US Centre Summer Research Grant

Recipient name: Xinchen Ma

Project title: Consumers' Demand for Digital Privacy

Summary of project:

This paper investigates how consumers and investors react to the standardized disclosure of data privacy practices. Since December 2020, Apple has required all apps to disclose their data collection practices by filling out privacy "nutrition" labels that are standardized and easy to read. We web-scrape these privacy labels and first document several stylized facts regarding the supply of privacy. Second, augmenting privacy labels with weekly app downloads and revenues, we examine how this disclosure affects consumer behaviour. Exploiting the staggered release of privacy labels and using the nonexposed Android version of each app to construct the control group, we find that after privacy label release, an average iOS app experiences a 14% (15%) drop in weekly downloads (revenue) when compared to its Android counterpart. The effect is stronger for more privacy-invasive and substitutable apps. Moreover, we observe negative stock market reactions, especially among firms that harvest more data, corroborating the adverse impact on product markets. We also find that consumers' annual WTP to avoid data collection by an average app is $7, which is mostly attached to data used for tracking and advertising.

US CENTRE SUMMER RESEARCH GRANT REPORT

1. Introduction and Objective

Recent decades have seen a digital revolution, shifting economic activities from offline to online markets. With this sweeping change, personal data has become an essential element of business, fueling a $227 billion-a-year data industry. The rise of social media and Big Tech showcases the potential of data monetization at scale. However, the risk of privacy intrusion and data breach looms large at the same time. In response to growing public concerns about data privacy issues and cybersecurity risks, bold regulatory moves have emerged in various jurisdictions. Meanwhile, a nascent set of academic works is starting to associate firm valuation and corporate policies with cybersecurity risks and data breaches. These discrete events originate from firm's continuous data harvesting, and their impact on firms depends crucially on consumers' attitudes towards privacy. Despite major regulatory efforts and heated public discussions, there is limited large-scale evidence on the market for data privacy. How much privacy do firms supply? Can we consistently measure the scope and the purpose of data collection? How much do consumers demand privacy, and does it translate into the valuation of firms that thrive on monetizing personal data?

This project aims to shed some light on data privacy's supply and demand side, especially focusing on customer reaction toward policy regulation changes. First, we aim to provide a comprehensive descriptive analysis of the supply side: how much data the firm collects, for what purposes, and what app characteristics are associated with more data collection. Second, we investigate how customers react following the release of the iOS privacy labels relative to apps' Android versions. Third, we check how broad the pattern is applicable across 95 economies and aim to find contributors to the country-level heterogeneity. Finally, we quantify customer's willingness to pay for online privacy using structural estimation.

2. Empirical Methodology

2.1 Reduced-form regression: Difference-in-difference

Using a difference-in-differences approach, we first estimate the causal impact of privacy label release on the demand for digital services. Our main regression sample is the top 10k US

sample. We begin by investigating the share of the iOS version of apps relative to the Android version over time. To take into account any potential pre-trends, we allow for platform-specific linear trends in the regression specification. We formally estimate the following regression specification:

$$Y_{i,p,t} = \beta_1 iOS_{i,p} \times Post_{i,t} + \beta_2 Post_{i,t} + \alpha_i + \varphi_{age,p} + \theta_t + \lambda iOS_p \times t + \varepsilon_{i,p,t}$$

in which the subscript i, p, and t denote app, platform, and week respectively. The app-specific event indicator, $Post_{i,t}$ equals one for all the weeks after the respective app's release date and zero otherwise. The treatment indicator, $iOS_{i,p}$ equals one (zero) if the observation corresponds to the iOS (Android) version of an app. The outcome variable, $Y_{i,p,t}$, is the logarithm of the weekly downloads or revenue of app i on platform p in week t. We add app fixed effects α_i so that the variation in the outcome variable comes from the difference between the iOS and Android versions of the same app. We include year-week fixed effects, θ_t, to control for seasonality and other common shocks to the consumption of all mobile apps. With $iOS_p \times t$ fixed effects and $platform \times t$ as an additional regressor, our specification also allows for different time trends for iOS and Android apps over their life cycle and calendar years. We double-cluster standard errors by app developer and year-week. Our coefficient of interest, β_1, captures the effect of privacy label release on users' consumption of mobile apps. The key variable of interest is $iOS_{i,p} \times Post_{i,t}$, and its coefficient captures the differential effect of privacy label release on app downloads and revenue from the iOS platform versus those from the Android platform. Based on this benchmark regression, we also conduct dynamic difference-in-difference regressions, and we explore heterogeneities through triple interactions.

2.2 Structural estimation: BLP model

To quantify customers' willingness to pay, we estimate a nested-logit BLP type discrete choice model in which we assume the same degree of substitutability among apps in the same nest, and consumers choose the apps to download, taking into account app price, data collection intensity, and other app features. Importantly, we assume that in the absence of privacy labels, consumer beliefs about an app's data collection are constant. We further assume the supply of data privacy by app developers, measured by the amount of data collected, to remain unchanged. This is consistent with the lack of time-series variations in privacy labels for any

given app during a 12-month period. We model a continuum of homogeneous consumers who choose from J apps to download among the top 10k iOS apps and an outside option. This outside option h can be interpreted as the option to download no apps or to download apps outside the top 10k iOS app sample. Consumers' utility from app j is given by

$$U_{jt} = \delta_{jt} + \varepsilon_{jt}$$

where δ_{jt} denotes the mean utility derived from app j at time t, and ε_{jt} denotes the error term. Specifically, the error term is $i.i.d.$ with the Type I Extreme Value distribution. To capture the features of the app markets, we model the mean utility term as

$$\delta_{jt} = X_{jt}\beta + \alpha p_{jt} + \gamma d_j + \xi_{jt}$$

where X_{jt} captures app characteristics (e.g., rating, recent updates, app size) that influence the user experience, p_{jt} denotes prices, d_j represents the amount and scope of data collected from consumers (e.g., whether the app tracks users, the number of data types or items collected), and ξ_{jt} represents the unobserved characteristics to econometricians. We expect negative coefficients for both p_{jt} and d_j, with the coefficient for the latter representing disutility from data being collected. After the release of privacy labels of app j, d_j becomes observable to consumers and directly impacts their decision to download the app. Before the release of privacy labels, while X_{jt} and p_{jt} are observed, data collection intensity d_j is not. Instead, consumers form beliefs of d_j based on the nest an app belongs to, which we assume to be the same for all apps within the nest. The pre- and post-label scenarios can be combined into and represented by the following estimation equation, where the pre- and post-label choices can be estimated simultaneously:

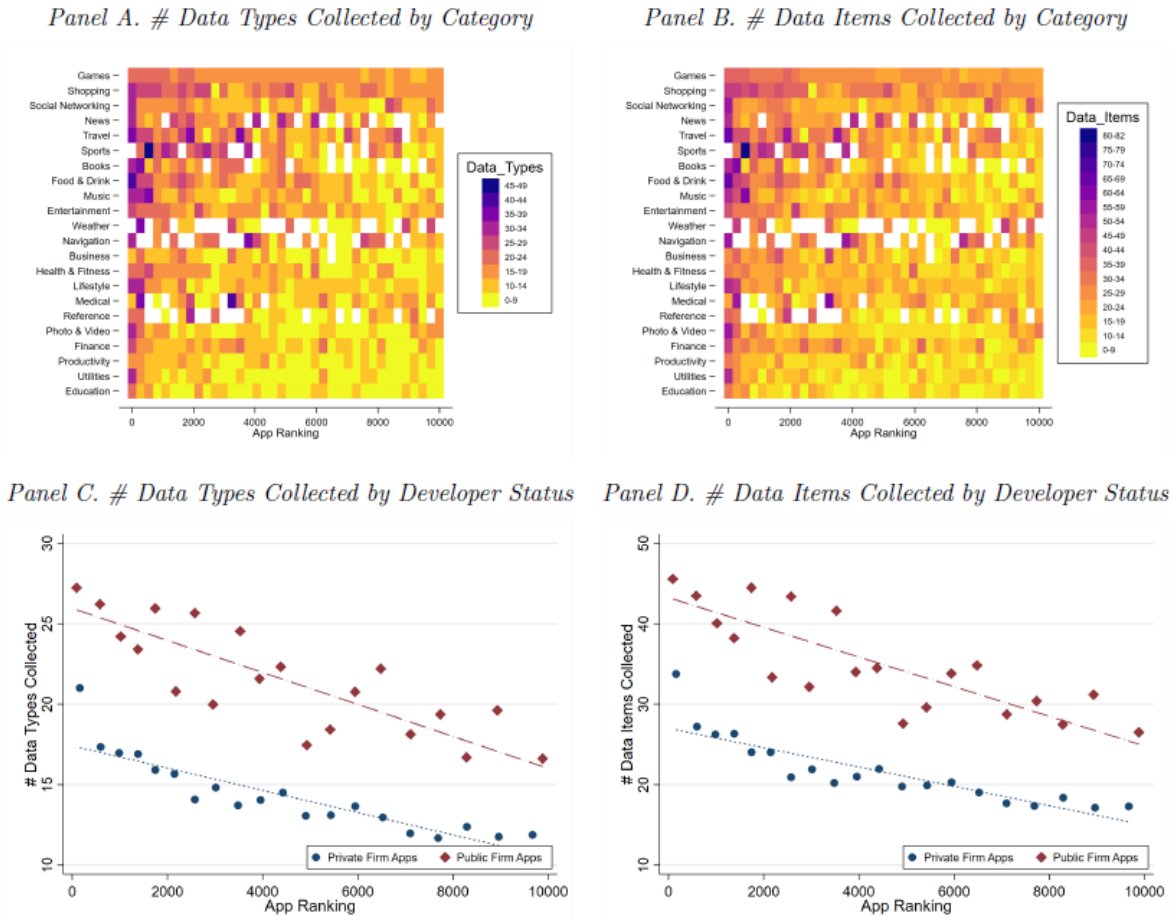
$$\ln(s_{jt}) - \ln(s_{0t}) = X_{jt}\beta + \alpha p_{jt} + \lambda Post_{jt} + \phi Post_{jt} \times d_j + \delta_j + \delta_t + \rho \ln(s_{j|h(j),t})$$

where s_{jt} and s_{0t} are market share of app j and the outside option, while $s_{j|h(j),t}$ denotes the app's market share within the nest. Coefficients λ measures the overall impact of nest-specific beliefs about data collection and ϕ captures the marginal disutility from data collection. Consumer surplus and willingness to pay can be calculated following standard BLP estimation procedure.

3. Results

3.1 Supply side: Privacy Labels

Figure 1. Data Collection Intensity

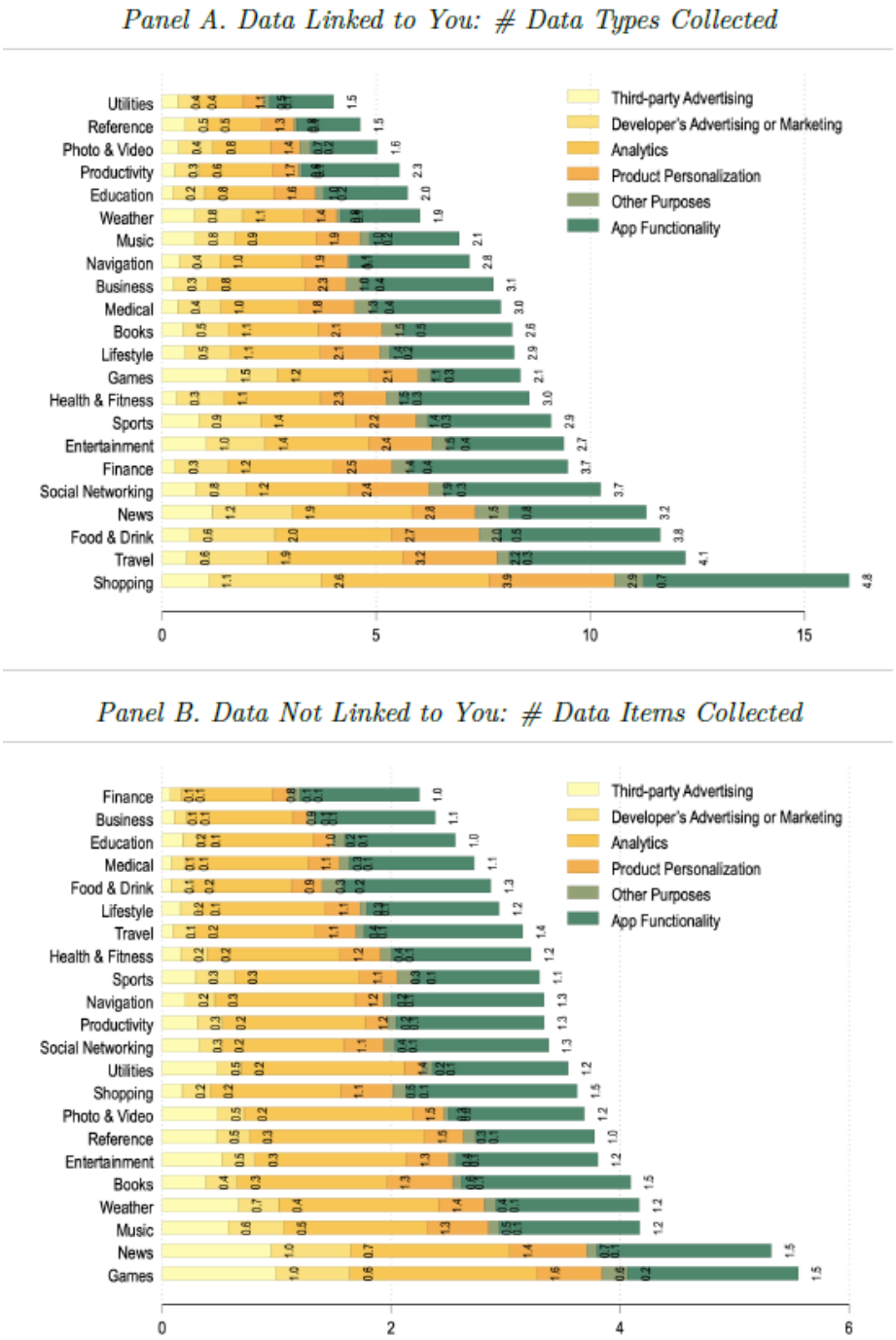


Intensity We find that an average app collects 24 data items across 16 data types, with substantial variations across apps. 80% of data items collected are used for purposes unrelated to the functionality of an app. Data are most frequently collected for product personalization and developer's advertising or marketing. More importantly, 60% of apps collect data to track users (or their devices) and share user data across different apps, advertising networks, and companies. Worse still, sensitive information collected within this category could be sold to data brokers.

Characteristics We found that apps that collect more data have a larger market share, a younger age, a higher rating, and more in-app purchases. They are also more likely to be developed by publicly listed firms (Figure 1). These results hold after controlling for app category fixed effects. Out of over 20 categories, we find that gaming apps gather the most

data for third-party advertising, while shopping apps are the top data collector for multiple data uses including developer's advertising or marketing, analytics, and product personalization. Apart from these two categories, news, food & drinks, and social networking apps are also heavy data collectors for purposes other than app functionality (Figure 2).

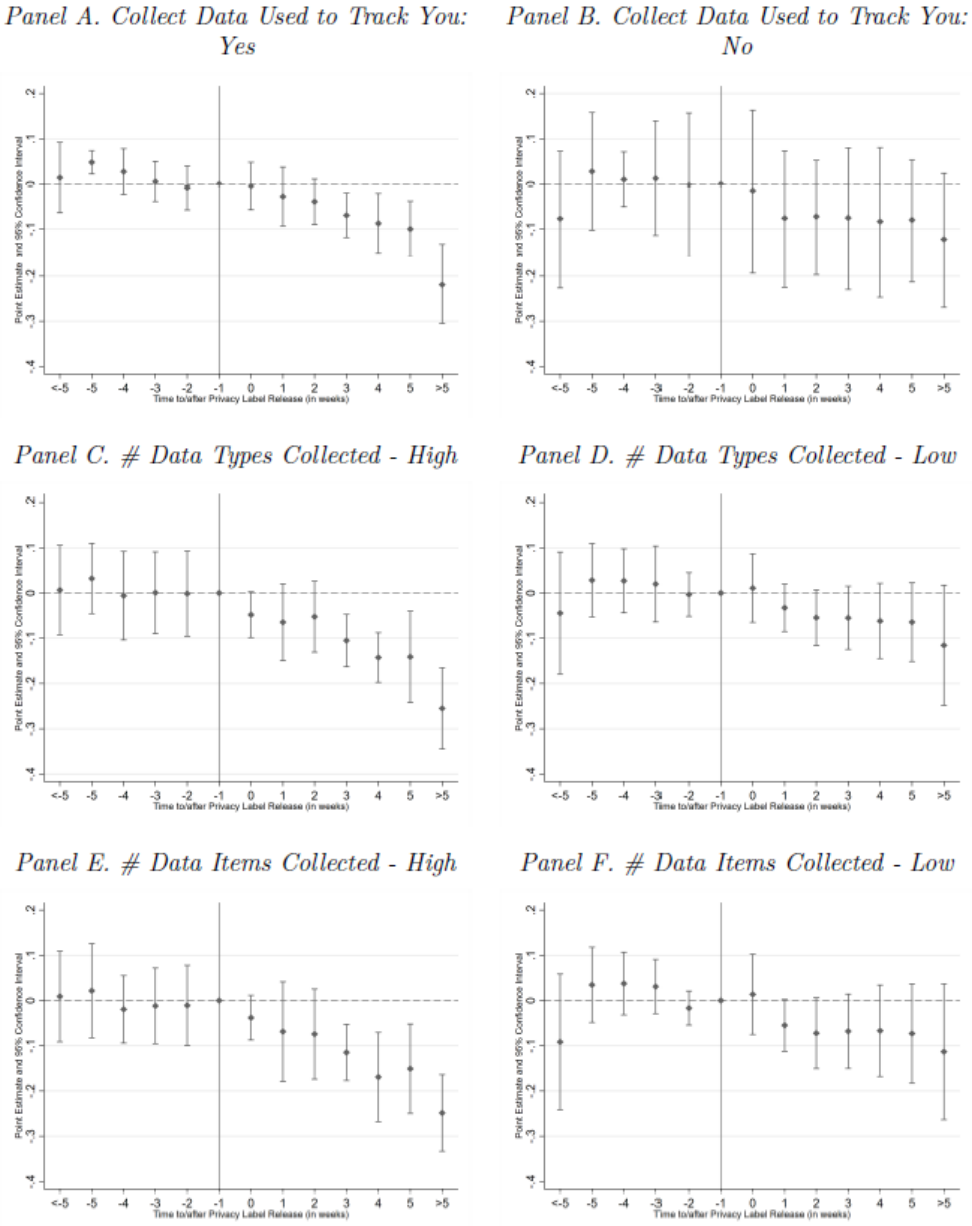
Figure 2. Data Collection Intensity and Data Use



3.2 Demand side: Difference in Difference

Exploiting the privacy label release, we estimate a difference-in-differences model, where iOS apps form the treatment group and corresponding Android apps the control. Our central finding is that, following the release of privacy labels, relative to its Android counterpart, the iOS version of a given app on average experiences a 14% decline in weekly download and a 15% decline in revenue from user subscriptions and in-app purchases. This result is robust to different combinations of fixed effects, clusters, and alternative sampling criteria. We show that the decline in downloads and revenue is greater when the data collection intensity is higher (Figure 3). For example, the treatment effect for apps that collect data to track consumers is 2.2 times that for apps that do not collect such data.

Figure 3. Impact of Privacy Label Release on Downloads



3.3 International Evidence

we expand the sample to 95 countries and repeat our baseline difference-in-differences analysis to obtain country- specific DiD estimators (Figure 4). We can associate these coefficients with country-level factors including legal environment, general trust, and survey-based data privacy concerns (Figure 5). We first document the role of legal institutions. Countries with stronger legal protection of privacy and better law enforcement react less negatively to privacy labels, presumably because consumers consider the current legal protection of their personal data adequate. Second, as proxies for general trust, confidence in the press and confidence in major companies negatively correlates with consumers' demand for privacy. In the end, consumer attitudes regarding data use and privacy also matter. Consumers in countries with more severe privacy concerns react more negatively to privacy label release.

Figure 4. Data Collection Intensity and Data Use

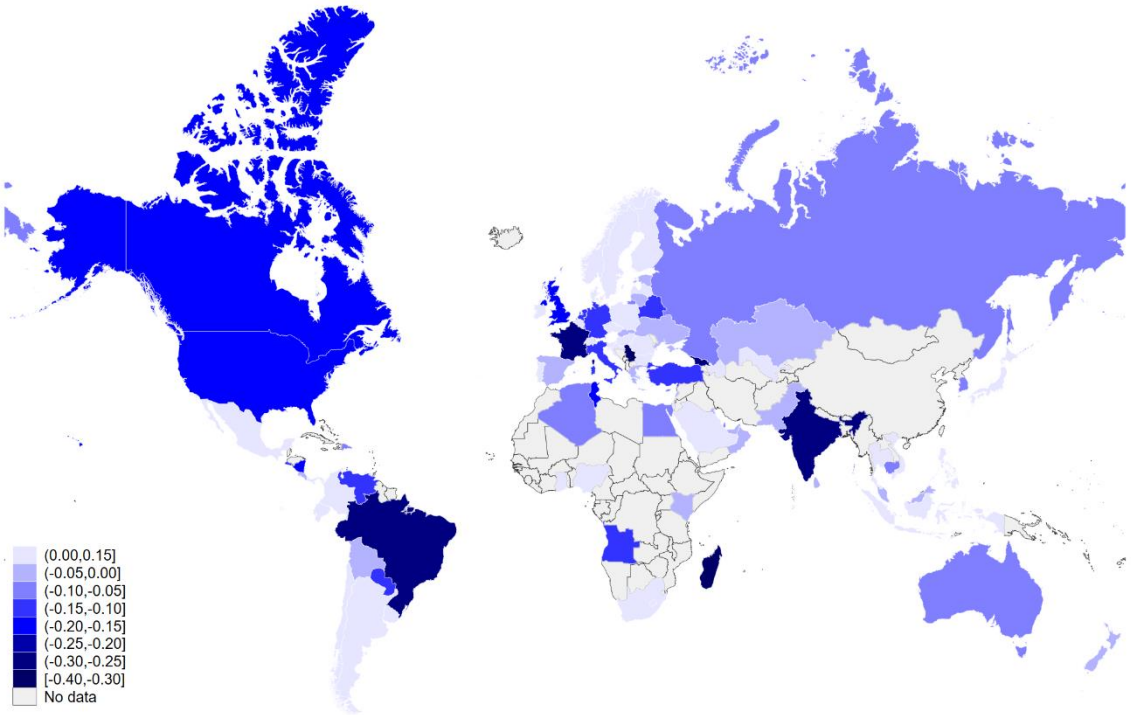
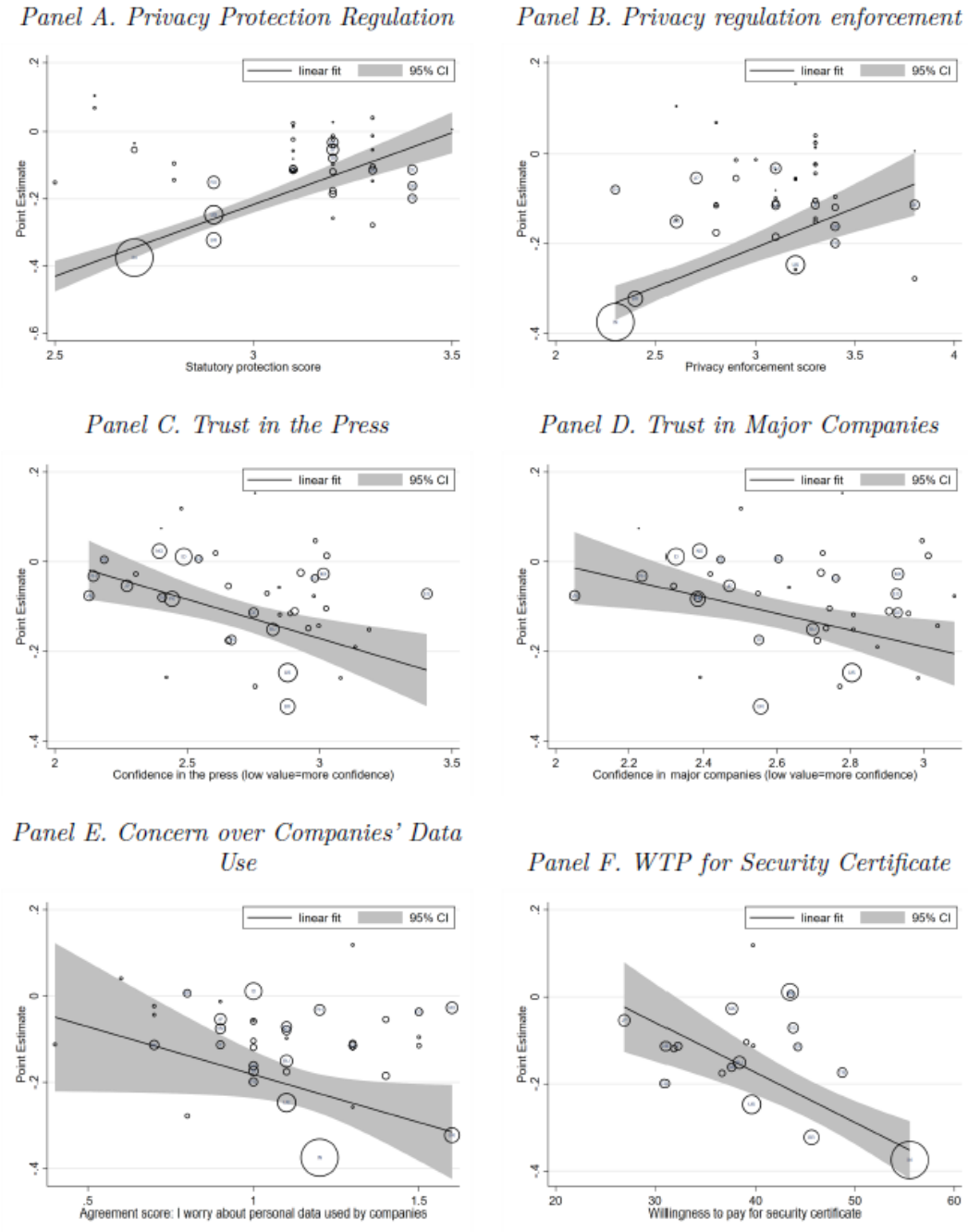


Figure 5. The Role of Regulation, Public Trust, and Privacy Concerns



4.4 Customer's willingness to pay

Motivated by our reduced-form evidence, we conjecture that consumers underestimate the number of collected data items on average, and therefore, overestimate the utility of using digital services. Our counterfactual analyses therefore provide answers to the question of "how the perceived consumer surplus would change had customers known the data collection intensity before privacy label releases, for a given prior belief and holding their app choices

unchanged." We consider a set of counterfactual prior beliefs by setting it to the minimum (i.e., zero), 25th, 50th, and 75th percentiles of the actual data collection intensity distribution among the top 10k gaming apps. We found that indeed, consumers underestimate data collection intensity, even when setting the prior belief to the median collection intensity, customers still overperceive welfare by $0-$24 per data use per annum (Figure 6). We also find that consumers' annual WTP to avoid data collection by an average app is $7, which is mostly attached to data used for tracking and advertising.

Figure 6. Counterfactual belief in data collection intensity in the Pre-label period

$(Actual welfare - Perceived Welfare)	#Tracking	#3P Ads	#1P Ads	#Analytics	#Personalization	#Others	#AppFn
No data collection	-18.77	-18.94	-35.38	-8.47	-22.16	-2.15	43.82
	(-0.91%)	(-0.92%)	(-1.72%)	(-0.41%)	(-1.08%)	(-0.10%)	(-2.12%)
25th percentile	-12.64	-13.08	-35.38	-4.30	-22.16	-2.15	33.85
	(-0.61%)	(-0.63%)	(-1.72%)	(-0.20%)	(-1.08%)	(-0.10%)	(-1.64%)
50th percentile	0.66	-7.14	-24.26	-0.05	-15.35	-2.15	6.38
	(-0.03%)	(-0.34%)	(-1.18%)	(0.00%)	(-0.74%)	(-0.10%)	(-0.3%)
75th percentile	7.86	4.97	-0.60	2.11	-1.38	-2.15	-2.01
	(-0.38%)	(-0.24%)	(-0.02%)	(-0.1%)	(-0.06%)	(-0.10%)	(-0.09%)

4. Conclusion

Following Apple's privacy label policy, iOS app developers are required to report the collection and use of customer data in a transparent and digestible "nutrition label" format. We scrape privacy labels for the most popular iOS apps in ten countries to provide a valuable measure of data collection intensity. Supplementing this dataset with weekly downloads and revenues from Sensor Tower, we investigate how consumers react to the standardized disclosure of data privacy practices - a key element of corporate digital responsibility. We show that consumers are averse to data collection by apps, especially when their data is collected for privacy-invasive uses. Our findings highlight the lack of consumer awareness about firms' data collection practices as one important explanation for the privacy paradox – the discrepancy between an individual's intentions to protect their privacy and how they actually behave in the online marketplace. We also document negative stock market reactions, in particular among firms in the retail and service sector that harvest more user data. Overall, our findings suggest that data play a central role in firm valuations in today's digital economy.

Reference:

Abrokwa, Desiree, Shruti Das, Omer Akgul, and Michelle L Mazurek, 2021, Comparing security and privacy attitudes between iOS and Android users in the US, in *SOUPS 2021: USENIX Symposium on Usable Privacy and Security*.

Acquisti, Alessandro, Leslie K. John, and George Loewenstein, 2013, What is privacy worth?, The Journal of Legal Studies 42, 249–274.

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman, 2016, The economics of privacy, Journal of Economic Literature 54, 442–492.

Akey, Pat, Stefan Lewellen, Inessa Liskovich, and Christoph Schiller, 2021, Hacking corporate reputations, Rotman School of Management Working Paper.

Al-Natour, Sameh, Hasan Cavusoglu, Izak Benbasat, and Usman Aleem, 2020, An empirical investigation of the antecedents and consequences of privacy uncertainty in the context of mobile apps, Information Systems Research 31, 1037–1063.

Amos, Ryan, Gunes Acar, Elena Lucherini, Mihir Kshirsagar, Arvind Narayanan, and Jonathan Mayer, 2021, Privacy policies over time: Curation and analysis of a milliondocument dataset, in Proceedings of the Web Conference 2021, 2165–2176.

Aridor, Guy, Yeon-Koo Che, and Tobias Salz, 2020, The economic consequences of data privacy regulation: Empirical evidence from GDPR, NBER Working Paper 26900, Columbia University, Massachusetts Institute of Technology.

Athey, Susan, Christian Catalini, and Catherine Tucker, 2017, The digital privacy paradox: Small money, small costs, small talk, NBER Working Paper 23488, Stanford University, Massachusetts Institute of Technology.

Bana, Sarah, Erik Brynjolfsson, Wang Jin, Sebastian Steffen, and Xiupeng Wang, 2021, Cybersecurity hiring in response to data breaches, Available at SSRN.

Bertrand, Marianne, and Emir Kamenica, 2018, Coming apart? cultural distances in the United States over time, Technical report, National Bureau of Economic Research.

Bessen, James E., Stephen M. Impink, Lydia Reichensperger, and Robert Seamans, 2020, GDPR and the importance of data to AI startups, Working paper, New York University, Boston University.

Chen, Long, Yadong Huang, Shumiao Ouyang, and Wei Xiong, 2021, The data privacy paradox and digital demand, NBER Working Paper 28854, Luohan Academy, Princeton University.

Chia, Pern Hui, Yusuke Yamamoto, and Asokan N., 2012, Is this app safe? A large scale study on application permissions and risk signals, in Proceedings of the 21st International Conference on World Wide Web, 311–320.

Chiou, Lesley, and Catherine Tucker, 2017, Search engines and data retention: Implications for privacy and antitrust, Technical report, National Bureau of Economic Research.

Comscore, 2019, Global state of mobile, Technical report.

eMarketer, 2020, US mobile time spent 2020, Technical report.

Florakis, Chris, Christodoulos Louca, Roni Michaely, and Michael Weber, 2020, Cybersecurity risk, Technical report, National Bureau of Economic Research.

Goldberg, Samuel, Garrett Johnson, and Scott Shriver, 2019, Regulating privacy online: The early impact of the GDPR on European web traffic & e-commerce outcomes, Available at SSRN 3421731.

Goldfarb, Avi, Shane M Greenstein, and Catherine E Tucker, 2015, Economic Analysis of the Digital Economy (University of Chicago Press).

Goldfarb, Avi, and Catherine Tucker, 2012a, Privacy and innovation, Innovation Policy and the Economy 12, 65–90.

Goldfarb, Avi, and Catherine Tucker, 2012b, Shifts in privacy concerns, American Economic Review 102, 349–353.

Goldfarb, Avi, and Catherine Tucker, 2019a, Digital economics, Journal of Economic Literature 57, 3–43.

Goldfarb, Avi, and Catherine Tucker, 2019b, Digital marketing, in Handbook of the Economics of Marketing, volume 1, 259–290 (Elsevier).

Goldfarb, Avi, and Catherine E Tucker, 2011, Privacy regulation and online advertising, Management Science 57, 57–71.

Huang, Henry He, and Chong Wang, 2021, Do banks price firms' data breaches?, The Accounting Review 96, 261–286.

Janssen, Rebecca, Reinhold Kesler, Michael Kummer, and Joel Waldfogel, 2021, GDPR and the lost generation of innovative apps, NBER Working Paper 146409, University of Zurich, University of Minnesota, University of East Anglia, Georgia Institute of Technology.

Jentzsch, Nicola, S¨oren Preibusch, and Andreas Harasser, 2012, Study on monetising privacy: An economic model for pricing personal information, Report for the European Network and Information Security Agency (ENISA), Heraklion: European Network and Information Security Agency.

Jia, Jian, Ginger Zhe Jin, and Liad Wagman, 2021, The short-run effects of the general data protection regulation on technology venture investment, Marketing Science forthcoming.

Johnson, Garrett, Scott Shriver, and Samuel Goldberg, 2022, Privacy & market concentration: Intended & unintended consequences of the GDPR, Working paper, Boston University, University of Colorado, Northwestern University.

Johnson, Garrett A, Scott K Shriver, and Shaoyin Du, 2020, Consumer privacy choice in online advertising: Who opts out and at what cost to industry?, Marketing Science 39, 33–51.

Kesler, Reinhold, 2022, The impact of Apple's app tracking transparency on app monetization, Available at SSRN 4090786.

Kesler, Reinhold, Michael E. Kummer, and Patrick Schulte, 2017, Mobile applications and access to private data: The supply side of the Android ecosystem, ZEW Discussion Paper 17-075, Unversity of Zurich, University of East Anglia.

Kummer, Michael, and Patrick Schulte, 2019, When private information settles the bill: Money and privacy in Google's market for smartphone applications, Management Science 65, 3470–3494.

Lin, Tesary, 2021, Valuing intrinsic and instrumental preferences for privacy, Marketing Science forthcoming.

Peukert, Christian, Stefan Bechtold, Michail Batikas, and Kretschmer Tobias, 2021, Regulatory spillovers and data governance: Evidence from the GDPR, Marketing Science forthcoming.

Preibusch, Soren, Dorothea K¨ubler, and Alastair R. Beresford, 2013, Price versus privacy: An experiment into the competitive advantage of collecting less personal information, Electronic Commerce Research 13, 423–455.

Prince, Jeffrey, and Scott Wallsten, 2021, How much is privacy worth around the world and across platforms?, Working paper, Indiana University, Technology Policy Institute.

Ramadorai, Tarun, Ansgar Walther, and Antoine Uettwiller, 2019, The market for data privacy, CEPR Discussion Paper DP13588, Imperial College London.

Sarma, Bhaskar P., Ninghui Li, Chris Gates, Rahul Potharaju, Cristina Nita-Rotaru, and Ian Molloy, 2012, Android permissions: A perspective combining risks and benefits, in Proceedings of the 17th ACM Symposium on Access Control Models and Technologies, 13–22.

Schmitt, Julia, Klaus M. Miller, and Bernd Skiera, 2020, The impact of privacy laws on online user behavior, Working paper, Goethe University Frankfurt, HEC Paris.

Tambe, Prasanna, Lorin Hitt, Daniel Rock, and Erik Brynjolfsson, 2020, Digital capital and superstar firms, Technical report, National Bureau of Economic Research.

Tang, Huan, 2019, The value of privacy: Evidence from online borrowers, Working paper, HEC Paris.

Tsai, Janice Y., Serge Egelman, Lorrie Cranor, and Alessandro Acquisti, 2011, The effect of online privacy information on purchasing behavior: An experimental study, Information Systems Research 22, 254–268.

Tucker, Catherine, A Agrawal, J Gans, and A Goldfarb, 2018, Privacy, algorithms, and artificial intelligence, The Economics of Artificial Intelligence: An Agenda 423–437.