The analyses utilize two policy events (the EU's General Data Protection Regulation and the enactment of local privacy ordinances in the San Francisco Metropolitan Statistical Area) along with five microeconomic models to showcase the relationship between data regulation and investment, prices, and welfare.

Jian Jia is a 5th year doctoral student in finance at the Stuart School of Business, Illinois Institute of Technology, 565 W Adams, 4th Flr, Chicago, IL 60661. Tel. 312-810-5168. Email: jjia5@hawk.iit.edu. Liad Wagman is Associate Professor of Economics at the Stuart School of Business, Illinois Institute of Technology, 565 W Adams St, Suite 412, Chicago, IL. Tel. 773-980-9883. Email: lwagman@stuart.iit.edu. Components of this report, in conjunction with co-authors, are from published manuscripts, and other components are being developed as part of ongoing projects by the authors as more data become available; all of the works are detailed in the References. Jia gratefully acknowledges support from the Net Institute (http://www.netinst.org/). Wagman gratefully acknowledges support from Data Catalyst (https://www.datacatalyst.org/) and from the Net Institute. The views expressed herein are the authors' own and should not be interpreted as influenced by nor representing of the views of any organization or collaborators with which or with whom the authors are affiliated.

The General Data Protection Regulation (GDPR) was adopted on April 14, 2016, becoming forceable two years later on May 25, 2018. The regulation aims to protect data by 'design and default,' whereby firms are obligated to handle data according to a set of principles and safeguards. GDPR mandates a higher degree of privacy, data management, and control, as well as requires informed opt-in consent for data collection and assigns substantial liability risks and penalties for data flow and data processing violations. The enactment of GDPR is particularly likely to influence technology firms, given an ever increasing need for the use of data as a core product input. This report begins with a positive economic analysis of the impact of the rollout of GDPR on new technology venture investment in the European Union. The findings indicate a negative differential effect on EU ventures after the rollout of GDPR relative to their US counterparts. The report then presents additional evidence on the relationship between data regulation and consumer-facing prices, where it is shown that opt-in consent can lead to higher prices. The report then presents multiple workhorse microeconomic models that show that data restrictions tend to harm a portion of the consumer population through higher prices.

DATA

We collect data on technology-venture related activity in the EU and US from July 2017 to September 2018 from Crunchbase—a platform for tracking information about technology businesses. In particular, this dataset tracks in detail the parameters of venture financing rounds, such as venture information (name, location, operating category, founding date, financing dates, and a range on the number of employees) and funding information (the size of the funding round in USD, the date each round was announced, the type of financing round such as seed, Series A, Series B, and the number, names, locations, and types of the participating investors). Each funding round observed is treated as a 'deal' event, with deals tallied in each week in each sub-industry in each EU country and in each US state.

We calculate a venture's time-variant age based on its founding date.[1] We consider four different age categories: new ventures/firms (0-3 years old), young firms (3-6 years old), established firms (6-9 years old), and mature firms (9+ year sold). Firms may consequently switch among age categories in our sample. Each venture in the Crunchbase dataset is also tagged with a few relevant product keywords (e.g., 'software', 'e-commerce', 'finance', etc). We add to this categorization by

[1] There are some cases where a founding date is unavailable or when a venture's first financing round predates its founding; in those cases, we use the venture's first financing round as its founding date.

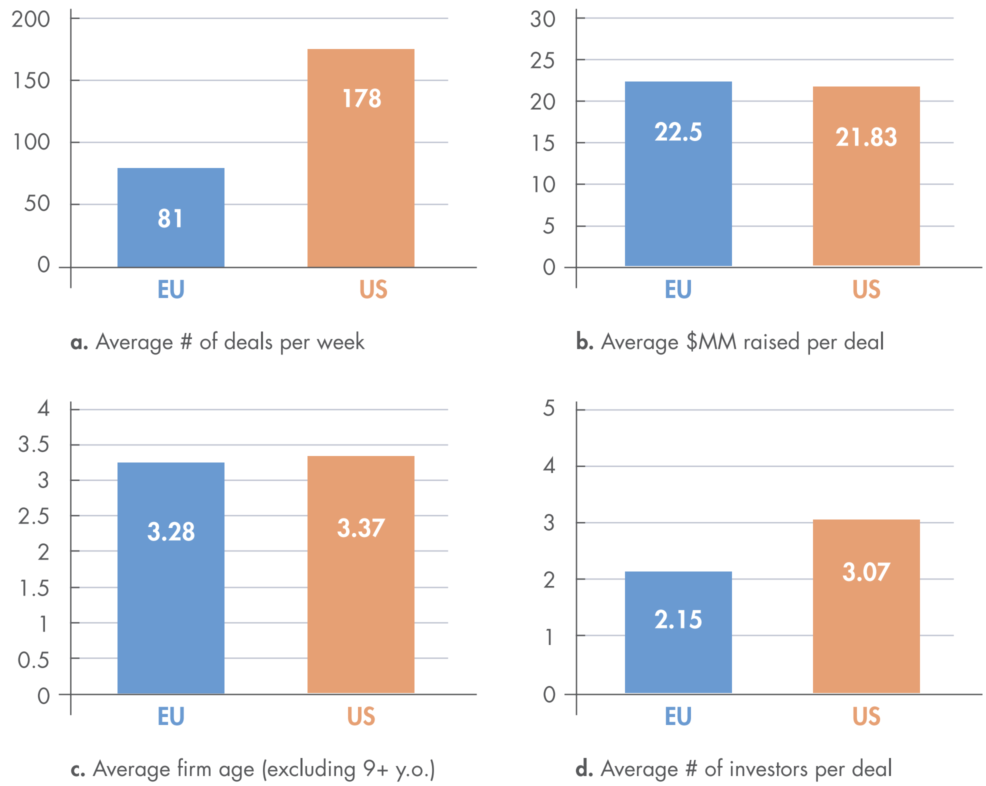


Figure 1: Whole sample summary statistics

grouping ventures, based on their tagging, into two unique categories of healthcare-financial, and all other technology.[2] We further collect local macroeconomic controls such as unemployment rate, CPI, interest rate, and GDP, for each US state and EU country in which a venture is located.[3]

Figures 1-3 along with Table 1 in Appendix A summarize the data. Figures 1(a) and 1(b) indicate that the average number of deals per week and the median amount raised per deal in the US are more than twice that in the EU. Figure 1(c) indicates a similar average age of US and EU technology ventures, though US venture deals appear to involve a higher number of investors on average relative to EU deals as indicated in Figure 1(d).

Figure 2 depicts the distribution of technology firms' ages in our dataset in the US and EU. While they are similar, the US has a larger proportion of 9+ year old firms. The EU, in contrast, has a larger proportion of firms in the 0-3 and 3-6 year age groups—firms which may be in the process of raising their initial financing rounds—and which may be particularly susceptible to higher costs of compliance. Of particular interest is the fact that near half of technology ventures in the EU and US are relatively new, 0-3 year old ventures.

[2] The main results are qualitatively unchanged when we use an unsupervised k-means clustering approach. We prefer our chosen categorization because it allows us to isolate groupings that may be of particular interest due to their tendency to handle personally identifiable and sensitive consumer information.

[3] For a few months in 2018 for which macroeconomic data is yet available, we extrapolate macroeconomic variables by using their corresponding growth averages from the past 2 years.

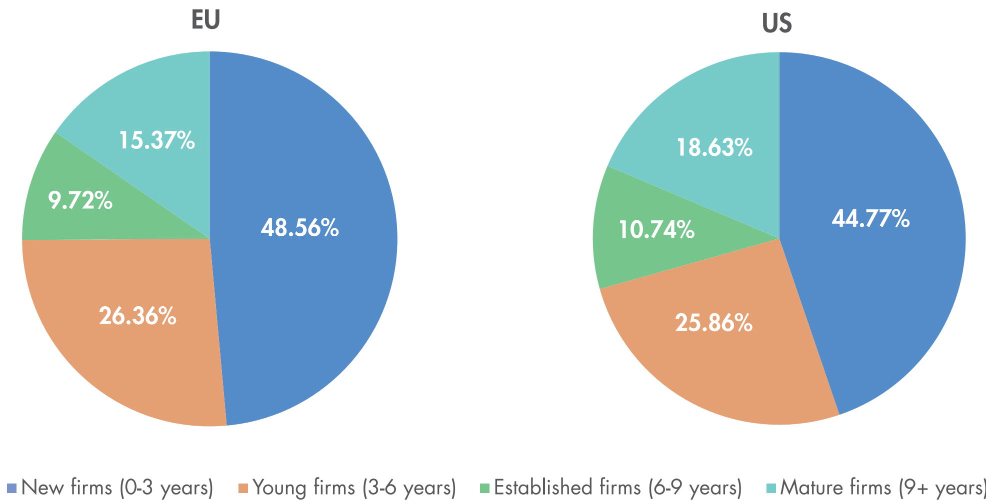


Figure 2: The proportion of technology ventures in each age group in the EU and US

A comparison by age subgroups is presented in Figure 3, indicating that more than 70% of US and EU deals are by new and young (0-6 year old) firms. It is also apparent that the older the firm, the higher the average dollar amount raised per deal. Other summary statistics are provided in Table 1.

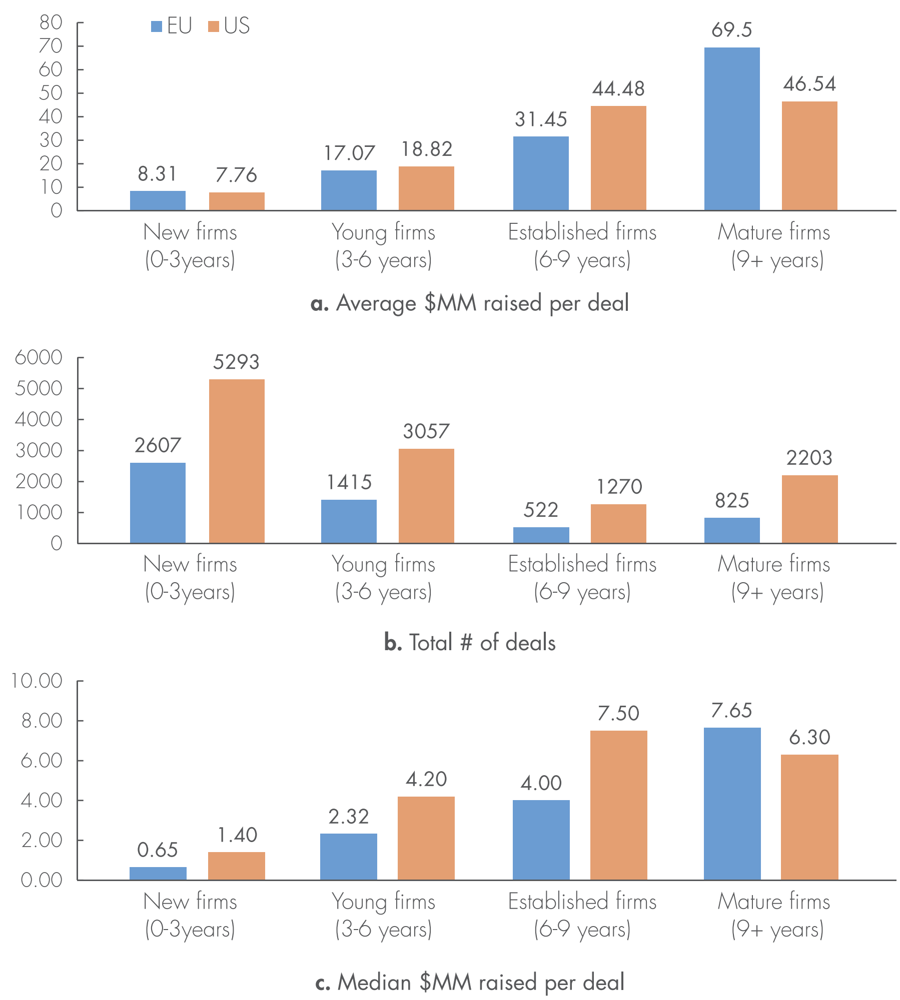


Figure 3: Summary statistics by technology venture age group

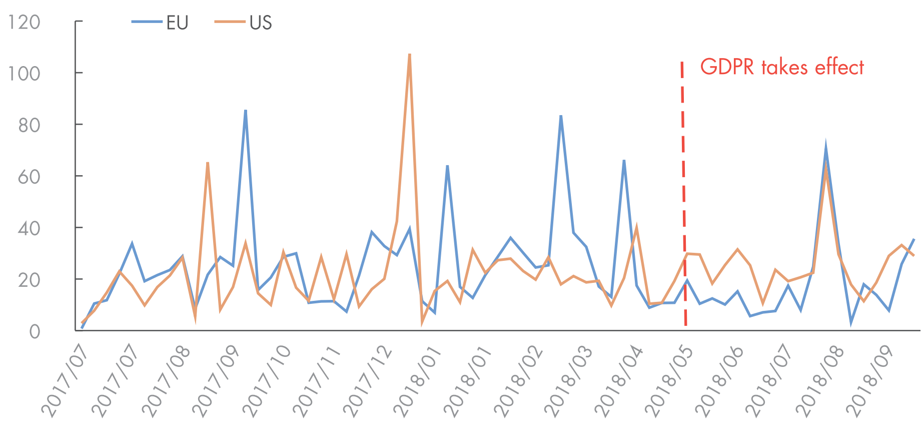
EMPIRICAL METHODOLOGY

We aim to study the effects of the rollout of GDPR in May 2018 on venture financing in the EU. We do so by contrasting venture activity in the EU with the US before and after the rollout of GDPR, taking advantage of preexisting trends in the EU and US venture environments. While GDPR was enacted in April 2016, its enforceability began to take hold in May 2018, with mandatory implementation by EU member states and mandatory compliance by firms.

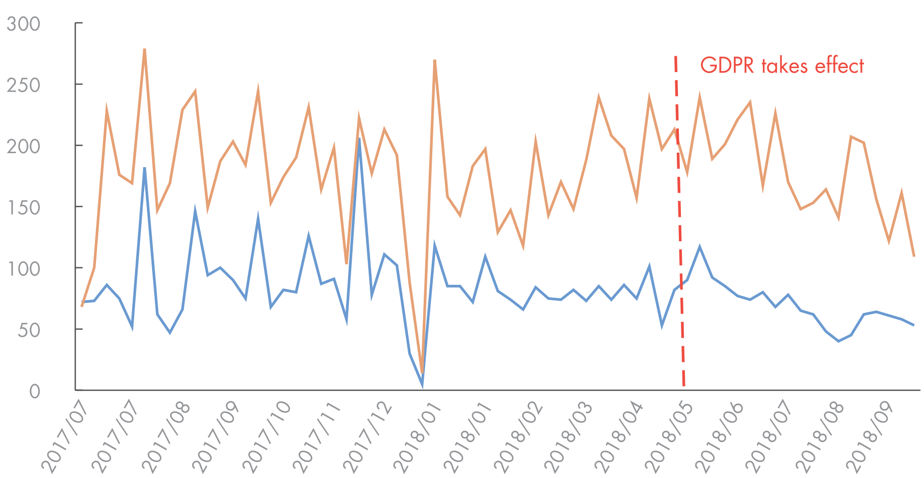
Our hypothesis is that as GDPR's enforceability was coming into place, entrepreneurs and investors both realized the actual compliance and implementation costs, as well as the ex-post implications of GDPR. This is particularly evident as major platforms like Google, Facebook, Amazon, and Apple, on which a vast number of technology ventures rely, began to reveal the ways in which they were tightening their platforms and app stores with new data sharing, data portability, and data liability rules. The dataset suggests immediate consequences to those revelations, with investor dollars differentially lessening and fewer venture deals being executed in the EU relative to the US following the rollout of GDPR.

Figure 4 depicts trend lines of the average amount raised per deal, and the average total number of deals per week from July 2017 to September 2018. Subfigures 4(a) and 4(b) both suggest that some sustained divergence is taking place between EU and US ventures after GDPR came into effect. Both EU and US trends also track each other closely otherwise, and particularly so up until May 2018.

Figure 4: Trends of the average amount invested per deal and average weekly number of deals



a. Average $MM amount per deal at weekly frequency



b. Average total number of weekly deals

We test the effect of the enactment of GDPR using a difference-in-difference methodology. We add macroeconomic control variables and time, state, and country fixed effects. Our treatment group comprises EU ventures and our control group comprises US ventures. While the treatment group does have lower levels of venture activity than the control group, there does not appear to be a differential pre-trend that would violate the common trend assumption in our difference-in-difference analysis. We use Tobit (for $ amount) and Poisson (for number of deals) regressions at the aggregate level for which:

$$y_{jkt} = \alpha_t + \alpha_k + \delta X_{jkt} + \beta GDPR_{kt} + \varepsilon_{jkt} ,$$

where *j* identifies ventures according to their assigned unique identifier, *k* denotes whether a firm belongs to the treatment (EU) or control group (US), *t* indexes time, y_{jkt} is the dependent variable of interest, which is either the dollar amount raised or the number of financing deals in each subindustry, α_t and α_k are week and state (US) or country (EU) fixed effects, X_{jkt} are macroeconomic control variables (monthly unemployment rate, CPI, interest rate, exchange rate, and quarterly GDP), $GDPR_{kt}$ is a dummy variable that equals one after May 25, 2018 if applicable to start-up *j* in group *k* and zero otherwise, and ε_{jkt}, is an error term. This methodology controls for fixed differences between treated and untreated ventures via state and country fixed effects, and week dummies control for aggregate fluctuations. Our estimate of the effect of the enactment of GDPR is the coefficient β. At the deal level, we use a log linear OLS specification and add deal-level controls such as funding stage, investor type, and firm age.

THE EFFECT OF GDPR ON INVESTMENT

We compare the total dollar amount raised per week per state/country per technology category, the average dollar amount raised per deal, and the average total weekly number of deals per state per category, before and after May 25, 2018, using US ventures as a control group. We identify a significant negative effect of GDPR against EU ventures (Tables 2-3). The proceeding figures depict how the rollout of GDPR affected the total amount raised (per week, per state/country, per category), the amount raised per deal, and the overall average number of deals per week per category.

Figure 5: Effect of GDPR on total $MM raised per week/country/category by EU firms

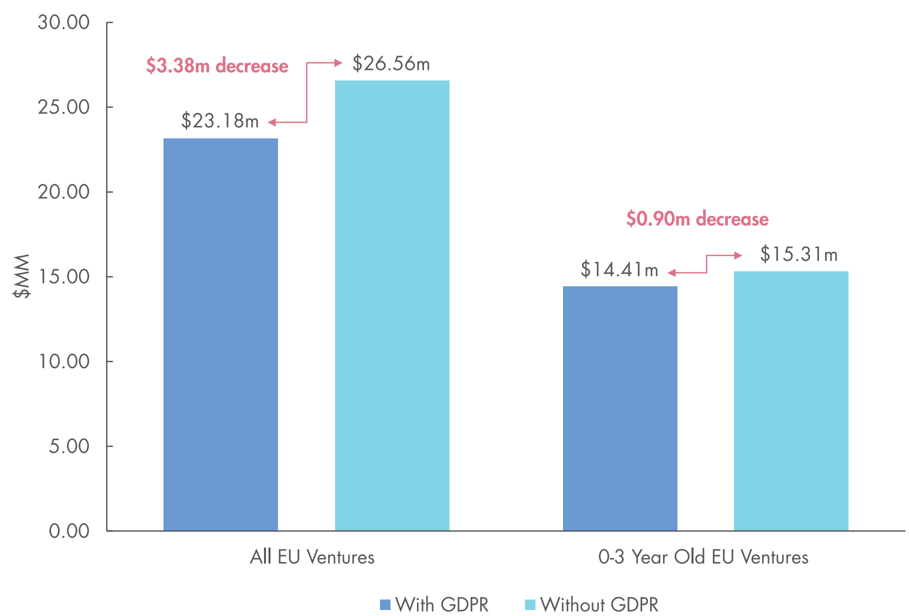
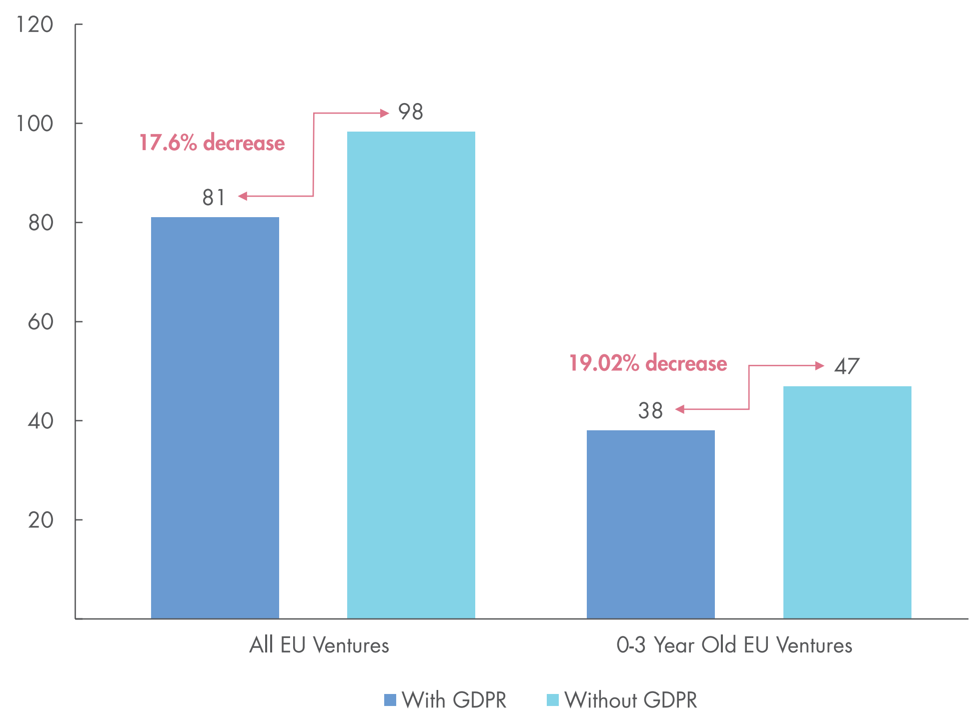


Figure 5 illustrates the negative effect at the margin of the enactment of GDPR on the dollar amount raised per week per EU country per category. The light blue bars give the predicted average dollar amounts in millions of US dollars per week/country/category invested in EU technology firms in our dataset if GDPR were not rolled out. The blue bars give the average dollar amounts invested per week/country/category that we observe post GDPR, a lower amount than if GDPR were not enacted. Specifically, the result is a reduction of $3.38 million at the margin across all EU firms (Column 2 of Table 3). New, 0-3 year old ventures incur a decrease of $0.90 million at the margin (Column 4 of Table 3) per week/country/category.

In addition to the negative effect on the total dollar amount raised per week/country/category is a reduction in the total average number of financing deals, as illustrated by Figure 6. The figure depicts the results of Poisson specifications for testing the effect of GDPR on the number of EU venture deals. In Figure 6, the light blue bars show the predicted average total weekly number of technology venture financing deals that would have taken place in the EU had GDPR not been rolled out, whereas the blue bars give the averages that we observe. Specifically, our analysis finds an average 17.6% decrease for the full sample in the number of financing deals of EU firms after the rollout of GDPR (the interpreted Poisson coefficient estimate in Column 1 of Table 3), and a 19.02% decrease for a subsample comprising only 0-3 year old technology ventures (the interpreted Poisson coefficient estimate in Column 3 of Table 3).

To corroborate these findings, we also run a log-linear regression at the deal-level to test the effect of GDPR on the amount raised per individual deal, and the results confirm a negative effect on EU venture deals post GDPR, as depicted in Figure 7 and summarized in Table 2.

Figure 6: Effect of GDPR on the number of weekly deals by EU technology ventures



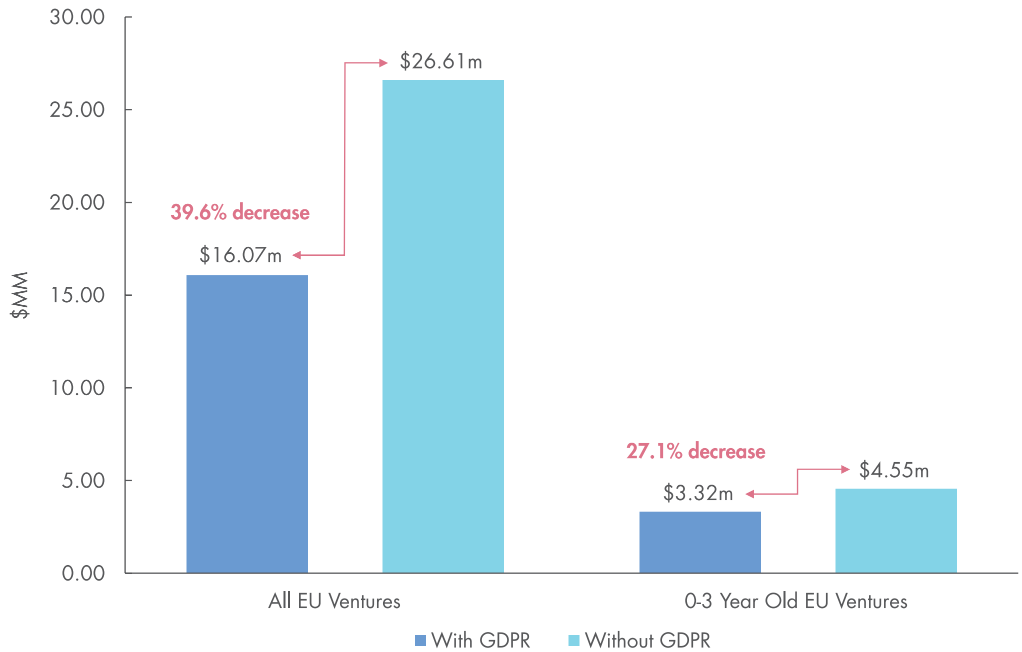


Figure 7: Effect of GDPR on the dollar amount raised per deal by EU technology ventures

THE EFFECT OF GDPR ON JOBS

For each venture, our dataset provides a range on the number of employees (e.g., 1-10, 10-50, 50-100, etc). However, this range is time-invariant as of October 1, 2018; that is, we do not have historical ranges. In other words, as of October 1, 2018, we have the total dollar amount raised by each firm and a range on its number of employees. We can use this information to provide a back-of-the-envelope measure of the average dollar amount raised per 'current' employee as a function of the firm's age. We pay particular attention to new (0-3 year old) ventures, because they tend to be the primary job creators,[4] and we focus our analysis on the EU to assess the potential for EU technology job losses as a result of GDPR.

In Figure 8(a), the blue bars represent average lower and upper bounds on the annualized dollar amount raised per employee across all 0-3 year old EU technology firms that were founded on or after 2015. We focus on those firms because we can observe in the data the total amount of venture financing they have raised since their founding. As the figure illustrates, the annualized range on the dollar amount raised per employee for 0-3 year old EU technology firms has an average lower bound of $123,246 and an average upper bound of $1,019,763.

Given the negative effect of GDPR on the total dollar amount raised by EU ventures in the four months post-GDPR period of our dataset, we can provide a rough estimate of the range of potential annual EU technology job losses incurred by 0-3 year-old firms as a result of GDPR. To obtain an estimate for annual job losses, we extrapolate the

[4] See, e.g., Haltiwanger et al., 2013.

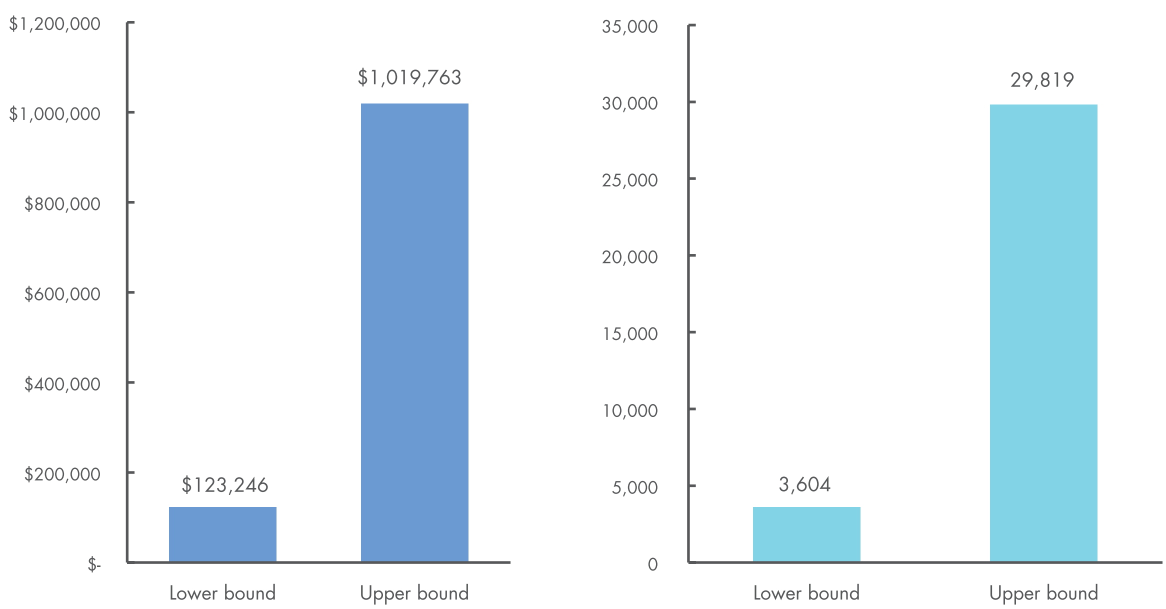
back-of-the-envelope job loss potential in June through September 2018 to an entire calendar year by examining the relationship between EU venture activity in June through September 2017 to the rest of calendar year 2017.

The outcome of this calculation is presented in Figure 8(b), which indicates that over a calendar year, post-GDPR, a lower bound back-of-the-envelope estimate on the number of technology venture jobs lost in the EU by 0-3 year-old ventures is 3,604 jobs, and an upper bound is 29,819 jobs. As a percentage of the total range on employment by EU ventures founded on or after 2015 in our sample, the estimated job losses by 0-3 year old ventures translate to a 4.09%-11.20% loss in the number of jobs employed by those nascent EU firms.

We emphasize that given our data, our measure can only provide rough back-of-the-envelope ranges on job loss estimates. This is because, on the one hand, we have no insight into whether investors are taking a wait-and-see approach nor do we know the outside options of affected firms or of those individuals who would have been employed by the firms in our dataset had it not been for GDPR. There may also be jobs created as a result of GDPR (for instance, data privacy compliance officers, data security and management ventures, etc).

On the other hand, the potential for job losses may well extend and intensify past our four months post-GDPR dataset period, in which case the effect on jobs is understated. Moreover, our estimates do not incorporate potential foregone downstream job creation (for instance, jobs to service the additional employees that may have been employed had it not been for the rollout of GDPR).

Figure 8: Range estimates of $ per employee and GDPR-related job losses for 0-3 year-old EU ventures



a. 0-3 y.o. firms average annual $/employee range

b. Estimate of annual job loss range by 0-3 y.o. firms

SUMMARY: THE EFFECT OF GDPR ON NEW VENTURE INVESTMENT

We presented a preliminary analysis of the effects of the rollout of GDPR on new technology venture financing in the EU. The findings indicate a significant negative differential effect on EU ventures after GDPR came into effect relative to their US counterparts, with fewer venture financing dollars, fewer venture financing deals, and considerable potential for job losses. The decrease in jobs and venture financing could mean (i) a decrease in innovation, particularly if investors choose not to invest in new ventures, and (ii) a shift of innovation from ventures to established firms, which can entrench incumbents and increase market power (Krasteva et al, 2015; Campbell et al., 2015). We caution that given the extent of our data, which only comprises four post-GDPR months, it is impossible to rule out that our estimates may be reflective of a short-term reaction and/or a wait-and-see attitude by investors and entrepreneurs, rather than a permanent long-term consequence of GDPR. The long-term impact of GDPR on the EU technology venture scene will be clearer as more data is available over time.

Data Restrictions and Prices

Consumer prices have long been a critical measure for evaluating the impact of policy. Assessing the impact of a potential data regulation—and in particular, opt-in consent—on consumer prices is challenging. However, it has been done theoretically with economic models and empirically through natural experiments in other markets where related laws exist. For instance, there is evidence (Kim and Wagman, 2015) that the prices (interest rates) of mortgages that are offered to consumers, everything else held equal, increase under an opt-in consent regime for data collection relative to an opt-out regime.

Consumer data is routinely collected by financial institutions, where a plethora of personally-identifiable information about customers is used to determine product eligibility. When consumers apply for a loan, for example, they provide information about their financial situation that is arguably of a personal nature. A lender may collect additional pieces of information from sources such as credit reports prepared by credit bureaus. Lenders and loan-servicing companies also have records of a customer's current balance, the frequency and timing of payments, and in some cases information about insurance policies obtained. Such information can be shared with affiliates or sold to outside companies (non-affiliates), including markets and data brokers, where it is used as lead generation to better target consumers who may be interested in related products and services.

REGULATORY BACKGROUND

In 1999, Congress enacted the Gramm-Leach-Bliley Act (GLBA), allowing a variety of financial institutions to sell, trade, share, or give out nonpublic personal information about their customers.[5] GLBA requires financial institutions to notify consumers about how their personal information is collected and used. In particular, financial institutions that share or sell consumer data to non-affiliated third parties must give customers a chance to opt out; that is, to request that their information not be shared (15 USC §§ 6801- 6809). However, there are several exemptions under the GLBA that can permit information sharing despite a consumer's objections.[6] Since the enactment of the GLBA in 1999, there has been much debate about whether GLBA privacy provisions meet increasing public concern surrounding consumer privacy.[7]

One of the main criticisms of GLBA's privacy provisions has been that most consumers do not (and likely will not) take advantage of the opt-out option to cease trade of their information. In 2002, three out of five counties in the San Francisco-Oakland-Fremont (SFOF), CA, Metropolitan Statistical Area (MSA) adopted a local ordinance (effective January 1, 2003) that was more protective than previous practices by pursuing an *opt-in* approach. Specifically, the local

[5] The GLBA partially repealed the Glass-Steagall Act of 1933 by allowing banking, insurance, and securities companies to operate under the same entity. Financial holding companies so created can have a variety of non-banking affiliates. Under the GLBA consumers have no right to stop sharing of nonpublic personal information among affiliates.

[6] A financial institution can, e.g., share information with a marketer in order to jointly offer products. In fact, many of the nation's leading banks use information about their customers' shopping habits to help retailers target offers to customers without actually releasing their data.

[7] The 107th Congress introduced bills that seek to modify the GLBA to require opt-in consent for information transfers. See, for instance, the Financial Institutions Privacy Protection Act of 2001, S. 450, § 3 (2001); Consumer's Right to Financial Privacy Act, H.R. 2720, § 2 (2001).

ordinance would require financial institutions to seek a written waiver before sharing consumer information with both affiliates and non-affiliates.[8]

OPT-IN CONSENT

Analyzing census tract-level and individual loan-level data on mortgage applications for conventional home purchase, as well as refinancing, it is possible to study the differential effect of an opt-in regime relative to an opt-out regime on loan denial rates, and, by direct implication, the impact on offered mortgage interest rates (prices). Comparisons of loan denial rates before, during, and after the adoption of local ordinance offer a unique opportunity for evaluating the effects of strengthening consumer financial data restrictions. Specifically, an opt-in consent regime is shown to have a statistically significant negative effect on loan denial rates (that is, approval rates increased), an indicator of higher mortgage rates. This finding indicatesthat the resultant mortgage interest rates are higher under opt-in than opt-out consent. The result is further reinforced by corroborating evidence about downstream foreclosures, as well as by an economic model of consumer lending and data flow restrictions that is provided in Appendix B. The model illustrates the direct link between lower loan denial rates and higher loan prices. A summary of the results of this combined analysis is provided in Figure 11.

Opt Out	Opt In
More data is collected	Less data is collected
More efficient matching between borrowers and loans	Less efficient matching between borrowers and loans
Lower mortgage prices	Higher mortgage prices

Figure 11: Summary of outcomes of opt-out and opt-in regimes for information sharing.

While an opt-in consent regime does appear to result in less data being collected about borrowers, it also appears to result in less efficient matching between borrowers and loans (higher default and foreclosure rates) as well as in higher consumer-facing mortgage prices (interest rates).

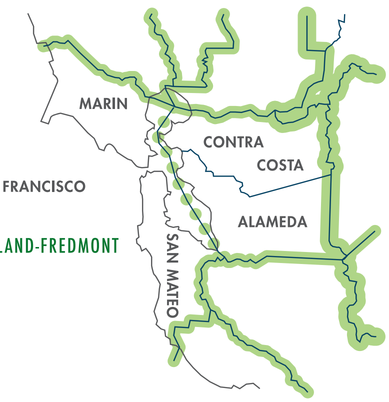
[8] The California Constitution allows a county or city to make and enforce within its limits all local, police, sanitary, and other ordinances and regulations that do not conflict with the state's own general laws. That is, an ordinance is a local law adopted with all the legal formalities of a statute. Further, the privacy ordinance applied to all 'financial institutions' that engaged in financial activities (as described in 12 USC §§ 1843) and conducted business in the county. Thus, regardless of the institutional type (e.g., regulatory agency or charter status) or its size, any institution that is significantly engaged in financial activities had an obligation to abide by the local ordinance, creating policy variation across counties until it was superseded by the California state law (effective July 1, 2004), which provided the consumer financial-information privacy the local governments had sought.

DATA AND BACKGROUND

ince the enactment of the GLBA, there have been considerable legislative activities in state governments in regards to privacy issues, financial privacy in particular. Some activities pertain to the adoption of an opt-in standard.[9] As a result, there exists significant state-level variation in the protection and trade of consumers' financial information by financial institutions, which lends itself to understanding the effects of data regulation.[10]

Five counties in the San Francisco-Oakland-Fremont Metropolitan Statistical Area (MSA) considered adoption of opt-in privacy ordinances. Three of them (Alameda, Contra Costa, and San Mateo) adopted the ordinance, while the other two (Marin and San Francisco) did not (see Figure 12 for a map of this MSA).

Figure 12: Map of the five counties in the San Francisco-Oakland-Fremont MSA.



The Home Mortgage Disclosure Act (HMDA) requires financial institutions (including banks, savings associations, credit unions, and other mortgage lending institutions) to annually report disclosures of their lending activities. Using the data submitted by these financial institutions, the Federal Financial Institutions Examination Council releases aggregate lending information on the disposition of mortgage applications by categories (e.g., loans on 1-to-4 family dwellings) at the Census tract level.

For each Census tract, the data show how many (1-to-4 family home) loans were originated, how many were approved but not accepted, and how many applications were denied, withdrawn, or closed for incom- pleteness. The data also show the aggregate dollar amounts in each of these five categories. Importantly, the HMDA aggregate reports do not include any resale loans (i.e., loans purchased by institutions, which has action-type code 6) as well as any preapproval denied or approved (loans which have action-type codes 7 and 8, respectively, in an institution's reporting). Thus, the data is based on snapshots at the time of origination or denial, and the focus is on conventional home-purchase loans and refinancing loans for 1-to-4 family dwellings.

[9] GLB permits states to formulate privacy protections that exceed federal law. The Federal Trade Commission (FTC) was granted sole authority to determine whether state statutes are inconsistent with (and therefore pre-empted by) the GLBA. In 2001 and 2004, the FTC issued a formal letter in which it determined that affirmative opt-in provisions in Illinois and North Dakota, respectively, were consistent with GLBA.

[10] Features of state privacy legislations are less suitable for studying state-level policy variation. Some states, including Alaska, Connecticut, Illinois, North Dakota, and Vermont, have strict laws that require opt-in consent for the sharing of consumer information with unaffiliated third parties while other states may not. These states adopted an opt-in approach in their banking laws long before GLB was enacted in 1999. An increasing number of states have enacted laws that limit the sale of personal information by financial institutions and impose stricter requirements for third-party uses. The best-known example is the California Financial Information Privacy Act (CalFIPA), which in partsuperseded the opt-out approach of the GLBA. The opt-out provision of the CalFIPA regarding affiliate sharing was preempted by the Fair Credit Reporting Act (American Bankers Association v. Gould, 412 F.3d 1081). However, in 2008 the Ninth Circuit revived part of California's law regarding affiliate sharing (American Bankers Association v. Lockyer, No. 05-17163), and the Supreme Court denied review in 2009.

Table 4 shows the variables. In constructing mortgage application denial rates, the number (or dollar amount) of applications that were denied is divided by the sum of the number (or dollar amount) of i) loans originated, ii) loans approved but not accepted, and iii) applications denied. Controls include tract-level economic characteristics (median income as % of MSA median; % of population below poverty threshold; an indicator for being inside a central city), population characteristics (% of Minority, or Asian, Black, and Hispanic population), and housing characteristics (median age of housing stock; % of owner-occupied housing units; number of households to housing units).

RESULTS

Table 5 and Table 6 contain estimation results from the following difference-in-differences specification:

$$Denial\ Rate_{it} = \beta\, Treat_{it} + \gamma X_{it} + Tract_i + Year_t + \varepsilon_{it},$$

where for each tract i and year t, $Treat_{it}$ is an indicator variable for tracts belonging to the three treatment counties during the intervention period and for all tracts in the five counties during the post-intervention period; and X_{it} is a set of Census tract-level characteristics. Standard errors are clustered at the Census tract-level and given in the parentheses.

Table 5 shows the estimation results for conventional home-purchase loans. Specifically, columns 1–3 show the estimation results for mortgage denial rates in numbers and columns 4–6 show them for denial rates in dollar amounts. The treatment effect of the privacy ordinance is statistically significant and negative in all specifications, where the magnitude of the effect on mortgage denial rates, both in number and in amount, seems to be around 1%. Similarly, Table 6 replicates the same models for home refinancing loans. The results here indicate that the treatment effect is about a 0.5% decrease in the denial rates for refinancing loans.

Using loan-level data, it is possible to classify individual loan applications into conforming and non- conforming loans based on the loan amount. This classification can then be used to examine the effects of securitization. Many lenders sell their loans to government-sponsored enterprises such as Fannie Mae and Freddie Mac shortly after originating them. This is often done because retaining a mortgage represents a significant risk for the lender. However, these government-sponsored enterprises only accept loans that meet certain standards (i.e., conforming loans). One of the most common types of non-conforming loans is the 'jumbo' loan, a loan for an amount that exceeds the limit set by Fannie Mae and Freddie Mac. Because these enterprises do not typically purchase non-conforming loans, lenders may have to retain a larger proportion of jumbo loans which would mean that lenders have greater incentives to screen non-conforming loan applications relative to conforming loans. Therefore, one would anticipate stronger robustness results for jumbo loans than for conforming loans.

Table 7 shows OLS estimation results using the following linear probability model:

$$Reject_i = \beta\, Treat_{st} + \gamma X_i + Year_t + County_s + Lender_j + \varepsilon_{ijst},$$

where for each loan application i, $Treat_{st}$ is an indicator for treatment status for county s and year t; X_i includes an applicant's race, sex, and mortgage-to-income (MTI) ratio; and a full set of lender dummies (of which there are 1483 in total). Standard errors are clustered by lender.

The first two columns of Table 7 show the estimation results by loan purpose, where the treatment effect is statistically significant for purchase loans but insignificant for refinancing loans. The next two columns show the estimation results using only conforming-size loans. The change is that treatment effect on purchase loans is now marginally insignificant at the 5% level which suggests that the opt-in consent regime might have had only marginal (little) effect on lenders' incentives to screen purchase and refinancing loans when they were likely to be subsequently purchased by government-sponsored enterprises. On the other hand, the last two columns contain estimation results using only jumbo loans. Here, the treatment effect on both purchase and refinancing loans is statistically significant and negative which indicates that the opt-in consent regime had a relatively larger impact on lenders' incentives to screen jumbo loans, presumably because they had to retain a larger proportion of them.

Table 8 presents reduced-form estimates of cross-sectional regressions where the dependent variable is 2007-08 foreclosure start rates. Columns 1–3 show that the coefficient on the indicator for the three treatment counties that adopted the opt-in consent regime are significant and positively associated with the estimated foreclosure start rates in 2007-08. This does not provide direct evidence in support of the notion that the 2003-04 local opt-in consent ordinance is causally associated with this rate differential because the foreclosure data is not decomposable by loan age. However, given that the other predictors such as house price change and tract-level loan characteristics absorb a significant amount of variation in the data, the 1% increase in foreclosure start rate (hence, < 0.5% in predicted foreclosure rate) is suggestive of additional downstream costs that are associated with opt-in consent regimes—specifically, the costs of potential mismatches between borrowers and lenders.

To explore this possibility further, columns 4–6 of Table 8 use the share of 2003-04 HMDA loans in the HUD estimated number of mortgages in 2008 interacted with the indicator for the treatment counties. The idea is that if those tracts with higher shares of 2003-04 loans in the treated counties are disproportionately more likely to have higher foreclosure start rates, then it would be consistent with additional downstream costs. In fact, the coefficient on the interaction term is indeed statistically significant and positive. The interactive effect is an increase of 1% (= .2222 .0452) in foreclosure start rates in the treatment counties that adopted the opt-in consent regime evaluated at the simple mean of the 2003-04 share. The result of additional downstream costs is further corroborated by a model in Appendix B..

SUMMARY: OPT-IN VS OPT-OUT CONSENT FOR DATA SHARING

Combined with the economic model in Appendix B, the preceding empirical analysis helps break down the dynamics behind opt-in and opt-out consent regimes and give rise to the following: Under opt-in consent regimes for data trade and commercialization, lenders are less able to monetize information about borrowers. Consequently, consumer-facing prices increase and lenders collect less information, which increases the potential of a mismatch between borrowers and loan products. The result is both higher prices and a higher potential for additional downstream costs as a result of product mismatch.

Data Restrictions and Welfare: A Four-Model Comparison

Given the unprecedented availability of consumer data, a recurrent question is who stands to win and who stands to lose from imposing data restrictions on firms. The economic literature shows that the answer is not clearcut and under many models some consumers stand to lose from data restrictions. To demonstrate this, the literature (e.g., Taylor and Wagman, 2014) has examined diverse models of competition and shows that who benefits and who loses from data restrictions largely depends on the specific industry structure and corresponding model under consideration. Therefore, a one-size-fits-all approach for different industries and competitive landscapes is likely to be economically inefficient.

Here, we showcase results from four fundamental "workhorse" models that are commonly used in the antitrust and consumer protection literatures: (i) a linear city duopoly model (LCM, Model 1), (ii) a circular city oligopoly model (CCM, Model 2), (iii) a vertical differentiation duopoly model (VDM, Model 3), and (iv) a multi-unit symmetric demand duopoly model (MSDM, Model 4). The effects of imposing data restrictions on firms that disallow them to use data to adjust their prices to match consumer preferences are summarized in the proceeding figure. The formal derivations of these results are in Appendix C.

Outcome of Imposing Data Restrictions	Model 1	Model 2	Model 3	Model 4
Consumer Surplus	Lower	Lower	Lower	Higher
Total Industry Profits	Higher	Same	Higher	Lower
Overall Welfare	Same	Lower	Lower	Mixed
Consumers Prefer Data Restrictions	None	Mixed	Mixed	Mixed

Figure 13: Summary of the predictions from the four oligopoly models.

As the figure indicates, of the four models, only one predicts a clear gain for (overall) consumer surplus as a result of data restrictions. In other words, consumers, in an overall sense, tend to benefit from greater access to firms' products under a data-access regime that imposes less restrictions on firms. Moreover, rather than benefitting consumers overall, two of the models predict that data restrictions stand to benefit firms' profits, rather than consumers—the reason for this profit gain is a reduction in firm-to-firm competition, which results in surplus being transferred away from consumers and to firms.

An opt-in consent regime essentially imposes additional data restrictions on firms (Johnson and Gold- stein, 2004), and can thus, at least in part, have consequences along the above lines. The four models, as summarized in Figure 13, thus provide further corroborating evidence to the results in the preceding analysis. While the effects of data restrictions are not equal across models, the outcome is often less efficient (by way of lower economic welfare) with data restrictions in place, and the preference for these restrictions among consumers usually vary. We emphasize that these results hold despite no assumptions being made about any intrinsic tastes for privacy among consumers (i.e., consumer preferences in the four models are driven purely by price considerations). In particular, high-value consumers for a given product tend to prefer that data restrictions are in place, whereas low-value consumers tend to prefer no data restrictions.

Imposing data restrictions can reduce low-value consumers' ability to purchase products—in the absence of data, firms will no longer be able to target those consumers with discounts. Moreover, if lower consumer product valuations are correlated with lower consumer incomes, then imposing data restrictions can reduce lower-income consumers' ability to purchase products, and may consequently hinder their overall access to the market.

SUMMARY: DATA RESTRICTIONS AND WELFARE

The findings from the four models caution that data restrictions must be understood within their individual market contexts and respective industries, and that the consequences of imposing data restrictions crucially depend on the specific competitive landscapes and markets at play, and may not necessarily apply more broadly. In the interest of economic efficiency, these findings call for a nuanced approach to data flow restrictions that is tailored specifically to each market. The risk of not doing so is that data restrictions can harm some consumers' access to the markets.

References

Campbell, J., A. Goldfarb, and C. Tucker (2015): "Privacy regulation and market structure," *Journal of Economics & Management Strategy*, 24(1), 47–73.

Haltiwanger, J., R. S. Jarmin, and J. Miranda (2013): "Who Creates Jobs? Small versus Large versus Young," *Review of Economics and Statistics*, 95(2), 347–361.

Jia, J., G. Jin, and L. Wagman (2018): "The Short-Run Effects of GDPR on Technology Venture Investment," *Working paper.*

Johnson, E., and D. Goldstein (2004). "Defaults and donation decisions," *Transplantation*, 78(12), 1713–1716.

Kim, J.-H. and L. Wagman (2015): "Screening Incentives and Privacy Protection in Financial Markets: A Theoretical and Empirical Analysis," *RAND Journal of Economics*, 46(1), 1–22.

Krasteva, S., P. Sharma, and L. Wagman (2015): "The 80/20 Rule: Corporate Support for Innovation by Employees." *International Journal of Industrial Organization*, 38: 32–43.

Taylor, C. R. and L. Wagman (2014): "Consumer Privacy in Oligopolistic Markets: Winners, Losers, and Welfare." *International Journal of Industrial Organization*, 34: 80–84.

Appendix A: Tables

Appendix A: Tables

Table 1: Summary Statistics

	EU			US		
	Mean	*Median*	*N*	*Mean*	*Median*	*N*
Panel A: Summary statistics of whole sample						
$ amount raised per deal (in million)	22.27	1.42	5,369	21.79	3	11,823
# of deals per week	81	-	67	178	-	67
Firm Age (exclude mature firms)	2.94	2.56	4,544	3.37	3	9,620
Firm Age (whole sample)	8.66	3.13	5,369	10.46	3.51	11,823
# of investors per deal	2.15	-	5,369	3.07	-	11,823
Panel B: Sub-group by firm age						
New firms (0 – 3 years):						
$ amount raised per deal (in million)	8.31	0.65	2,607	7.76	1.40	5,293
# of deals per week	38	-	67	79	-	67
Young firms (3 – 6 years):						
$ amount raised per deal (in million)	17.07	2.32	1,415	18.82	4.20	3,057
# of deals per week	22	-	67	46	-	67
Established firms (6 – 9 years):						
$ amount raised per deal (in million)	31.45	4	522	44.48	7.5	1,270
# of deals per week	8	-	67	19	-	67
Mature firms (9+ years):						
$ amount raised per deal (in million)	69.50	7.65	825	46.54	6.30	2,203
# of deals per week	13	-	67	33	-	67

Note: the # of deals summary is calculated for each week from July 2017 to September 2018. We include funding rounds with undisclosed $ raise amounts in calculating # of deals. The sample thus differs between # of deals and $ amount raise per week period.

Table 2: GDPR effect on investment in technology ventures ($ amount per deal)

	(1)	(2)	(3)
GDPR effect	-0.396***	-0.397***	-0.271*
	(0.074)	(0.074)	(0.159)
Firm Age	0.004**	0.004**	0.212***
	(0.001)	(0.001)	(0.027)
Sub-sample group	No	No	New firms (0-3 y.o.)
State FE	Yes	Yes	Yes
Week FE	Yes	Yes	Yes
Linear Trend	No	Yes	No
Observations	17,192	17,192	7,900
Adjusted R2	0.382	0.382	0.414

Note: Dependent variable is the $ amount raised of each deal in columns (1) and (2), and the log($ amount raised) of each deal in column (3) and (4). Treatment group comprises EU ventures with deals after May, 2018. The first 3 columns are for the entire sample. The sub sample in column (4) comprises only 0-3 year old ventures. The control group comprises US ventures. Standard errors are clustered by state and reported in parentheses. Fixed effects include week, US state, and EU countries. Top-coding is done at the 95% level to remove outliers. Macroeconomic controls include unemployment rate, GDP, interest rate, and CPI. ***, **, and * indicate significance at the 1%, 5%, and 10% levels.

Table 3: GDPR effect aggregate level weekly # of deals and $ MM raised amount

	(1)	(2)	(3)	(4)
GDPR effect	-0.194***	-11.574***	-0.211**	-3.204*
	(0.069)	(4.436)	(0.087)	(1.863)
Effect at post GDPR mean		-3.380***		-0.902**
		(1.074)		(0.422)
Sample	Whole Sample	Whole Sample	0-3 y.o. firms	0-3 y.o. firms
Specification	Poisson Regression	Tobit with top-coded	Poisson Regression	Tobit with top-coded
State FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
Observations	10,050	10,050	5,025	5,025

Note: Note: Dependent variable is the # of deals in column 1 and 3, aggregate $ MM raised amount in column 2 (top-coded at the 95% level to remove outliers. Treatment group comprises EU venture deals after May 2018. Control group comprises US deals. Standard errors are robust and reported in parentheses. Fixed effects include the firm's US state or EU country. Macroeconomic controls include a country's monthly interest rate, unemployment rate, GDP, and CPI. ***, **, and * indicate significance at the 1%, 5%, and 10% levels.

Table 4: Summary Statistics

Variable	Treatment Group		Control Group	
	Mean	*Std. Dev.*	*Mean*	*Std. Dev.*
Panel A: Pre-intervention (2001, 2002)				
Purchasing loan denial rate (in number)	.1387	(.0883)	.1077	(.0940)
Purchasing loan denial rate (in $000's)	.1342	(.0891)	.1043	(.1026)
Number of purchasing loans originated	123.6	(143.8)	74.01	(55.26)
Refinancing loan denial rate (in number)	.1478	(.0804)	.1392	(.1104)
Refinancing loan denial rate (in $000's)	.1487	(.0765)	.1405	(.1086)
Number of refinancing loans originated	429.7	(391.7)	272.2	(209.5)
Panel B: During intervention (2003, 2004)				
Purchasing loan denial rate (in number)	.1649	(.0826)	.1406	(.0945)
Purchasing loan denial rate (in $000's)	.1615	(.0856)	.1354	(.0919)
Number of purchasing loans originated	158.4	(177.3)	86.89	(81.84)
Refinancing loan denial rate (in number)	.1796	(.0929)	.1784	(.1286)
Refinancing loan denial rate (in $000's)	.1914	(.0951)	.1854	(.1253)
Number of refinancing loans originated	478.5	(426.0)	291.6	(244.8)
Panel C: Post-intervention (2005, 2006)				
Purchasing loan denial rate (in number)	.2173	(.0972)	.1815	(.1258)
Purchasing loan denial rate (in $000's)	.2178	(.1014)	.1782	(.1265)
Number of purchasing loans originated	170.5	(201.3)	87.65	(84.10)
Refinancing loan denial rate (in number)	.2467	(.0946)	.2280	(.1265)
Refinancing loan denial rate (in $000's)	.2604	(.0973)	.2368	(.1266)
Number of refinancing loans originated	275.3	(267.5)	131.5	(97.12)
Panel D: Control variables (2001-2006)				
Median income, % of MSA median	104.4	(45.36)	98.27	(44.52)
% of population below Poverty Line	9.295	(9.209)	10.80	(8.581)
Inside central city?	.3820	(.4859)	.7922	(.4059)
% of Minority population	47.70	(26.47)	45.18	(26.86)
% of Asian population	.1600	(.1347)	.2297	(.1998)
% of Black population	.1227	(.1851)	.0768	(.1323)
% of Hispanic population	.1643	(.1412)	.1202	(.1304)
Median age of housing stock	35.41	(13.89)	47.29	(13.91)
% of owner-occupied units	.6838	(.2056)	.5162	(.2396)
Ratio of households to housing units	.9674	(.0851)	.9459	(.0681)
Number of observations (tract x year)	3793		1304	

Note: Treatment group comprises all Census tracts in Alameda, Contra Costa, and San Mateo Counties; Control group comprises all Census tracts in Marin and San Francisco Counties. All data come from Federal Financial Institutions Examination Council.

Table 5: Difference-in-Difference Models (Purchase Loans)

	Denial Rate in Number			Denial Rate in Amount		
Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Treatment	–.0087**	–.0107**	–.0102**	–.0090**	–.0111**	–.0109**
	(.0040)	(.0042)	(.0043)	(.0042)	(.0043)	(.0044)
Median income %		–.0001	–.0002		–.0001	–.0002
		(.0001)	(.0001)		(.0001)	(.0001)
Below poverty %		–.0002	–.0002		.0000	.0001
		(.0005)	(.0005)		(.0006)	(.0006)
Inside city?		–.0017	–.0016		–.0010	–.0006
		(.0033)	(.0033)		(.0035)	(.0035)
Minority %		.0006***			.0006***	
		(.0002)			(.0002)	
Asian %			.0679*			.0652*
			(.0351)			(.0367)
Black %			.0647*			.0566
			(.0383)			(.0376)
Hispanic %			.0537			.0710*
			(.0356)			(.0375)
Median house age			–.0004			–.0002
			(.0004)			(.0005)
Owner-occupied %			–.0007			.0176
			(.0225)			(.0248)
HHD to housing			–.0228			–.0109
			(.0442)			(.0461)
Tract dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	5,055	5,055	5,035	5,055	5,055	5,035
R^2	.7590	.7597	.7607	.7580	.7588	.7596

Note: All specifications are Weighted Least Squares where the weight is the number of total purchase loan applications (the denominator). Standard errors are clustered by tract and reported in the parentheses. Significance level: ***1%, **5%, *10%.

Table 6: Difference-in-Difference Models (Refinancing Loans)

	Denial Rate in Number			Denial Rate in Amount		
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	–.0030	–.0051**	–.0051**	–.0034	–.0059**	–.0062**
	(.0020)	(.0021)	(.0022)	(.0023)	(.0024)	(.0025)
Median income %		–.0001	–.0001		–.0000	–.0000
		(.0000)	(.0001)		(.0001)	(.0001)
Below poverty %		–.0005	–.0006		–.0004	–.0004
		(.0004)	(.0004)		(.0004)	(.0004)
Inside city?		.0047**	.0044*		.0038*	.0036
		(.0020)	(.0020)		(.0023)	(.0023)
Minority %		.0005***			.0006***	
		(.0001)			(.0001)	
Asian %			.0718***			.0813***
			(.0188)			(.0224)
Black %			.0419*			.0441*
			(.0221)			(.0239)
Hispanic %			.0556***			.0897***
			(.0196)			(.0210)
Median house age			–.0001			–.0002
			(.0002)			(.0003)
Owner-occupied %			–.0072			–.0079
			(.0197)			(.0216)
HHD to housing			–.0203			–.0113
			(.0309)			(.0341)
Tract dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	5,084	5,084	5,054	5,083	5,083	5,053
R^2	.9117	.9126	.9130	.8919	.8928	.8933

Note: All specifications are Weighted Least Squares where the weight is the number of total refinancing loan applications (the denominator). Standard errors are clustered by tract and reported in the parentheses. Significance level: ***1%, **5%, *10%.

Table 7: Linear Probability Models (Loan Rejections)

	Whole Sample		Conforming Loans		Jumbo Loans	
Variable	**Purch.**	**Refi.**	**Purch.**	**Refi.**	**Purch.**	**Refi.**
Treatment	–.0146***	–.0028	–.0141*	.0026	–.0153***	–.0098***
	(.0050)	(.0023)	(.0074)	(.0022)	(.0040)	(.0036)
Appl. Native	.0392***	.0677***	.0376***	.0685***	.0381***	.0703***
	(.0097)	(.0082)	(.0108)	(.0085)	(.0086)	(.0083)
Appl. Asian	.0277***	.0205***	.0259***	.0215***	.0273***	.0216***
	(.0029)	(.0035)	(.0037)	(.0033)	(.0035)	(.0040)
Appl. African	.0705***	.0677***	.0803***	.0681***	.0548***	.0753***
	(.0058)	(.0058)	(.0061)	(.0064)	(.0087)	(.0071)
Appl. Pacific	.0362***	.0523***	.0325***	.0544***	.0402***	.0616***
	(.0089)	(.0093)	(.0097)	(.0097)	(.0080)	(.0084)
Appl. male	–.0055***	–.0109***	.0002	–.0092***	–.0098***	–.0189***
	(.0016)	(.0024)	(.0020)	(.0025)	(.0017)	(.0020)
MTI	.0016***	.0025***	.0014	.0033***	.0018***	.0020***
	(.0006)	(.0008)	(.0012)	(.0013)	(.0007)	(.0006)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
County dummies	Yes	Yes	Yes	Yes	Yes	Yes
Institution dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	529,913	1,267,089	260,011	718,485	269,902	548,604
R^2	.1188	.1923	.1155	.2025	.1261	.1914

Note: All specifications are Ordinary Least Squares with a full set of dummy variables. Standard errors are clustered by institution and reported in the parentheses. Significance level: ***1%, **5%, *10%.

Table 8: Reduced-Form Models (Foreclosure Starts)

	Estimated Foreclosure Start Rate					
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Treatment counties	.0708**	.0130***	.0116**			
	(.0158)	(.0010)	(.0029)			
Treat x 0304 share				.2996***	.0547***	.0452***
				(.0581)	(.0033)	(.0080)
2003-04 share				–.0780	–.0158	–.0359
				(.1055)	(.0080)	(.0172)
House price change		–.2453***	–.1097		–.2534***	–.1229*
		(.0345)	(.0519)		(.0372)	(.0472)
Unemployment rate		–.1784*	1.410*		–.2789	1.292
		(.0761)	(.6153)		(.1447)	(.6157)
Low-cost high-lev %		–.1509***	–.1708***		–.1522***	–.1715***
		(.0110)	(.0134)		(.0096)	(.0121)
high-cost low-lev %		.2515***	.2560***		.2192***	.2514***
		(.0366)	(.0251)		(.0395)	(.0226)
high-cost high-lev %		.3863***	.3857***		.3897***	.3826***
		(.0348)	(.0128)		(.0331)	(.0132)
Median income %			–.0000**			–.0000
			(.0000)			(.0000)
Below poverty %			–.0002			–.0002
			(.0003)			(.0002)
Inside city?			–.0084**			–.0084**
			(.0022)			(.0022)
Asian %			–.0029			–.0026
			(.0106)			(.0108)
Black %			–.0073			–.0059
			(.0044)			(.0043)
Hispanic %			.0122			.0144
			(.0185)			(.0179)
Median house age			–.0002***			–.0002***
			(.0000)			(.0000)
Owner-occupied %			.0028			.0015
			(.0046)			(.0040)
HHD to housing			.0136			.0188
			(.0110)			(.0115)
N	868	868	865	868	868	865
R^2	.3269	.9561	.9715	.3602	.9568	.9710

Note: All specifications are Weighted Least Squares where the weight is the HUD estimated number of mortgages. Standard errors are clustered by county and reported in the parentheses. Significance level: ***1%, **5%, *10%.

Appendix B:
Data Restrictions and Prices

Appendix B: Data Restrictions and Prices

Modeling Data Collection

The supply side of the market is composed of two types of symmetric firms, A and B. Firms of type A offer financial products such as home loans (good A), and firms of type B offer related products, such as personal credit lines and insurance policies (good B). All firms are risk neutral and maximize expected profits. The demand side of the market consists of a continuum of ex-ante identical individuals with measure $M > 1$. A fraction of this mass, normalized to 1, is applying to purchase a loan from firm A. Individuals are assumed to be risk neutral and have unit demands (separately) for goods A and B. Let v_A and v_B denote the incremental utilities from consuming one unit of good A and B, respectively.

Suppose that there is uncertainty about cost-relevant consumer characteristics. Specifically, the cost of serving a consumer either turns out to be low (c_L^m) or high (c_H^m), with $c_L^m < c_H^m$ for $m \in \{A, B\}$. For technical simplicity, we assume that consumer types for firm B's products are perfectly correlated with those of firm A's.[12] We assume that $c_L^m < v_m < c_H^m$ for $m \in \{A, B\}$; that is, for both goods, it is efficient to serve only low-cost consumers. However, if screening is imperfect, firms may end up approving some high-cost consumers. The proportion of high-cost consumers in the population is $\lambda > 0$. At the onset, information is incomplete and symmetric, and, in particular, consumers and firms do not observe the realization of their types.

The game unfolds in several stages. First, each firm j of type A announces a price, $p_{j,A} \in \mathbb{R}_+$, at which it will sell the good to a consumer whose application is ultimately approved. Price announcements are made publicly and simultaneously. Next, each consumer applies to purchase the good from a firm A of his choice. Then, each firm of type A acquires information about its applicants and chooses which applicants to qualify. After selling to qualified applicants, each firm of type A makes a take-it-or-leave-it offer to a firm B for purchasing its list of applicants and the information it acquired about them, including whether they were approved. Firm B then decides whether to accept or reject this offer and proceeds to make targeted sale offers to potential customers.

Data Acquisition

A firm A chooses a sample size, or search intensity $n \geq 0$, which is treated for simplicity as a continuous variable. The cost to the firm of acquiring information about an applicant is kn, where $k > 0$. By choosing a search intensity n, firm A receives n conditionally independent Bernoulli signals, $\{X_1, \ldots, X_n\}$, where

$$\Pr\{X_i = 1 | c^A\} = \begin{cases} 1, & \text{if } c^A = c_L^A, \\ 1 - \alpha, & \text{if } c^A = c_H^A \end{cases}$$

The parameter $\alpha \in (0, 1)$ represents intrinsic signal strength. If $\alpha = 1$, then a signal is fully informative, and if $\alpha = 0$, then signals contain no information. This process is interpreted as follows: A firm A chooses a search report containing $i = 1, \ldots, n$ records, $\{X_1, \ldots, X_n\}$, for each of its applicants, and each record is either positive ($X_i = 1$) or negative ($X_i = 0$). Since the firm is in effect searching for bad news about its applicants' creditworthiness, it is possible to summarize all the information contained in an applicant's search report with the sufficient statistic $S_n \equiv \min\{X_1, \ldots, X_n\}$. That is, if $S_n = 0$, then at least one of the records was negative, and the applicant (now disqualified) is certainly type c_H^A; whereas if $S_n = 1$, then all records were positive, and the applicant (referred to as qualified) is type c_H^A with probability

$$\mu(n) = \frac{\lambda(1-\alpha)^n}{\lambda(1-\alpha)^n + (1-\lambda)} < \lambda \tag{1}$$

and type c_L^A with the complementary probability $1 - \mu(n)$.

After acquiring information about a consumer, a firm j of type A decides whether to approve the consumer's application (i.e., sell him the good at its posted price $p_{j,A}$). Approval results in an expected payoff of $p_{j,A} - E[c^A | S_n] - kn$ for the firm. Rejection results in a payoff of zero for the consumer and

[12] Incorporating imperfect correlation would not change our qualitative results but it would complicate the exposition without a major gain in intuition.

$-kn$ for the firm. A measure of the efficacy of firm A's information-acquisition technology is thus given as follows:[13] $m \equiv -\frac{k}{\ln(1-\alpha)}$.

The Value of Data

To focus on the effects of information trade, assume that firms of type B do not possess the technology to independently acquire information about potential customers. This is consistent with the observation that credit card and insurance policies are often approved without pursuing high levels of direct screening. Moreover, issuers often make pre-screened offers based on summary information (e.g., credit scores) that they purchase from primary sources. Suppose that a firm of type B (henceforth, firm B) randomly targets a fraction of the total mass of potential consumers, M. Let ξ denote the probability that a customer in firm B's initial target set overlaps with a firm of type A's approved set of applicants.

The benefit to firm B of learning about a *qualified* consumer from A's list is the following: With probability ξ, this consumer is already targeted by firm B, whereby B's benefit is zero. With probability $1-\xi$, this consumer would not have been targeted by B, in which case B's benefit is $E[c^B]-E[c^B|S_n=1]$ in expected cost savings (where $E[c^B]=\lambda c_H^B+(1-\lambda)c_L^B$ and $E[c^B|S_n=1]=\mu(n)c_H^B+(1-\mu(n))c_L^B$). Thus, firm B's overall expected benefit from learning about a qualified consumer is $(1-\xi)(E[c^B]-E[c^B|S_n=1])$. Analogously, firm B's expected benefit from learning about a randomly-selected *disqualified* consumer is given by $\xi(c_H^B-E[c^B])$. Formally, firm B's willingness to pay per application contained in firm A's applicant list (given a level of screening intensity n) is given as follows:[14]

$$\underbrace{[(1-\lambda)+\lambda(1-\alpha)^n](1-\xi)(E[c^B]-E[c^B|S_n=1])}_{\text{Qualified consumer information}}+\underbrace{\left(1-[(1-\lambda)+\lambda(1-\alpha)^n]\right)\xi(c_H^B-E[c^B])}_{\text{Disqualified consumer information}}.$$

Substituting for $\mu(n)$ from (1) and simplifying yields

$$(1-\lambda)\lambda(1-(1-\alpha)^n)(c_H^B-c_L^B). \tag{2}$$

From the expression in (2), firm B's expected benefit from information about firm A's applicant list is independent of ξ and is increasing in n. Intuitively, firm B benefits from firm A's information by being able to fine tune its target customer set. Specifically, a higher screening intensity enables B to more effectively avoid high-cost consumers and to better target low-cost consumers. Since the overall benefit is a combination of improving the targeting of individual consumers both in and out of B's initial target set, it is independent of the likelihood that A's information directly overlaps with this initial set.

Equilibrium in a Competitive Market

In the following analysis, it is assumed that the information-acquisition technology for firms of type A is sufficiently effective (i.e., m is sufficiently small) so that an interior solution obtains with positive screening intensity (i.e., $n^*>0$).[15] The super- and sub-scripts T and NT denote cases with and without information trade, respectively. For notational simplicity, the subscript j is omitted when referring to a representative firm of type A. Lemma 1 derives a preliminary result that hold in equilibrium both when information trade is permitted and when it is not.

Lemma 1 *In a symmetric equilibrium, a consumer's expected utility is decreasing in firm A's price, p_A.*

Proof. Whether or not trade is permitted, a consumer's expected utility from applying to purchase the good at a price p_A is

$$U(p_A,n)=(\lambda(1-\alpha)^n+(1-\lambda))(v_A-p_A).$$

Without information trade, firm A's expected profit from each of its applicants is given by

$$\Pi^{NT}(p_A,n)=(\lambda(1-\alpha)^n+(1-\lambda))p_A-(1-\lambda)c_L^A-\lambda(1-\alpha)^n c_H^A-kn.$$

[13] Lower values of m correspond to better technologies involving low sampling costs and/or a high intrinsic signal strength. The logarithm term is due to differentiating the probability $(1-\alpha)^n$ with respect to n.

[14] Firm B could directly swap out a disqualified consumer and swap in a qualified consumer. This would generate the same expected benefit as when firm B swaps out a disqualified consumer with a new consumer (for whom it has no information), and swaps out a consumer for whom it has no information with a qualified consumer from outside its simple.

[15] A condition sufficient to guarantee this is that $m<\lambda(c_H^A-v_A)$. This follows given that a firm A's expected profit without information trade is $(\lambda(1-\alpha)^n+(1-\lambda))p_A-(1-\lambda)c_L^A-\lambda(1-\alpha)^n c_H^A-kn$ and that its derivative evaluated at $n=0$ has to be positive. The condition for an interior solution with information trade is weaker. See the proof of Lemma 1.

The first-order condition with respect to n (given p_A) is

$$\frac{\partial \Pi^{NT}}{\partial n} = \lambda \ln(1-\alpha)(p_A - c_H^A)(1-\alpha)^n - k \leq 0.$$

Since $\frac{\partial^2 \Pi^{NT}}{\partial n^2} = \lambda(\ln(1-\alpha))^2(p_A - c_H^A)(1-\alpha)^n < 0$, an interior solution obtains if $\frac{\partial \Pi^{NT}}{\partial n} > 0$ at $n = 0$, or $m < \lambda(c_H^A - v_A)$. Thus, the expected utility assuming interior solutions can be rewritten as

$$U^{NT}(p_A, n) = [\frac{m}{c_H^A - p_A} + (1-\lambda)](v_A - p_A).$$

Differentiating with respect to p_A yields

$$\frac{dU^{NT}}{dp_A} = \frac{m(v_A - c_H^A)}{(c_H^A - p_A)^2} - (1-\lambda) < 0.$$

Similarly, in the case with trade, firm A's profit per application is given by

$$\Pi^T = (\lambda(1-\alpha)^n + (1-\lambda))p_A - (1-\lambda)c_L^A - \lambda(1-\alpha)^n c_H^A - kn+$$
$$(1-\lambda)\lambda(1-(1-\alpha)^n)(c_H^B - c_L^B).$$

The first-order condition with respect to n (given p_A) is

$$\frac{\partial \Pi^T}{\partial n} = \lambda \ln(1-\alpha)(1-\alpha)^n[p_A - c_H^A - (1-\lambda)(c_H^B - c_L^B)] - k \leq 0.$$

Since $\frac{\partial^2 \Pi^T}{\partial n^2} = \lambda[\ln(1-\alpha)]^2(1-\alpha)^n[p_A - c_H^A - (1-\lambda)(c_H^B - c_L^B)] < 0$, an interior solution obtains if $\frac{\partial \Pi^T}{\partial n} > 0$ at $n = 0$, or $m < \lambda[c_H^A - v_A + (1-\lambda)(c_H^B - c_L^B)]$. Expected utility when information trade is permitted is thus given by

$$U^T(p_A, n) = [\frac{m}{c_H^A - p_A + (1-\lambda)(c_H^B - c_L^B)} + (1-\lambda)](v_A - p_A).$$

Differentiating with respect p_A yields

$$\frac{dU^T(p_A, n)}{dp_A} = \frac{m[v_A - c_H^A - (1-\lambda)(c_H^A - c_L^A)]}{[c_H^A - p_A + (1-\lambda)(c_H^B - c_L^B)]^2} - (1-\lambda) < 0,$$

which gives the result. ■

Lemma 1 states that, taking into account a firm A's subsequent choice of screening intensity, an applicant's expected utility decreases monotonically in firm A's posted price. It follows that consumers choose to apply to a firm A that posts the lowest price. As the next result indicates, this choice has an important implication for a firm's level of consumer screening. In particular, the firms that post the lowest prices are also those that will be screening consumers most intensely. Our first result is the following:

Proposition 2 *With information trade, (i) consumers purchase from a firm A at lower prices, (ii) more information is collected about consumers, and (iii) more consumers are disqualified from purchasing good A relative to when consumer information cannot be traded*

Proof. The proof proceeds in two steps.

Step 1: First, it is shown that for a given p_A, under information trade, more information is collected about consumers and the amount of information collected is decreasing in p_A.

Without information trade, given a price p_A, firm A's optimal search intensity $n^*_{NT}(p_A)$ is given by

$$n^*_{NT}(p_A) = \frac{\ln m - \ln \lambda - \ln(c_H^A - p_A)}{\ln(1-\alpha)}.$$

With trade, given a price p_A, its optimal search intensity $n^*_T(p_A)$ is

$$n^*_T(p_A) = \frac{\ln m - \ln \lambda - \ln[c_H^A - p_A + (1-\lambda)(c_H^B - c_L^B)]}{\ln(1-\alpha)}.$$

Only the last term in the numerators need be compared. Since $\ln(c_H^A - p_A) < \ln[c_H^A - p_A + (1-\lambda)(c_H^B - c_L^B)]$ and the denominator $\ln(1-\alpha)$ is negative for $\alpha \in (0,1)$, it holds that $\frac{-\ln(c_H^A - p_A)}{\ln(1-\alpha)} < \frac{-\ln[c_H^A - p_A + (1-\lambda)(c_H^B - c_L^B)]}{\ln(1-\alpha)}$. Hence, it follows that $n_{NT}^*(p_A) < n_T^*(p_A)$. If $p_{A,T} < p_{A,NT}$, then the gap between $c_H^A - p_{A,NT}$ and $c_H^A - p_{A,T} + (1-\lambda)(c_H^B - c_L^B)$ expands.

From the first-order condition of the firm's expected profit without trade,

$$\lambda(1-\alpha)^{n_{NT}(p_{A,NT})}(c_H^A - p_{A,NT}) = m,$$

it follows that $n_{NT}^*(p_{A,NT})$ is decreasing in $p_{A,NT}$.

Similarly, from the first-order condition of the firm's expected profit with trade,

$$\lambda(1-\alpha)^{n_T(p_{A,T})}[c_H^A - p_{A,T} + (1-\lambda)(c_H^B - c_L^B)] = m,$$

it follows that $n_T^*(p_{A,T})$ is decreasing in $p_{A,T}$.

Step 2: Next, it is shown that under information trade, consumers purchase from firms A at a lower price.

Since firms are competing in price, equilibrium profits will be driven down to zero. Therefore, $p_{A,NT}^*$ and $p_{A,T}^*$, along with $n_{NT}^* = n_{NT}(p_{A,NT})$ and $n_T^* = n_T(p_{A,T})$, satisfy

$$p_{A,NT}^* = \frac{(1-\lambda)c_L^A + \lambda(1-\alpha)^{n_{NT}^*}c_H^A + kn_{NT}^*}{\lambda(1-\alpha)^{n_{NT}^*} + (1-\lambda)}$$

and

$$p_{A,T}^* = \frac{(1-\lambda)c_L^A + \lambda(1-\alpha)^{n_T^*}c_H^A + kn_T^* - (1-\lambda)\lambda[1-(1-\alpha)^{n_T^*}](c_H^B - c_L^B)}{\lambda(1-\alpha)^{n_T^*} + (1-\lambda)}.$$

Since $\Pi_A^{NT}(p_{A,NT}^*, n_{NT}^*) = 0$, once information trade is permitted, profits are positive at the price and search intensity $(p_{A,NT}^*, n_{NT}^*)$; that is, $\Pi_A^T(p_{A,NT}^*, n_{NT}^*) > 0$. By Lemma 1, consumers apply to the lowest priced firm. It follows that $p_{A,T}^* < p_{A,NT}^*$ must be satisfied in equilibrium. Combined with Step 1, it further follows that $n_T^* > n_{NT}^*$ in a symmetric equilibrium. ■

Proposition 1 points to an inverse relationship between price and screening intensity. From the standpoint of firms, price competition dissipates profits from selling good A. However, once information trade is permitted, firms A are able to lower their prices even further due to profits from selling applicants' information. This price reduction is coupled with a stricter screening of applicants. On the one hand, consumers benefit from lower posted prices; on the other hand, more consumers are disqualified compared to the case where information trade is prohibited. The next result addresses allocative efficiency:

Proposition 3 *There exists a constant $\bar{c} > 0$, such that for $c_H^B - c_L^B \geq \bar{c}$, allowing for information trade strictly increases ex-ante social welfare.*

Proof. The proof proceeds in two steps.

Step 1: It is first shown that there exists a lower bound $b(c_L^B, c_H^B)$ on the difference in the equilibrium prices with and without trade, $p_{A,NT}^* - p_{A,T}^*$.

For a given search intensity n, the gain from information trade is $(1-\lambda)\lambda(1-(1-\alpha)^n)(c_H^B - c_L^B)$. Since firm A's expected profits are zero in equilibrium without trade, the following relationship must hold when information trade is permitted:

$$p_{A,NT}^* - p_{A,T}^* \geq \frac{(1-\lambda)\lambda(1-(1-\alpha)^{n_{NT}^*})(c_H^B - c_L^B)}{\lambda(1-\alpha)^{n_{NT}^*} + 1 - \lambda},$$

else a firm A possesses a profitable deviation by undercutting its competitors and attracting all consumers while benefiting from a positive expected profit.

Step 2: Next, social welfare with and without information trade are compared.

Welfare without trade, W_{NT}, is given by

$$W_{NT} = (\lambda(1-\alpha)^{n_{NT}} + (1-\lambda))v_A - (1-\lambda)c_L^A - \lambda(1-\alpha)^{n_{NT}}c_H^A - kn_{NT}.$$

Welfare with trade, W_T, is given by

$$W_T = (\lambda(1-\alpha)^{n_T} + (1-\lambda))v_A - (1-\lambda)c_L^A - \lambda(1-\alpha)^{n_T} c_H^A - kn_T + (1-\lambda)\lambda(1-(1-\alpha)^{n_T})(c_H^B - c_L^B).$$

From firm A's first-order conditions, it holds that $\lambda(1-\alpha)^{n^*_{NT}}(c_H^A - p^*_{A,NT}) = m$ and $\lambda(1-\alpha)^{n^*_T}[c_H^A - p^*_{A,T} + (1-\lambda)(c_H^B - c_L^B)] = m$ for each case. Substituting n^*_{NT} and n^*_T, welfare with information trade is greater than without trade when

$$\frac{m(v_A - p^*_{A,T})}{c_H^A - p^*_{A,T} + (1-\lambda)(c_H^B - c_L^B)} + (1-\lambda)(v_A - p^*_{A,T}) \geq \frac{m(v_A - p^*_{A,NT})}{c_H^A - p^*_{A,NT}} + (1-\lambda)(v_A - p^*_{A,NT}).$$

Since $m < \lambda(c_H - v_A)$ by assumption (interior solutions), a sufficient condition for the above inequality to hold is

$$\frac{1-\lambda}{\lambda(c_H^A - v_A)}(p^*_{A,NT} - p^*_{A,T}) \geq \frac{v_A - p^*_{A,NT}}{c_H^A - p^*_{A,NT}} - \frac{v_A - p^*_{A,T}}{c_H^A - p^*_{A,T} + (1-\lambda)(c_H^B - c_L^B)}.$$

This inequality can be rewritten as

$$(1-\lambda)\frac{c_H^A - p^*_{A,NT}}{c_H^A - v_A}(p^*_{A,NT} - p^*_{A,T})(c_H^A - p^*_{A,T} + (1-\lambda)(c_H^B - c_L^B)) + \lambda(p^*_{A,NT} - p^*_{A,T})(c_H^A - v_A)$$

$$\geq \lambda(v_A - p^*_{A,NT})(1-\lambda)(c_H^B - c_L^B).$$

Since $p^*_{A,NT} > p^*_{A,T}$ and $0 < c_H - v_A < c_H - p^*_{A,NT}$, whereby $\frac{c_H - P_{A,NT}}{c_H - v_A} > 1$, a sufficient condition for the above is

$$p^*_{A,NT} - p^*_{A,T} \geq \lambda(v_A - p^*_{A,NT}).$$

Using $p^*_{A,NT} - p^*_{A,T} \geq \frac{(1-\lambda)\lambda(1-(1-\alpha)^{n^*_{NT}})(c_H^B - c_L^B)}{\lambda(1-\alpha)^{n^*_{NT}} + 1 - \lambda}$ from Step 1 and simplifying, a sufficient condition is given by

$$c_H^B - c_L^B \geq \frac{(1-\lambda+\lambda(1-\alpha)^{n^*_{NT}})(v_A - p^*_{A,NT})}{(1-\lambda)(1-(1-\alpha)^{n^*_{NT}})} \equiv \bar{c}.$$

The right-hand side of the above does not depend on $c_H^B - c_L^B$, giving a lower bound $\bar{c}$. ■

That is, when firm B's benefit from applicant information for product A is significant, the price reductions consumers enjoy ex ante for good A offset consumers' disutility from higher rates of ex-post rejections. If $c_H^B - c_L^B$ is sufficiently high, qualified consumers can even be paid to purchase good A; that is, the price for good A when information trade is permitted, $p^*_{A,T}$, could be negative. Intuitively, price commitments, whereby screening is conducted after consumers apply, lead firms selling good A to compete away their downstream profits from selling their applicants' information. In turn, more consumer data is collected and traded, more consumers are rejected from buying good A, but those who are ultimately approved enjoy significant discounts.

Thus far, it has been assumed that consumers differ only in terms of sellers' costs of serving their respective types, where all consumers derive the same utilities from acquiring sellers' products. However, consumers' tastes may differ, whereby each consumer type derives a different value v_A from purchasing good A. For example, net utility from a loan may reflect a consumer's likelihood of default, where high-cost consumers are more likely to experience disutility from meeting mortgage payments and risk default incurring additional costs. That is, let $v_{A,H}$ and $v_{A,L}$ denote a high and low type's incremental utility, respectively. It may be the case that $v_A = v_{A,H} = v_{A,L}$ (as in the base model) no longer holds. The preceeding analysis readily extends to the case in which high-cost consumers experience a higher level of expected disutility from default (and thus place a lower valuation on good A) than low-cost consumers, that is, $v_{A,H} < v_{A,L}$.

Proposition 4 *Suppose high-cost consumers derive a lower utility from good A than low-cost consumers. Then, allowing for information trade leads to higher ex-ante social welfare.*

Proof. The proof proceeds in two steps.

Step 1: It is first shown that a consumer's expected utility from applying to purchase good A is decreasing in the price P_A. To save notation, let $v_A = v_{A,L}$.

Whether or not information trade is permitted, a consumer's expected utility from applying to a firm A is now given by

$$U(p_A, n) = (\lambda(1-\alpha)^n + (1-\lambda))\,(v_A - p_A) - \lambda(1-\alpha)^n(v_A - v_{A,H}),$$

and for a given p_A a firm A chooses $n(p_A)$ from its first-order condition, $\lambda(1-\alpha)^n(c_H^A - p_A) = m$ (without trade) and $\lambda(1-\alpha)^n[c_H^A - p_A + (1-\lambda)(c_H^B - c_L^B)] = m$ (with trade). Thus, a consumer's expected utility is

$$U(p_A, n) = \left(\frac{m}{c_H^A - p_A + \mathbf{1}\{\text{trade}\}(1-\lambda)(c_H^B - c_L^B)} + (1-\lambda)\right)(v_{A,H} - p_A) + (1-\lambda)(v_A - v_{A,H}),$$

where $\mathbf{1}\{\text{trade}\}$ is the indicator function for information trade. Differentiation with respect to p_A yields

$$\frac{dU}{dp_A} = -\frac{m(c_H^A - v_{A,H} + \mathbf{1}\{\text{trade}\}(1-\lambda)(c_H^B - c_L^B))}{(c_H^A - p_A + \mathbf{1}\{\text{trade}\}(1-\lambda)(c_H^B - c_L^B))^2} - (1-\lambda),$$

which is negative since $c_H > v_{A,H}$ and $c_H > p_A$.

Step 2: Next, it is shown that a symmetric equilibrium consists of an analogous characterization to the base model. Notice that the planner's problem is equivalent to the monopolist's problem.

Without information trade, the highest price that a monopolist can charge has to satisfy consumers' participation constraints, $(1-\lambda)(v_A - p_A) + \lambda(1-\alpha)^n(v_{A,H} - p_A) \geq 0$. In equilibrium,

$$p_A^* = \frac{(1-\lambda)v_A + \lambda(1-\alpha)^n v_{A,H}}{1-\lambda+\lambda(1-\alpha)^n}.$$

The monopolist's profit is given by $(1-\lambda+\lambda(1-\alpha)^n)p_A - (1-\lambda)c_L^A - \lambda(1-\alpha)^n c_H^A - kn$. Substituting for p_A^*, the monopolist's profit can be rewritten as $\lambda(1-\alpha)^n(v_{A,H} - c_H^A) + (1-\lambda)(v_A - c_L^A) - kn$. The first term is negative and represents the cost of selling to high-cost consumers. Differentiating to obtain the first-order condition gives

$$\lambda(1-\alpha)^n(c_H^A - v_{A,H}) = m$$

A simple comparison between the optimal search condition employed by the monopolist above and the optimal search condition employed by a firm A in equilibrium, given by $\lambda(1-\alpha)^n(c_H^A - p_{A,NT}) = m$, reveals that in equilibrium applicants are not sufficiently screened relative to the efficient level since $p_A > v_{A,H}$.

It is straightforward to show that the problem is analogous in the case of information trade. That is, comparing the optimal search condition employed by a monopolist, $\lambda(1-\alpha)^n(c_H^A - v_{A,H} + (1-\lambda)(c_H^B - c_L^B)) = m$, to the optimal search condition employed by a firm A, given by $\lambda(1-\alpha)^n(c_H^A - p_A + (1-\lambda)(c_H^B - c_L^B)) = m$, reveals the following: Since $p_{A,T} < p_{A,NT}$ is satisfied by Proposition 1, allowing for information trade would move the outcome closer to the social optimum. ∎

Proposition 3 extends the findings in the baseline case by showing that information trade brings the outcome closer to the social planner's solution when taking into account high-cost consumers' higher likelihood of (and hence disutility from) default. In such cases, allowing for information trade is even more important in terms of improving welfare. This is because the social cost of misallocating good A (that is, of mistakenly qualifying high-cost consumers) is now strictly higher. Since firms screen applicants more intensely when information trade is permitted, allocative efficiency improves due to a greater likelihood of avoiding more costly defaults. In other words, when factoring in consumers' disutilities from default, there are greater benefits associated with not restricting data flows.

Appendix C :
Data Restrictions and Welfare

Appendix B: Data Restrictions and Welfare

Analysis of Four Microeconomic Models

Model 1: Linear City Model

Consider a linear city model on the unit interval, where firm A is located at 0 and firm B at 1. Both firms' unit costs are $c > 0$, and consumers' locations (preferences), $\alpha \in [0, 1]$, specify their distances from 0 and are uniformly distributed. Consumers have unit-demands with valuations v, incur transportation costs t per unit distance, and the market is assumed to be covered in equilibrium, which is ensured by $v > c + \frac{3t}{2}$.

Equilibrium with Privacy: When firms have no information about consumers' types, they set uniform prices. A consumer of type α^* is indifferent between purchasing from firms A and B if and only if $v - p_A - t\alpha^* = v - p_B - t(1 - \alpha^*)$. That is, given

$$\alpha^* = \frac{1}{2} + \frac{p_B - p_A}{2t}. \tag{3}$$

Consumers located below (above) α^* purchase from firm A (firm B). Taking the marginal consumer into account, firms A and B maximize profits with their objectives specified by $\max_{p_A} \pi_A = \alpha^*(p_A - c)$ and $\max_{p_B} \pi_B = (1 - \alpha^*)(p_B - c)$, respectively.

Proposition 5 *In equilibrium with privacy, prices satisfy $p_A^* = p_B^* = c + t$ and the marginal type is $\alpha^* = \frac{1}{2}$. Profits satisfy $\pi_A + \pi_B = t$, consumer surplus is $v - c - \frac{5t}{4}$, and the outcome is efficient. The minimum and maximum consumer utilities are $U(\frac{1}{2}) = v - c - \frac{3t}{2}$ and $U(0) = U(1) = v - c - t$, respectively.*

Proof. Substituting (3) into $\max_{p_A} \pi_A = \alpha^*(p_A - c)$ and $\max_{p_B} \pi_B = (1 - \alpha^*)(p_B - c)$ and taking the first-order conditions yields $p_A = (c+t+p_B)/2$ and $p_B = (c+t+p_A)/2$. Solving for the equilibrium prices yields $p_A^* = p_B^* = c+t$, resulting in $\alpha^* = \frac{1}{2}$ and $\pi_A = \pi_B = \frac{t}{2}$, whereas $CS = 2\int_0^{\frac{1}{2}}[v - c - t - t\alpha]d\alpha = v - c - \frac{5t}{4}$. Since all consumers buy from the closer firm, the outcome is efficient. ■

Equilibrium without Privacy: If consumer types are common knowledge and arbitrage is infeasible, then firms compete for each consumer, and prices are driven downward as follows:

	$p_A(\alpha)$	$p_B(\alpha)$
$\alpha \le 0.5$	$c + t(1 - 2\alpha)$	c
$\alpha \ge 0.5$	c	$c + t(2\alpha - 1)$

As indicated above, the resultant prices are the cost of production plus the difference in transportation costs.

Proposition 6 *In equilibrium without privacy, profits satisfy $\pi_A + \pi_B = \frac{t}{2}$, consumer surplus is given by $v - c - \frac{3t}{4}$, and the outcome is efficient. The minimum and maximum consumer utilities are $U(0) = U(1) = v - c - t$ and $U(\frac{1}{2}) = v - c - \frac{t}{2}$, respectively.*

Proof. In equilibrium, $\pi_A = \pi_B = \int_0^{\frac{1}{2}} t(1 - 2\alpha)d\alpha = \frac{t}{4}$ and $CS = 2\int_0^{\frac{1}{2}}[v - c - t(1 - \alpha)]d\alpha = v - c - \frac{3t}{4}$. Since all consumers buy from the closer firm, the outcome is efficient. ■

Notice that in comparison to the outcome with complete privacy where firms charge uniform prices, all consumers are better off with individualized pricing (consumers located at points 0 and 1 are offered the same prices under both privacy regimes and are indifferent, whereas other consumers are strictly better off without privacy). Rather than compete for the marginal consumer, firms now compete for each consumer on an individual basis. Consequently, prices decrease and some rents are transferred from firms to consumers.

Model 2: Circular City Model

Consider a circular city model with unit circumference and identical firms with unit production costs $c > 0$ and entry costs $f > 0$. Firms are located equidistant from each other. A unit mass of consumers is uniformly distributed along the circle. Consumers continue to have unit-demands with valuations v, incur linear transportation costs t per unit distance, and the market is assumed to be covered in equilibrium; that is, $v > c + \frac{3}{2}\sqrt{tf}$.

Equilibrium with Privacy: When firms have no information about consumers' types, they set uniform prices. Suppose there are n firms and consider a firm i and its nearest clockwise neighbor along the circle, j. Refer to firm i's address as 0 and firm j's as $\frac{1}{n}$. A consumer with an address $\alpha \in (0, \frac{1}{n})$ is indifferent between purchasing from firms i and j if and only if $v - p_i - t\alpha = v - p_j - t(\frac{1}{n} - \alpha)$. That is, the marginal consumer is located at address $\alpha = \frac{1}{2n} + \frac{p_j - p_i}{2t}$. Suppose that firm i's nearest clockwise and counter-clockwise competing neighbors charge a price p. On the relevant range for p, firm i's demand is given by

$$D_i(p_i, p) = 2\alpha = \frac{1}{n} + \frac{p - p_i}{t} \tag{4}$$

Taking its demand into account, firm i's objective is the following:

$$\max_{p_i} \pi_i = (p_i - c)\left(\frac{1}{n} + \frac{p - p_i}{t}\right) - f. \tag{5}$$

Let n^*_{pr} denote the equilibrium number of firms under the outcome with privacy. Then the following holds in equilibrium.

Proposition 7 *In equilibrium with privacy, firms' prices satisfy $p^* = c + \frac{t}{n}$, with $n^*_{pr} = \sqrt{t/f}$ firms entering and realizing zero profits. Consumer surplus is $v - c - \frac{5}{4}\sqrt{tf}$, and the outcome is inefficient due to excessive entry, with deadweight loss $\frac{1}{4}\sqrt{tf}$. Minimum and maximum consumer utilities are $U(\frac{1}{2n^*_{pr}}) = v - c - \frac{3}{2}\sqrt{tf}$ and $U(0) = U(\frac{1}{n^*_{pr}}) = v - c - \sqrt{tf}$, respectively.*

Proof. Taking the first-order condition of (5) and setting $p_i = p$ (symmetric firms), it holds that $p^* = c + \frac{t}{n}$. Each firm's profit is then given by $\pi = \frac{t}{n^2} - f$. Firms enter as long as there are profits, giving $n^*_{pr} = \sqrt{t/f}$. Consumer surplus satisfies $CS = 2n^*_{pr}\int_0^{\frac{1}{2n^*_{pr}}}[v - c - \frac{t}{n} - \alpha t]d\alpha = v - c - \frac{5t}{4n^*_{pr}} = v - c - \frac{5}{4}\sqrt{tf}$. A social planner, in contrast, would choose n to maximize $2n\int_0^{\frac{1}{2n}}(v - c - \alpha t)d\alpha - nf = v - c - \frac{t}{4n} - nf$, resulting in $n^{SP} = \frac{1}{2}\sqrt{t/f} = \frac{1}{2}n^*_{pr}$; that is, half as many firms would enter relative to the market equilibrium. Social welfare under a planner is then $v - c - \frac{t}{4n^{SP}} - n^{SP}f = v - c - \sqrt{tf}$. Deadweight loss is $DWL = \frac{1}{4}\sqrt{tf}$.
■

Equilibrium without Privacy: If consumer types are common knowledge and arbitrage is infeasible, then neighboring firms compete for each consumer, and prices are driven downward. For $\alpha \in [0, \frac{1}{n}]$, it then holds:

	$p_i(\alpha)$	$p_j(\alpha)$
$\alpha \leq \frac{1}{2n}$	$c + t(\frac{1}{n} - 2\alpha)$	c
$\alpha \geq \frac{1}{2n}$	c	$c + t(2\alpha - \frac{1}{n})$

The resultant prices are thus once more the cost of production plus the difference in transportation costs. Let n^*_{np} denote the equilibrium number of firms under the outcome without privacy.

Proposition 8 *In equilibrium without privacy, $n^*_{np} = \sqrt{\frac{t}{2f}} < n^*_{pr}$ firms enter and realize zero profits. Consumer surplus is $v - c - \frac{3}{2}\sqrt{\frac{tf}{2}}$, and the outcome is inefficient, with deadweight loss $\left(\frac{3}{2\sqrt{2}} - 1\right)\sqrt{tf}$. Minimum and maximum consumer utilities are given by $U(0) = U(\frac{1}{n^*_{np}}) = v - c - \sqrt{2tf}$ and $U(\frac{1}{2n^*_{np}}) = v - c - \sqrt{\frac{tf}{2}}$, respectively.*

Proof. In equilibrium, each firm's profit is $\pi = 2\int_0^{\frac{1}{2n}} t(\frac{1}{n} - 2\alpha)d\alpha - f = \frac{t}{2n^2} - f$, resulting in $n^*_{np} = \frac{1}{\sqrt{2}}\sqrt{t/f}$ firms entering the market. Consumer surplus satisfies $CS = 2n^*_{np}\int_0^{\frac{1}{2n^*_{np}}}[v - c - \frac{t}{n} + \alpha t]d\alpha = v - c - \frac{3t}{4n^*_{np}} = v - c - \frac{3}{2\sqrt{2}}\sqrt{tf}$. A social planner, in contrast, sets $n^{SP} = \frac{1}{2}\sqrt{t/f} < n^*_{np}$, leading to a deadweight loss in the market equilibrium given by $DWL = \left(\frac{3}{2\sqrt{2}} - 1\right)\sqrt{tf}$. ■

Notice that in comparison to the outcome with complete privacy where firms charge uniform prices, consumers are overall better off with individualized pricing—despite the fact that fewer firms enter the market in equilibrium. However, some individual consumers are worse off without privacy—particularly those consumers who are located nearest to firms. While firms realize zero profits in both cases, increased competition among firms due to the targeting of individual consumers leads to lower gross profits, and thus induces fewer firms to enter. Consequently, the outcome under the no-privacy regime is closer to the efficient outcome and results in a smaller deadweight loss in equilibrium.

Model 3: Vertical Differentiation Model

Consider now a vertical-differentiation model in which two firms, L and H, produce products that are differentiated by their qualities, q_L and q_H, respectively, such that $0 < q_L < q_H$. Both firms' unit costs are constant at c as before. Consumers are differentiated by their willingness to pay. In particular, consumer have types $\theta \in [\underline{\theta}, \overline{\theta}]$, $0 < \underline{\theta} < \overline{\theta}$, and utilities $U(q_j, p_j; \theta) = \theta q_j - p_j$ for $j \in \{L, H\}$. To focus on interior solutions, assume that $q_L > \frac{2q_H(\overline{\theta} - 2\underline{\theta}) + 6c}{3\overline{\theta}}$ and $\overline{\theta} > 2\underline{\theta}$.

Equilibrium with Privacy: Given that products differ only in consumers' willingness to pay, it is efficient for all consumers to purchase product H. However, since firms have no information about consumers' types under the privacy regime, they set uniform prices. As a result, some consumers will buy product L. The marginal consumer type θ^* is indifferent between purchasing products L and H if and only if $\theta^* q_H - p_H = \theta^* q_L - p_L$. That is, at

$$\theta^* = \frac{p_H - p_L}{q_H - q_L}, \tag{6}$$

consumers with willingness to pay below (above) θ^* purchase product L (H). Taking the marginal consumer into account, firms L and H maximize profits with their objectives specified by $\max_{p_H} \pi_H = (\overline{\theta} - \theta^*)(p_H - c)$ and $\max_{p_L} \pi_L = (\theta^* - \underline{\theta})(p_L - c)$, respectively. In the following, let $\Delta_q = q_H - q_L$, and let CS_L and CS_H denote the surplus of consumers who buy product L and H, respectively.

Proposition 9 *In equilibrium with privacy, prices are given by* $p^*_H = \frac{1}{3}(2\overline{\theta} - \underline{\theta})\Delta_q + c$ *and* $p^*_L = \frac{1}{3}(\overline{\theta} - 2\underline{\theta})\Delta_q + c$, *and the marginal type is* $\theta^* = \frac{\overline{\theta} + \underline{\theta}}{3}$. *Profits satisfy* $\pi_H = \frac{1}{9}(2\overline{\theta} - \underline{\theta})^2\Delta_q$, *and* $\pi_L = \frac{1}{9}(\overline{\theta} - 2\underline{\theta})^2\Delta_q$. *The outcome is inefficient, with deadweight loss* $\frac{1}{18}(\overline{\theta} - 2\underline{\theta})(\overline{\theta} + 4\underline{\theta})\Delta_q$. *The minimum and maximum consumer utilities are* $U(q_L, p_L; \underline{\theta}) = \underline{\theta} q_L - \frac{1}{3}(\overline{\theta} - 2\underline{\theta})\Delta_q - c$ *and* $U(q_H, p_H; \overline{\theta}) = \overline{\theta} q_H - \frac{1}{3}(2\overline{\theta} - \underline{\theta})\Delta_q - c$, *respectively.*

Proof. Substituting (6) into $\max_{p_H} \pi_H = (\overline{\theta} - \theta^*)(p_H - c)$ and $\max_{p_L} \pi_L = (\theta^* - \underline{\theta})(p_L - c)$ and taking the first-order conditions yields $p_H = (p_L + c + \overline{\theta}\Delta_q)/2$ and $p_L = (p_H + c - \underline{\theta}\Delta_q)/2$. Solving for the equilibrium prices yields $p^*_H = \frac{1}{3}(2\overline{\theta} - \underline{\theta})\Delta_q + c$ and $p^*_L = \frac{1}{3}(\overline{\theta} - 2\underline{\theta})\Delta_q + c$, resulting in $\theta^* = \frac{\overline{\theta} + \underline{\theta}}{3}$, $\pi_H = \frac{1}{9}(2\overline{\theta} - \underline{\theta})^2\Delta_q$, and $\pi_L = \frac{1}{9}(\overline{\theta} - 2\underline{\theta})^2\Delta_q$, whereas $CS_L = \int_{\underline{\theta}}^{\theta^*}(\theta q_L - p_L)d\theta = \frac{\overline{\theta} - 2\underline{\theta}}{9}\left(\overline{\theta}(\frac{3q_L}{2} - q_H) + 2\underline{\theta} q_H - 3c\right)$ and $CS_H = \int_{\theta^*}^{\overline{\theta}}(\theta q_H - p_H)d\theta = \frac{2\overline{\theta} - \underline{\theta}}{9}\left(\frac{3}{2}\underline{\theta} q_H + (2\overline{\theta} - \underline{\theta})q_L - 3c\right)$. Since some consumers buy product L, the outcome is inefficient, with deadweight loss $DWL = \int_{\underline{\theta}}^{\theta^*}\Delta_q \theta d\theta = \frac{1}{18}(\overline{\theta} - 2\underline{\theta})(\overline{\theta} + 4\underline{\theta})\Delta_q$. ■

Equilibrium without Privacy: If consumer types are common knowledge and arbitrage is infeasible, then firms compete for each consumer individually. As a result, the following holds:

$$\begin{aligned} p_L(\theta) &= c, \quad \theta \in [\underline{\theta}, \overline{\theta}] \\ p_H(\theta) &= c + \theta\Delta_q, \quad \theta \in [\underline{\theta}, \overline{\theta}] \end{aligned} \tag{7}$$

Consequently, all consumers buy from firm H, which leads to an efficient outcome.

Proposition 10 *In equilibrium without privacy, profits satisfy* $\pi_L = 0$ *and* $\pi_H = \frac{1}{2}(\overline{\theta}^2 - \underline{\theta}^2)\Delta_q$*, consumer surplus is given by* $(\overline{\theta} - \underline{\theta})\left(\frac{1}{2}q_L(\overline{\theta} + \underline{\theta}) - c\right)$*, and the outcome is efficient. The minimum and maximum consumer utilities are given by* $U(q_H, p_H(\underline{\theta}); \underline{\theta}) = \underline{\theta}q_L - c$ *and* $U(q_H, p_H(\overline{\theta}); \overline{\theta}) = \overline{\theta}q_L - c$*, respectively.*

Proof. Given firms' pricing strategies as specified in (7), we have $\pi_L = 0$, $\pi_H = \int_{\underline{\theta}}^{\overline{\theta}} \theta\Delta_q d\theta = \frac{1}{2}(\overline{\theta}^2 - \underline{\theta}^2)\Delta_q$, and $CS = \int_{\underline{\theta}}^{\overline{\theta}} \theta q_H - p_H(\theta)d\theta = \int_{\underline{\theta}}^{\overline{\theta}} \theta q_L - cd\theta = (\overline{\theta} - \underline{\theta})\left(\frac{1}{2}q_L(\overline{\theta} + \underline{\theta}) - c\right)$. Since all consumers buy product H, the outcome is efficient. ■

Notice that in comparison to the outcome with complete privacy where firms charge uniform prices, some consumers are better off with individualized pricing, while others — particularly those with higher valuations — prefer privacy. The following corollary formalizes this observation.

Corollary 1 *Under vertical differentiation, consumers with types* $\theta > \frac{2\overline{\theta}-\underline{\theta}}{3}$ *prefer the privacy regime, and those with types* $\theta < \frac{2\overline{\theta}-\underline{\theta}}{3}$ *prefer no privacy. Overall profits decrease whereas consumer surplus increases under no privacy.*

Proof. In the equilibrium under no privacy, $U(q_H, p_H(\theta); \theta) = \theta q_L - c$. With privacy, consumers with $\theta > \theta^* = \frac{\overline{\theta}+\underline{\theta}}{3}$ have utility $U(q_H, p_H; \theta) = \theta q_H - \frac{1}{3}(2\overline{\theta} - \underline{\theta})\Delta_q - c$. Notice that a consumer of type $\theta > \theta^*$ prefers privacy iff $\theta q_H - \frac{1}{3}(2\overline{\theta} - \underline{\theta})\Delta_q - c > \theta q_L - c$; that is, if $\theta > \frac{2\overline{\theta}-\underline{\theta}}{3}$. Maintaining $\overline{\theta} > 2\underline{\theta}$, it follows that $\frac{2\overline{\theta}-\underline{\theta}}{3} > \frac{\overline{\theta}+2\underline{\theta}-\underline{\theta}}{3} = \theta^*$. Finally, under the privacy regime, for types $\theta < \theta^*$, consumer utility is given by $U(q_L, p_L; \theta) = \theta q_L - c - \frac{1}{3}(\overline{\theta} - 2\underline{\theta})\Delta_q$, which is evidently lower than the utility of a consumer in this type range without privacy, given by $U(q_H, p_H(\theta); \theta) = \theta q_L - c$.

Profits with privacy are given by $\pi_L + \pi_H = \frac{1}{9}(\overline{\theta} - 2\underline{\theta})^2\Delta_q + \frac{1}{9}(2\overline{\theta} - \underline{\theta})^2\Delta_q = \frac{5}{9}\Delta_q(\overline{\theta}^2 - \frac{8}{5}\overline{\theta}\underline{\theta} + \underline{\theta}^2) > \frac{5}{9}\Delta_q(\overline{\theta} - \underline{\theta})^2$, which is evidently greater than total profits without privacy, $\frac{1}{2}\Delta_q(\overline{\theta} - \underline{\theta})^2$. Since overall profits decrease but the outcome is efficient without privacy, it follows that overall consumer surplus rises. ■

In the absence of privacy, more consumers purchase product H. While this clearly leads to a rise in allocative efficiency, it also results in an overall increase in consumer surplus—despite lower payoffs for consumers with higher willingness to pay. Moreover, while firm H is able to monopolize the market with individualized pricing, the potential entry of product L forces it to keep its prices below consumers' reservation values and leads to a reduction in overall industry profits.

Model 4: Multi-Unit Symmetric Demand Model

Consider two symmetric firms with unit production costs of c. There is a population of consumers with demand parameters $\gamma \in [\underline{\gamma}, \overline{\gamma}]$. A type γ consumer has demands for each good specified by

$$x_i = \gamma - p_i + bp_j, \quad i = 1, 2.$$

The measure of consumers with $\tilde{\gamma} \leq \gamma$ is specified by the cumulative-density function $F(\gamma)$, where F is continuously differentiable with $\int_{\underline{\gamma}}^{\overline{\gamma}} \gamma\, dF(\gamma) \equiv \mu$. To ensure existence of an interior equilibrium, assume $b \in (0, 1)$ and $\underline{\gamma} > \frac{(1-b)(\mu+c)}{2-b}$.

Equilibrium with Privacy: Suppose types $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ are private information. When firms have no information about consumers' types, they set uniform prices. In particular, firm i, $i \in \{1, 2\}$, solves

$$\max_{p_i} \; E[\pi_i] = (\mu - p_i + bp_j)(p_i - c). \tag{8}$$

Proposition 11 *In equilibrium with privacy, prices satisfy* $p^* = \frac{\mu+c}{2-b}$*. Expected profit for each firm is* $\left(\frac{\mu-(1-b)c}{2-b}\right)^2$ *and consumer surplus is* $E[\gamma^2] - 2\mu\left(\frac{(1-b)(\mu+c)}{2-b}\right) + \left(\frac{(1-b)(\mu+c)}{2-b}\right)^2$*. The outcome is inefficient, with minimum and maximum consumer utilities given by* $U(\underline{\gamma}) = \left(\underline{\gamma} - \frac{(1-b)(\mu+c)}{2-b}\right)^2$ *and* $U(\overline{\gamma}) = \left(\overline{\gamma} - \frac{(1-b)(\mu+c)}{2-b}\right)^2$*, respectively.*

Proof. Taking the first-order condition of (8), we obtain $p_i = \frac{\mu + bp_j + c}{2}$ for $i \in \{1, 2\}$. Solving for the Nash equilibrium gives $p^* = \frac{\mu+c}{2-b}$ and $E[\pi] = \int_{\underline{\gamma}}^{\overline{\gamma}} (\gamma - (1-b)\frac{\mu+c}{2-b})(\frac{\mu+c}{2-b} - c) dF(\gamma) = \left(\frac{\mu-(1-b)c}{2-b}\right)^2$. Consumer surplus for type γ in each market is given by the area under the equilibrium demand curve and above the equilibrium price. Consumer surplus for type γ summed across both markets is then $\left(\gamma - \frac{(1-b)(\mu+c)}{2-b}\right)^2$. Aggregate consumer surplus is then given by $\int_{\underline{\gamma}}^{\overline{\gamma}} \left(\gamma - \frac{(1-b)(\mu+c)}{2-b}\right)^2 dF(\gamma) = E[\gamma^2] - 2\mu \left(\frac{(1-b)(\mu+c)}{2-b}\right) + \left(\frac{(1-b)(\mu+c)}{2-b}\right)^2$. ■

Equilibrium without Privacy: If firms observe γ, then they price discriminate by tailoring prices to consumers based on their type. In particular, firm i, $i \in \{1, 2\}$, sets individual prices to maximize the proceeds from each consumer γ by solving

$$\max_{p_i} \ \pi_i(\gamma) = (\gamma - p_i(\gamma) + bp_j(\gamma))(p_i(\gamma) - c). \tag{9}$$

Proposition 12 *In equilibrium without privacy, prices satisfy $\hat{p}(\gamma) = \frac{\gamma+c}{2-b}$. Each firm's profit is given by $\frac{E[\gamma^2]-2(1-b)c\mu+(1-b)^2c^2}{(2-b)^2}$ and consumer surplus is $\frac{1}{(2-b)^2}\left(E[\gamma^2] - 2(1-b)c\mu + (1-b)^2c^2\right)$. The outcome is inefficient, with minimum and maximum consumer utilities given by $U(\underline{\gamma}) = \left(\frac{\underline{\gamma}-(1-b)c}{2-b}\right)^2$ and $U(\overline{\gamma}) = \left(\frac{\overline{\gamma}-(1-b)c}{2-b}\right)^2$, respectively.*

Proof. Taking the first-order condition of (9), it follows that $p_i(\gamma) = \frac{\gamma+bp_j+c}{2}$ for $i \in \{1, 2\}$. Solving for the Nash equilibrium gives $\hat{p} = \frac{\gamma+c}{2-b}$ and $E[\pi] = \int_{\underline{\gamma}}^{\overline{\gamma}} (\gamma - (1-b)\frac{\gamma+c}{2-b})(\frac{\gamma+c}{2-b} - c) dF(\gamma) = \frac{E[\gamma^2]-2(1-b)c\mu+(1-b)^2c^2}{(2-b)^2}$. Consumer surplus for type γ summed across both markets is given by $\left(\gamma - \frac{(1-b)(\gamma+c)}{2-b}\right)^2$. Aggregate consumer surplus is then $\int_{\underline{\gamma}}^{\overline{\gamma}} \left(\gamma - \frac{(1-b)(\gamma+c)}{2-b}\right)^2 dF(\gamma) = \frac{1}{(2-b)^2}\left(E[\gamma^2] - 2(1-b)c\mu + (1-b)^2c^2\right)$. ■

Notice that in comparison to the outcome with privacy where firms charge uniform prices, consumers with types $\gamma < \mu$ are better off with individualized pricing, while those with types $\gamma > \mu$ benefit from privacy. The following corollary compares consumer surplus, profits, and welfare with and without privacy.

Corollary 2 *In the multi-unit symmetric demand model, profits are higher and consumer surplus is lower without privacy. For $b \in [0, 2-\sqrt{3})$, overall welfare is higher with privacy, and for $b \in (2-\sqrt{3}, 1]$, welfare is higher without privacy.*

Proof. Notice that in the absence of privacy, each firm's profit rises by $\frac{E[\gamma^2]-\mu^2}{(2-b)^2}$, an amount that is strictly positive by Jensen's inequality. Similarly, the difference between consumer surplus with privacy and without privacy is given by $\frac{1}{(2-b)^2}(3-b)(1-b)(E[\gamma^2] - \mu^2)$, also positive by Jensen's inequality. Subtracting total surplus under no privacy from total surplus with privacy gives $\frac{1}{(2-b)^2}(1 - (4-b)b)(E[\gamma^2] - \mu^2)$. This difference is positive for $b \in [0, 2-\sqrt{3})$ and negative for $b \in (2-\sqrt{3}, 1]$. ■

Firms' pricing strategies in this model are strategic complements. Lower values of b reduce this complementarity feature. Firms, in effect, then behave as near-monopolists. Under the privacy regime, the welfare loss of near-monopoly pricing is partially mitigated by firms' incomplete information about consumers' demands, as firms are forced to set uniform prices. Under the no privacy regime, firms, in effect, bring the welfare distortion of a near-monopoly to individual consumers, which results in decreased consumer surplus and overall welfare but higher profits.

Higher values of b increase the complementarity of firms' pricing strategies. This feature is enhanced under the no-privacy regime, as firms take advantage of this complementarity when pricing to individual consumers, leading to an overall increase in consumers' demands. As a result, despite price discrimination and lower consumer surplus, overall welfare is higher under no privacy.

DATA Catalyst