Data Collection, Product Quality and Privacy Discrimination in Competitive Markets

Allouard, Hugo

Université Paris-Saclay

March 27, 2023

Abstract

Innovations in new technologies and information system has raised concerns in consumer privacy and leaded to an academic and policy debate about its regulation. Although privacy regulations have shown to be effective for the consumer privacy, adverse effects may accompany their introduction. In this paper, we focus on the privacy discrimination of digital markets. We build and estimate a structural model of supply and demand on the mobile app industry where developers cannot set different data collection between countries. Our contribution to the economic literature is two-fold. First, we show that using a structural model of demand, we can estimate the privacy sensitivity of consumers using observational data. The main advantage of this measure is to avoid the privacy paradox of surveys, and reproducibility issues of experiments. We find that French, German and Canadian consumers are privacy-sensitive while US consumers are not. Second, using counterfactual analyses we find that the market dominance of the US coupled with the non-privacy discrimination negatively impacts the privacy in other countries. We find that implementing a regulation for privacy discrimination between Europe and North America should increase consumer and firm surplus in all countries, but also the average product quality. However, such regulation increases the privacy in Europe but decreases it in North America.

1 Introduction

In the past decades, innovations in new technologies and information systems increased the firm interest in consumer data. Firm benefits of using consumer data are multiple. It can be used to improve product quality, to offer new services and personalized content based on consumer's characteristics (Acquisti *et al.*, 2016). Consumer data can also be as monetization strategies by increasing the effectiveness of advertising (Bleier & Eisenbeiss, 2015; Johnson *et al.*, 2017) or by selling data to other companies. Finally, the use of data allows to price discriminate consumers (Mikians *et al.*, 2013). However, the increased use of consumer data by firms has raised concerns about consumer privacy. In fact, the data collection and the misuse of this data can increase consumer costs. First, individuals may experience a psychological discomfort when personal information is revealed. Second, using consumer data for price discrimination allows the firm to capture a part of the consumer surplus (Taylor, 2004; Valentino-Devries *et al.*, 2012; Mikians *et al.*, 2012, 2013; Ichihashi, 2020). Third, consumers can be harmed by other forms of discrimination (Manant *et al.*, 2019). Finally, an increase in consumer costs can be induced by risk of data breaches, resulting in potential identity or credit card thefts for example. Costs and benefits associated to the use of consumer data results in trade-off between data disclosure and protection, and to a political and academic debate. What are the costs and benefits of consumer data? How do we compute them? Should we implement privacy regulations?

A part of the economic literature focused on estimating the privacy sensitivity of consumers. An obvious way would be to use of surveys asking people whether they care about their privacy. However, the *privacy paradox* makes impossible to estimate the privacy sensitivity of individuals using surveys. While individuals often claim that they are interested in data protection, their privacy concerns are rare translate into their decision-making (Joinson *et al.*, 2010; Athey *et al.*, 2017; Barth & De Jong, 2017; Barth *et al.*, 2019). Instead of estimating the privacy sensitivity of individuals, the use of surveys give an estimate of the intention to protect privacy. In order to estimate privacy sensitivity of individuals, researchers relied on experiments. Using laboratory experiments, a way to measure the privacy sensitivity is the price that individuals are willing to pay to protect their privacy. Such experiments have found evidences of a positive willingness to pay for privacy protection (Hann *et al.*, 2007; Tsai *et al.*, 2011; Preibusch *et al.*, 2013; Savage & Waldman, 2014). The inherent issue of laboratory experiments is the reproducibility of results and the external validity. These results are highly dependent of the contexts of the experiment and participants' characteristics and cannot be used to create temporal measure of the privacy sensitivity or reused in other economic analysis. We contribute to this stream of the literature showing that we can measure privacy sensitivity of consumers using observational data of products on competitive markets. Using observational data of mobile apps, we use the substitution effect between them to measure the privacy sensitivity of consumers. Using a structural model of demand (McFadden, 1986; Berry *et al.*, 1995) we estimate the effect of data collection on the demand for the mobile app controlling for all app characteristics. We find that collecting consumer data in France, Germany and Canada decreases the demand, but we do not find any evidence of privacy sensitivity in the US. Moreover, the privacy sensitivity is higher in France and Germany than in Canada.

Another part of the economic literature on privacy focused on the impact of privacy regulations on economic outcomes, and more specifically to adverse consequences of data regulations. On the health sector, Miller & Tucker (2011) find that the ePrivacy Directive reduced the hospital's adoption of electronic medical records systems, which has shown to deliver better health outcomes to patients. Another field of study is the online ad effectiveness which is a data intensive sector. Goldfarb & Tucker (2011); Johnson *et al.* (2021) find that the introduction of privacy regulations reduced ad effectiveness and firm profits. Finally, a part of this literature focused on the effect of privacy regulations on innovation and market structure. Goldfarb & Tucker (2012) emphasize that privacy regulations can change the market structure and decrease the innovation. Using a theoretical framework, Campbell *et al.* (2015) formalizes these predictions. The authors find that a privacy regulation based on the consumer consent for sharing data increases the cost of all firms in a competitive market, but small and new firms are the most adversely affected, leading to a concentration of data-intensive markets.

The empirical literature has validated these predictions showing that privacy regulations more costly for small and new firms, and find a decrease in innovation and an increase in market concentration (Janssen *et al.*, 2022; Johnson *et al.*, 2021; Peukert *et al.*, 2022). Our paper is closely related to this stream of the literature. In this paper we focus on studying the consequences of a (non-)privacy discrimination. We define the privacy discrimination as a situation where firms can set different data collection between markets. Our research questions are the following. What the effect of a privacy discrimination situation, compared to a non-privacy discrimination situation on consumer surplus, firm surplus, average product quality, and privacy in competitive markets? Should policymakers compel digital platforms to let firms privacy discriminate between markets?

To answer these questions, we build and estimate a structural dynamic choice model on the mobile app industry where developers strategically choice whether to invest in product quality and to collect consumer data. The interesting feature of the mobile app market, is the impossibility for firms to privacy discriminate between countries. More specifically, developers in the markets have to inform consumers about data collection before downloading the app. However, he has to inform consumers about data collection in all countries the app is commercialized, even it is not collecting and using consumer data in some of them. In addition, the part of the US in the total mobile game revenues is about 55% and we show in the paper that US consumers are not sensitive to data collection[1] while French, German and Canadian consumers are privacy-sensitive. In this situation, developers can have incentives to collect consumer data in order to increase the quality of the app and revenues in the US, even if it harms its demand and privacy in other countries. Our results show evidences of such behaviors. Finally, we estimate counterfactual scenarios where developers can privacy discriminate between countries. Our results suggest that a privacy discrimination has, at worst, no impact on consumer/firm surplus and the average product quality in all countries. With scale economies on costs, we find positive effects on these outcomes. Allowing developers to set a different data collection in North America and in Europe shows that the European Union should compel mobile app platforms to allow firms to privacy discriminate. In fact, our

[1] The data collection do not reduces the demand for the app

counterfactual shows an increase of privacy in european countries without harming consumer and firm surplus. However, the introduction of this regulation could decrease the privacy in the US and Canada.

The paper is divided as follows. In section 2 we write a model of demand and supply of the mobile app industry. In section 3 we describe the estimation procedure used to solve the model. In section 4 we describe the database, and show some descriptive statistics. In section 5 we describe the empirical strategy to solve the model, and show our results. In section 6 we show the results of our counterfactual analyses. Finally, in section 7 we discuss our results and the policy recommendation.

2 Model

We write a simple model of app industry equilibrium. Developers in our model take forward-looking decisions on their collection of consumer data, but also on their investments in product quality. In the app industry firms cannot discriminate in data collection by country. Hence, firms take decisions based on their current state but also on the industry state over all markets. We also make the hypothesis that products in the industry are experience goods and free to download and consume. However, consumers can purchase additional features through in-app purchases.

In the dynamic problem, time t is discrete. Each app j is commercialized in a set of country C_j and defined by a set of state variables ω_j. Since mobile apps are experience goods, we define two state variables describing the quality of the product: ξ_j, is the *perceived* quality of the app before downloading it, and q_j id the *actual* quality of the app, revealed only after downloading it. We define DC_j, the data collection type, indicating whether the app is collecting consumer data. Finally, $dl_{jt} \equiv [dl_{jt,t}, dl_{jt,t-1}, ..., dl_{jt,0}]$ where $dl_{jt,\tau}$ is the number of time an app j at time t has been downloaded during the period τ. In each market m, combination of a country c and a period t, n_m commercialized apps are observed by the econometricians.

We define the utility of a consumer i in country c at time t for app j as follows:

$$U_{ijtc} = -\alpha_c DC_{jt} + \gamma X_{jt} + \tilde{\xi}_{jtc} + \epsilon_{ijtc} \tag{1}$$

where X_{jt} are observed app characteristics. $\tilde{\xi}_{jtc}$ is a dimension of quality differentiation unobserved by the econometrician. We allow the parameter α, measuring the effect of consumer privacy concerns on demand, to vary across country c. Assuming that the idiosyncratic taste of consumers ϵ_{ijtc} is i.i.d extreme value distributed, we obtain the following equation of market share:

$$S_{jtc} = \frac{exp(-\alpha_c DC_{jt} + \gamma X_{jt} + \tilde{\xi}_{jtc})}{1 + \sum_{k \in n_m} exp(-\alpha_c DC_{kt} + \gamma X_{kt} + \tilde{\xi}_{ktc})} \tag{2}$$

Normalizing the outside option of downloading another app than observed ones S_{0tc} to 0 and using the Berry *et al.* (1995) 's inversion, we obtain the following estimable equation:

$$ln\left(\frac{S_{jtc}}{S_{0tc}}\right) = -\alpha_c DC_{jt} + \underbrace{\gamma X_{jt} + \tilde{\xi}_{jtc}}_{\xi_{jtc}} \tag{3}$$

We define ξ_{jtc} as the mean app quality *perceived* by consumer before downloading the app.

On the supply side, we do not add structure to the per-period profit equation. First, all observed apps are free to download[2] and do not require to model price competition between firms. Second, we have a lack of data to properly model firm revenues. We do not have data on the number of active users, time usage, information about in-app purchases characteristics and prices or ad frequency necessary to properly model the revenue. However, we emphasize that a reduced form equation should not significantly impact our results. After the download of an app, competition resulting from the use of different monetization strategies, with different intensities should be completely captured by a measure of app quality. In fact, an app with intrusive advertisements should have a lower quality resulting in a lower per-download usage, spending and henceforth, revenue. We define as follows the revenue equation of an

[2]In May 2022, more than 95% of mobile games are free to download.

app j at time t:

$$R_{jt} = \sum_{c \in C_j} R_{jtc}(dl_{jtc}, q_{jt}|\theta_r) \tag{4}$$

where R_{jtc} is the revenue of app j at time t in country c. $dl_{jtc} \equiv [dl_{jtc,t}, dl_{jtc,t-1}, ..., dl_{jtc,0}]$ is a vector of present and past number of downloads. Specifically $dl_{t,\tau}$ is the number of time an app j at time t has been downloaded during the period τ. q_{jt} the *actual* quality of the app discovered only after downloading and consuming the app. In our model the *perceived* quality impacts only indirectly revenue through the demand for app and the number of downloads.

We now discuss the dynamic problem of the firm. In our model, we consider that developers take forward-looking decisions about data collection, and quality investments. We denote by *ndc* an app with the *non-data collection* type, and *dc* apps with *data collection* type. At each period t, each app j must choose an action $a \in A^k$ with $k \in [ndc, dc]$ where A^k is the set of possible actions for an app of type k. The set of possible actions is different between data collection types:

$$\begin{aligned} A^{dc}_{jt} &= \begin{cases} Update^{dc}_{jt} \\ dc_{jt-1} \rightarrow ndc_{jt} \ \& \ Update^{ndc}_{jt} \\ \emptyset_{jt} \end{cases} \\ A^{ndc}_{jt} &= \begin{cases} Update^{ndc}_{jt} \\ ndc_{jt-1} \rightarrow dc_{jt} \ \& \ Update^{dc}_{jt} \\ \emptyset_{jt} \end{cases} \end{aligned} \tag{5}$$

$Update^{k}_{jt}$ is taking value 1 if the developer of the app of type k decides to update the app. In our model, update decision refers to the decision of investing in app quality. In fact to introduce new features or to correct bugs in the app, developers have to release a new version which is only possible with an update of the app. The second possible action for developers is to change his data collection type, and is denoted $k_{jt-1} \rightarrow k'_{jt}$ where k is the app data collection type and k' the other type. In the case where the developer decides to start or stop the data collection mobile app platforms compel them to release a new app version. Finally, developers can decide to do nothing (i.e to not update the app and to not change the data collection type).

We now discuss how our state variables $\omega_{jtc} \equiv [DC_{jt}, \xi_{jt}, q_{jt}, dl_{jtc}]$ that evolve from one period to the next one. The evolution of DC_{jt} and dl_{jtc} is deterministic. DC_{jt} is only determined by the data collection type of the previous period, and the action chosen by the developer. The data collection type remains unchanged unless the developer decides to change it. The actual number of downloads is the product of the market share S_{jtc} and the market size I_{tc}: $dl_{jtc,t} = S_{jtc} I_{tc}$ and past downloads are defined by $dl_{jtc,\tau} = dl_{j\tau c,\tau}, \forall \tau < t$. For ξ_{jt} and q_{jt}, the evolution from one period to next one is probabilistic. The evolution of ξ_{jt} depends on its own *perceived* and *actual* quality, and whether the app is updated. An increase in *actual* quality should increase the *perceived* quality of the app because of word to mouth or press reviews. Another potential mechanism explaining the positive effect of the *actual* on *perceived* quality comes from the algorithm of the *Google Play Store* to rank apps. While algorithms are not well known, it is probably the case that high *actual* quality apps are more likely to be promoted by the platform resulting in a decrease of search costs. In the same spirit, updating the app has a positive effect on the *perceived* quality even if there is no quality improvement. Mobile app platforms have categories for recently updated app only. Only apps updated in the last 30 days are published, resulting in a decrease of search costs to find these apps. We assume that the *perceived* quality is bounded and can only take a finite number of discrete values. We denote the evolution function of ξ_{jtc} as follows:

$$p_\xi(\xi_{jtc}|\xi_{jt-1c}, q_{jt}, Update_{jt}) = g_\xi(\xi_{jt-1c} + \Delta_\xi|\xi_{jt-1c}, q_{jt}, Update_{jt}) \qquad (6)$$

where Δ_ξ is the change in ξ from one period to the next one, and $g_\xi(.)$ is a non-parametric function. The evolution of q_{jt} is quite similar. The change in q_{jt} from one period to the next one is function of the *actual* quality and whether the app is updated. However, we consider two types of updates according to the data collection type $Update^{DC} \equiv [Update^{dc}, Update^{ndc}]$. We make the hypothesis that using consumer data allows developers to introduce new features that are impossible to implement without consumer data. Consumer data can also be used to understand better the preferences of consumers, and to improve existing features with personalized contents and recommendations. Finally, consumer data can also be used for targeted ad that are preferred by consumers than non-targeted ads. We assume that the *actual* quality is bounded and can only take a finite number of discrete values. We denote the evolution function of q_{jt} as follows:

$$p_q(q_{jt}|q_{jt-1}, Update^{DC}_{jt}) = g_q(q_{jt-1} + \Delta_q|q_{jt-1}, Update^{DC}_{jt}) \tag{7}$$

where Δ_q is the change in q from one period to the next one, and $g_q(.)$ is a non-parametric function.

The decision to invest in app quality and the choice of the data collection type are decided strategically based on the current costs θ_c and the expected future stream of revenues. Assuming stationary, dropping indexes c and t and denoting by prime the next period for simplicity, we write the dynamic problem of developers as follows :

$$\begin{aligned} V_j(\omega_j, \omega_{-j}|\theta) = \max_{a^k_j \in A^k} \Big\{ & R_j(\omega_j, \omega_{-j}|\alpha_c, \theta_r) - c_j(a^k_j, Update_DC_j, dl_j|\theta_c) \\ & + \beta E\big[V_j(\omega'_j, \omega'_{-j}|\omega_j, \omega_{-j}, a^k_j, \theta)\big]\Big\} \end{aligned} \tag{8}$$

where V_j is the value function. β is the discount factor. $c(.)$ is the cost function depending on the chosen action, the type of update, and the number of downloads, a proxy for app usage. Finally, θ is the whole set of parameters: $\theta \equiv [\alpha_c, \theta_c, \theta_r]$. In the dynamic problem, the value function depends on his own state ω_j, but also to the state of the entire industry ω_{-j}.

The solution of equation (8) is a strategy profile $a_j = \Omega_j(\omega_j, \omega_{-j}|\Omega_{-j})$. We assume the existence of Markov Perfect Equilibrium (*MPE*) satisfying the following condition:

$$V_j(\omega_j, \omega_{-j}|\tilde{\Omega}_j, \tilde{\Omega}_{-j}) \geq V_j(\omega_j, \omega_{-j}|\Omega^*_j, \tilde{\Omega}_{-j}) \tag{9}$$

where $\tilde{\Omega}_j$ is the markov perfect equilibrium strategy for the developer of app j and Ω^*_j any alternative strategy and represents a unilateral deviation from the equilibrium.

3 Estimation procedure

3.1 Oblivious equilibrium

Different methods have been proposed to find the *MPE* of the dynamic game such as Bajari *et al.* (2007); Aguirregabiria & Mira (2007); Pakes *et al.* (2007) methods. However, due to the large number of applications competing in the industry these methods are computationally intractable. In fact, it would require to compute all unilateral deviations from the equilibrium of each developer leading to a computational burden. Instead, we use the concept of Oblivious Equilibrium (*OE*) introduced by Weintraub *et al.* (2008). The *OE* can be used to approximate the *MPE* in the particular context of a large number of players. More specifically, Farias *et al.* (2012) states that *OE* approximates better *MPE* as the number of players increases and the market is not overly concentrated. The idea behind the notion of *OE* is that, when the number of player is large, changes in the state of the industry would average and remains constant over time. In this case, making a decision looking at the industry steady state is nearly the same as looking at the state of all competitors. Moreover, this assumption is particularly suitable for our analysis. Considering the very large number of mobile apps on mobile app platforms, it is very unlikely that developers have information about the state of all competitors. Hence, they are more likely to make decision based on aggregated measures of the industry state. We rewrite equation (2) using the notion of *OE*:

$$S_{jtc} = \frac{exp(-\alpha_c DC_{jt} + \xi_{jtc})}{1 + \sum_{k \in DC} \sum_{l \in \xi} n_{tc}^{k,l} \; exp(-\alpha DC_t^k + \xi_{tc}^l)} \tag{10}$$

where $n_{tc}^{k,l}$ is the expected number of apps at steady state with a data collection k and a *perceived* quality l in country c at time t. The demand side equation no longer depends on the data collection and *perceived* quality of each competitor, but only the expected number of apps at each possible state. It simplifies the dynamic game model to a single agent model which is much easier to solve and avoid the computational burden. Thus, any method proposed in the literature can be used to solve the single agent model.

3.2 Method

Although many methods can be used to solve this single agent dynamic choice model, we rely on a modified version of the method proposed by Hotz *et al.* (1994) to recover all parameters of the model. Since we have a large number of states in our model, using simulations of value functions allow us to decrease the computational burden. The only difference with their method, is the computation of the industry steady state. The method is divided in two steps. In the first step, we estimate the demand and revenue equations. We also estimate a matrix of state transitions for each possible states in our model, and estimate policy functions. Using these estimations, we approximate the value function for each possible action in each observed state. More precisely, we simulate random draws of actions and state transitions for T periods, and repeat the process with S simulations. We average the S simulations to have an estimate of the value function. In the second step, we estimate the cost parameters using our approximation of value functions.

4 Data

4.1 Mobile games

We restrict our sample to mobile games for several reasons. First, the estimation of the demand side equation (2) requires to have a precise definition of markets and substitute apps. Leyden (2019) shows that all non-game applications are not direct competitors, even among apps in the same category. Moreover, some non-game apps are complements to other services or products and do not compete directly with other apps in the market. In that case, the demand for the app do not reflect apps' competition but rather a competition between services or product in another industry. For example, the choice to download a particular bank app rather than another is not related to the app characteristics, but rather to the individual's choice of a bank made before installing the app. Another issue with non-game apps is to know whether the data collected is strictly necessary for its functioning. For example, GPS apps need to collect the consumer location to operate, and developers have no other choice than collecting and use this data. The choice of the data collection in that case is no longer a strategic behavior. Estimating the demand model with non substitute apps could lead to biased estimates. We argue that these issues are less a concern in the mobile game industry. First, all mobile games are used and downloaded for an entertainment purpose making them all potential substitutes. Second, data collection is almost never necessary for its functioning. We drop games from our dataset for which the data collection is mandatory, such as *PokemonGo*

4.2 Sample

Our sample is composed of mobile games commercialized on the *Google Play Store*. We construct our dataset using scrapped data from the *Google Play Store* and *AppTweak*. It includes free games appearing at least once in the top 500 mobile games between September 2015 and March 2021 in Canada, France, Germany or in the US. For each app in our sample, we observe its characteristics[3] but also the number of downloads and the revenue from in-app purchases by country, each day. We also know which consumer data are collected through the game, and we observe changes in the data collection of the game. Our model requires to define the *actual* quality q_{jt} and the data collection DC_{jt}. As a measure of the *actual* quality of the game, we use the average of new grades of the game j during the period t[4]. These grades are left by the users who downloaded the game and range from 1 to 5 where 1 is the worst rating. As a measure of the data collection, we use a dummy variable equal to one if the game collects the identity[5], the location[6] of the user, or both of them. Although other type of personal data are collected by games, the identity and the location represent

[3]Observed characteristics are those displayed on the google page of the game.

[4]We only take grades left by consumers during the period t

[5]Refers to the permission "*read phone status and identity*" on the *Google Play Store*

[6]Refers to permissions "*precise location (GPS and network-based)*", "*approximate location (network-based)*", "*access extra location provider commands*" and "*mock location sources for testing*" on the *Google Play Store*

a large share of collected data[7]. We decided to define the period t at the month level, and aggregate our data to the month. Our measure of the *actual* quality is based on grades left by consumers. For smallest games in our dataset, the number of grades by day or week is very low and do not necessarily reflect the quality of the game. It would be very imprecise to measure the state transition of the *actual* quality with daily or weekly data.

Table 1: Descriptive statistics

	Mean	Std. Dev.	Min	Max
Revenue (x1000) - R_{jt}	374	1594	0.003	37207
Downloads (x1000) - $dl_{jt,t}$	91.7	148	0.003	2200
Grade - q_{jt}	3.91	0.949	1	5
Data collection - DC_{jt}	0.26	-	0	1
Update - $Update_{jt}$	0.55	-	0	1
Ads	0.89	-	0	1
Obs.		42174		

R_{jt} and $dl_{jt,t}$ are the total revenue and number of downloads of the four countries.

Table 1 presents some descriptive statistics. The average revenue of an app is close to \$374000. However, revenues are extremely heterogeneous between games. A game is downloaded 91700 times by month in average, and the spending in in-app purchases by download is around \$4. Th average grade of an app is close to 3.91. 26% of games are collecting the identity, the location of consumers or both of them. Finally, in our dataset, 89% of games display ads. Our dataset also includes the category of the game.

5 Estimation methodology and results

5.1 Demand

We estimate the demand equation (3) to recover parameters α_c and values ξ_{jtc}. In our model, α_c is the privacy sensitivity of consumers. To obtain an unbiased estimate of the privacy sensitivity of consumers, we need to ensure that DC_{jt} is uncorrelated to unobserved

[7] Any other permissions are used by less than 1% of games

characteristics $\tilde{\xi}_{jtc}$ in our estimation. Most of unobserved characteristics of mobile games are invariant from a period to the next one and the only way for developers to change these unobserved characteristics is to update the game. We take advantage of that to control for time-fixed unobserved characteristics. Following Leyden (2019); Sweeting (2010) we assume an AR(1) process on the error term:

$$\tilde{\xi}_{jtc} = \rho\tilde{\xi}_{jt-1c} + \mu Update_{jt} + \sigma_{jtc} \tag{11}$$

where $\tilde{\xi}_{jt-1c}$ are unobserved characteristics of the previous period and σ_{jtc} is an i.i.d mean-zero shock. With this specification, $\tilde{\xi}_{jt-1c}$ controls for time-fixed unobserved game characteristics and $Update_{jt}$ controls for a change in those characteristics, but also in search costs as explained in section 2. In addition, we use Arellano & Bond (1991) instruments on the data collection variable. More precisely, we instrument the data collection at time t by the data collection at time $t-2$. We estimate the demand model using non-linear least square (NLLS) and two stage non-linear least square (TSNLLS) models.

Table 2 shows the results of the demand estimations. Coefficients associated to DC_{jt} variables in the table are our estimates of α_c. Our estimated coefficients show a negative effect of data collection on demand for the product in France, Canada and Germany. Nevertheless, the privacy sensitivity of consumers in Canada is lower than in France and Germany. In the opposite of these results, we do not find any evidence of a privacy sensitivity of consumers in the US. Both estimations mostly show the same results. The only difference is the magnitude of the coefficients where the use of instruments increases the negative effect of data collection on demand. This change can be explained by a bias introduced by the unnecessary use of instruments or because the instruments controls for additional unobserved characteristics. However, magnitudes are quite similar, and statistical significances and the privacy sensitivity ranking among countries are similar. Our results also show a large coefficient ρ meaning that unobserved characteristics are highly correlated and invariant from one period to the next one. Finally, coefficients associated to the $Update_{jt}$ variables indicate that an update increases the

Table 2: Demand Estimation

	NLLS	2SNLLS
DC_{jt} - fr	-0.0210 ***	-0.0305***
	(0.0055)	(0.0089)
DC_{jt} - de	-0.0211***	-0.0299***
	(0.0058)	(0.0087)
DC_{jt} - us	-0.0011	-0.0064
	(0.0054)	(0.0084)
DC_{jt} - ca	-0.0104**	-0.0173**
	(0.0056)	(0.0084)
q_{jt}	0.0260***	0.0258***
	(0.00130)	(0.00134)
ads_j	-0.0090**	-0.0087***
	(0.0044)	(0.0039)
$Update_{jt}$	0.0469***	0.0470***
	(0.0011)	(0.0025)
ρ	0.9427***	0.9426***
	(0.0011)	(0.0010)
Category FE	Yes	Yes
Time FE	Yes	Yes
Country FE	Yes	Yes
Obs.	118981	118981

Bootstrap standard errors.
P-values: *** <0.01, ** <0.05 * <0.1.
Dependent variable: $log\left(\frac{S_{jtc}}{S_{0tc}}\right)$

unobserved quality of the game. The release of a new version of the game allow developers to implement new features, to improve the existing ones and to correct bugs. In that case updates can increase the quality of the app, and consumers have access to some information about that through ratings, reviews and the description.the search cost for consumer The update also reduces consumer search costs as discussed in section 2 and developers internalize it during the decision-making Comino *et al.* (2019).

5.2 Revenues

Unfortunately we do not have the data to model the revenue of the developer. First, all our games are free to download. We cannot model the competition in price which is standard in demand-supply structural models. Second, we don't have any data about the per consumer spending in in-app purchase and the number of active users. Without these data, it's impossible to properly build a revenue equation. Instead, we estimate app revenues using a reduced form equation:

$$log(R_{jtc}) = \phi_0 + \phi_{1c} log(dl_{jtc,t}) + \phi_{2c} log(\sum_{\tau=1}^{4} dl_{jtc,t-\tau}) + \phi_{3c} log(\sum_{\tau=5}^{8} dl_{jtc,t-\tau}) + \phi_4 q_{jt} + \epsilon_{jtc} \quad (12)$$

With this specification, the revenue is a function of actual and past downloads. A consumer downloading a game now can still spend money to the game in future periods. Our specification implicitly assumes that after 9 months consumers no longer bring money to the firm. We sum revenues from periods $t-1$ to $t-4$ and $t-5$ to $t-8$ to avoid multicollinearity issues. We also allow the *actual* quality to have an effect on revenues. Consumers are likely to spend more money to higher quality games. We estimate the previous equation with an OLS.

Table 3 shows the result of the revenue estimation. Results show that an increase in the number of downloads increases the revenue of the app. We also find that the per-download spending is decreasing with the time. We explain this result by a decreasing number of users over time but also because active users already bought additional features in the past. Our results also show that per consumer spending is higher in the US than in other countries. Finally, the *actual* app quality increases consumers' spending in the app. Using equation (4) we can compute the total revenue of the game.

5.3 Policy functions

To solve the dynamic agent model we need to recover the probability of each action $P(a_{jt}|\omega_{jt-1c})$. In our specific case, we can't estimate directly the probability to change the data collection

Table 3: Revenue estimation

dep. variable : $log(R_{jtc})$	Coefficient	std. err.
$log(dl_{t,t})$	0.5356***	(0.023)
fr	-0.0082	(0.034)
de	0.0655*	(0.035)
us	0.1674***	(0.043)
$log(\sum_{\tau=t-4}^{t-1} dl_{t,\tau})$	0.2545***	(0.031)
fr	-0.0210	(0.045)
de	-0.0020	(0.047)
us	-0.0670	(0.058)
$log(\sum_{\tau=t-8}^{t-5} dl_{t,\tau})$	0.0962***	(0.021)
fr	0.0085	(0.030)
de	0.0139	(0.032)
us	0.0067	(0.039)
q_{jt}	0.0304***	(0.008)
ads	-2.1405***	(0.020)
Category FE	Yes	
Time FE	Yes	
Country FE	Yes	
R^2	0.45	
Obs.	118981	

Robust standard errors
P-values: *** <0.01, ** <0.05 * <0.1.
Reference group for $log(.)$ variables are the Canada (ca)
Coefficients and standard errors associated to variables fr, de and us are interaction effects.

of the game because we rarely observe this decision. Instead, we estimate the probability of collecting data $P(DC_{jt}|\omega_{jt-1c})$ and updating the game $P(Update_{jt}|\omega_{jt-1c})$. Using estimated probabilities, we compute action probabilities using the following formulas:

$$\begin{aligned} P(k_{jt-1} \to k'_{jt} \ \& \ Update^{k'}_{jt}|\omega_{jt-1c}) &= P(DC_{jt} = k'|DC_{jt-1} = k, \omega_{jt-1c}) \\ P(Update^{k}_{jt}|\omega_{jt-1c}) &= P(Update^{k}_{jt}|\omega_{jt-1c}) - P(k_{jt-1} \to k'_{jt} \ \& \ Update^{k'}_{jt}|\omega_{jt-1c}) \\ P(\emptyset_{jt}|\omega_{jt-1c}) &= 1 - P(k_{jt-1} \to k'_{jt} \ \& \ Update^{k'}_{jt}|\omega_{jt-1c}) - P(Update^{k}_{jt}|\omega_{jt-1c}) \end{aligned} \tag{13}$$

where $k_{jt} \in [ndc_{jt}, dc_{jt}]$ and k'_{jt} the alternative to k. We estimate probabilities of collecting data and updating the game using logistic regressions. As denoted in our previous equations, probabilities are estimated conditionally on the state of the game.

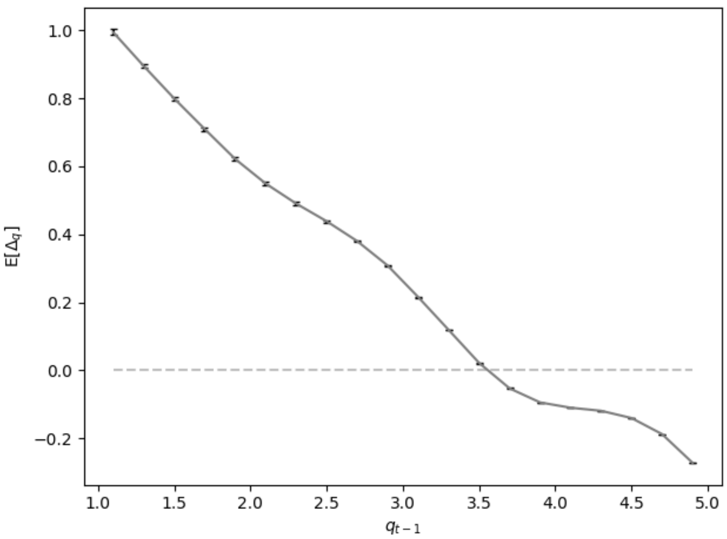
Table 4: Empirical policy function

	DC_{jt}	$Update_{jt}$
q_{jt-1}	-0.2880	-1.0582***
	(1.176)	(0.337)
q^2_{jt-1}	0.1172	0.3837***
	(0.387)	(0.111)
q^3_{jt-1}	-0.0161	-0.0371***
	(0.040)	(0.011)
DC_{jt-1}	7.5371***	-0.2293***
	(0.075)	(0.025)
Ads	0.0842	-0.4935***
	(0.122)	0(0.036)
Category FE	Yes	Yes
Time FE	Yes	Yes
Country FE	Yes	Yes
R^2	0.84	0.06
Obs.	42174	42174

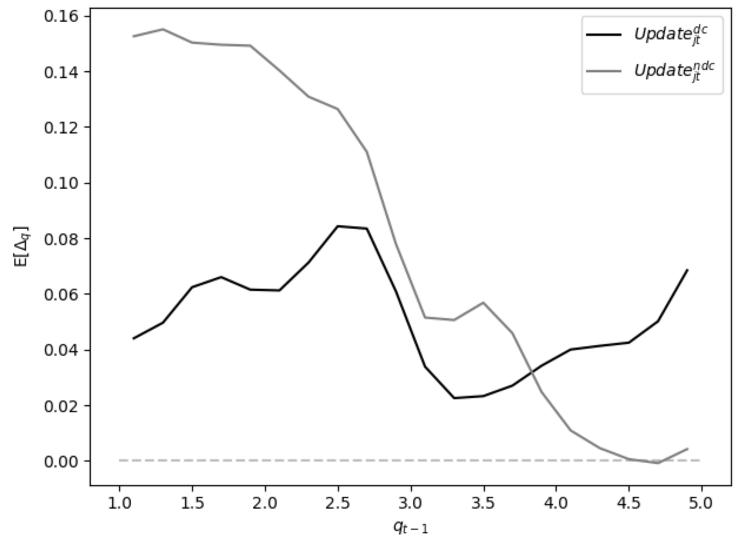
P-values: *** <0.01, ** <0.05 * <0.1.

Table 4 shows the results of policy function estimations. The first column shows estimated coefficients for the data collection variable. Results indicate that the *actual* quality of the game has no effect on the probability to collect consumer data. However, the parameter associated to DC_{jt-1} is positive and significant meaning that collecting data at $t-1$ increases the probability to continue so in the next period. The second column shows estimated coefficients for the update decision. In this estimation, we add a control $Change\ collect_{jt}$ taking value 1 if the data collection changed between $t-1$ and t. Omitting this variable could lead to biased estimates because updating an app is mandatory when changing the data collection. Results show that the *actual* quality of the game has a significant impact on the updating decision. However, the interpretation of the coefficients is not straightforward and the relation is complex. Our results also indicates that games collecting consumer data are less likely to be updated. There are two potential explanations for that. First, as we will show latter in the paper, the collection and use of consumer data rise costs. Updates using consumer data not necessarily replace those without consumer data, increasing the

Figure 1: State transition of the *actual* quality



1. Expected change in q_{jt}



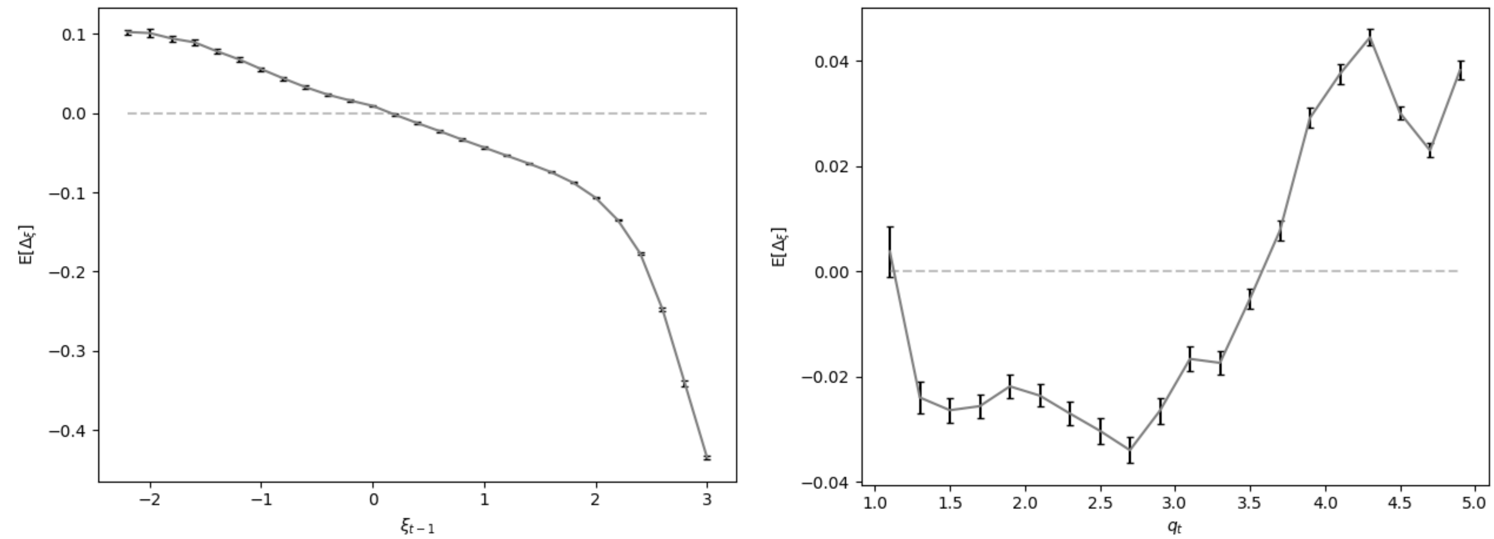
2. Average effect of updates on expected change in q_{jt}

total cost. Moreover, the use of consumer data can increase the update cost for example with an increased use of servers and an additional cost of data processing. Second, fewer updates with consumer data than without consumer data are required to reach the same app quality. It is a possible explanation if updates with data collection increase more the quality than without data collection. Using these estimates, we can compute the probability of each action at each observed state using equations (13).

5.4 State transitions

To estimate state transitions, we use a conditional kernel density estimation. This model allows us to obtain an estimate of the density of changes in state variable conditional on some explanatory variables. The use of kernel density potentially allows a positive probability for each potential state, as defined in our model. Since our model is non-parametric, we use graphics to show main results of our estimations. To do so, we discretize ξ_{jt} and q_{jt} using equal steps of 0.2. The value of each discretion is equal to the mean between the lower and upper bound of the discretion. The discretion leads to 20 possible values for q_{jt} and 31 for ξ_{jtc}. Until the end, we only use the discrete version of these variables. Figure 1.1. shows the average effect of *actual* quality q_{jt-1} on the expected change of *actual* quality

Figure 2: State transition of the *perceived* quality



1. Expected change in ξ_{jtc}

2. Average effect of q_{jt} on expected change in ξ_{jtc}

$E[\Delta_q]$ from period $t-1$ to t. We find that the expected change in the app *actual* quality is strictly decreasing with the app quality. We also find a cutoff at a q_{jt-1} close to 3.5 where the expected change is zero. This negative relationship is intuitive and found in other industries (Hashmi & Biesebroeck, 2016). Figure 1.2. shows the effect of an update with and without data collection, according to the *actual* quality q_{jt-1}, on the expected change of *actual* quality $E[\Delta_q]$ from period $t-1$ to t. First, we show that updates always have a positive effect on the expected change of the *actual* quality. Second, the average effect of an update is decreasing with the *actual* app quality. It implies that it is harder to innovate for higher quality games. Figure 2.1. is the same as figure Figure 1.1. replacing the *actual* quality but the *perceived* quality on both axis. We find that the expected change in the app *perceived* quality is strictly decreasing with the *perceived* app quality. We also find a cutoff in ξ_{jt-1c} close to 0.3 where the change in *perceived* quality equals 0. Figure 2.2. shows the average effect of *actual* quality q_{jt} on the expected change in *perceived* quality $E[\Delta_\xi]$. Our results show a positive effect of *actual* quality on *perceived* quality. This result was expected. First, some new features introduced in the game can be observed on the *Google* page of the game before downloading the game. For example with the description of the game. Second, an increase in quality should positively impact reviews and ratings from users, partially observed before downloading the game. Results also show that effect of *actual* quality is quite stable for low-quality app. The result indicates that consumers do not make differences between low and very low-quality apps. Finally, our results show non-intuitive results for extreme *actual* quality apps. Two reasons can explain these results. First, the number of observation is small for extreme *actual* qualities leading to non-representative results. Second, apps with extreme *actual* qualities have different characteristics than others. In fact, games with a low number of reviews are more likely to obtain extreme qualities and an unexpected bug in the app can lead temporarily to an extremely low quality.

5.5 Costs

We define the cost function as follows:

$$c_{jt}(Update_{jt}, dl_{jt,t}, \nu_{jt}|\theta_c) = \theta_{c1}Update^{dc}_{jt} + \theta_{c2}Update^{ndc}_{jt} + \theta_{c3}dl_{jt,t} + \theta_{c4}\nu_{jt} \tag{14}$$

The total cost is function of updates with ($Update^{dc}_{jt}$) and without ($Update^{ndc}_{jt}$) data collection. We also allow the number of actual download to affect the cost. In fact, an increase in users can raise costs because of an increase in computation capacity, push notifications, storage space to save user profile etc. We assume that the error term ν_{jt} is T1 extreme value distributed, and we add a random coefficient on errors to control for the unobserved heterogeneity in payoffs.

As detailed in section section 3 we use all previous results to approximate value functions of each observed state. Using linearity of parameters θ_c in the cost function, we can write the value function for a specific action a_{jt} as follows:

$$
\begin{aligned}
V_j(a_j|\omega_j,\omega_{-j}) = {} & E_{a_j,\omega_j,\omega_{-j}}\sum_{t=0}^{\infty}\beta^t R_{jt} \\
& - \theta_{c1}E_{a_j,\omega_j,\omega_{-j}}\sum_{t=0}^{\infty}\beta^t Update_{jt}^{ndc} \\
& - \theta_{c2}E_{a_j,\omega_j,\omega_{-j}}\sum_{t=0}^{\infty}\beta^t Update_{jt}^{dc} \\
& - \theta_{c3}E_{a_j,\omega_j,\omega_{-j}}\sum_{t=0}^{\infty}\beta^t dl_{jt,t} \\
& + \theta_{c4}E_{a_j,\omega_j,\omega_{-j}}\sum_{t=0}^{\infty}\beta^t \nu_{jt}
\end{aligned}
\tag{15}
$$

The advantage of a linear cost function in parameters is that we have to compute only once the value function. Otherwise, we have to recompute value functions for each guess of θ_c during the minimization procedure. We assume a discount factor $\beta = 0.95$. We approximate value functions using $T = 100$ periods. For each observed states, we simulate paths for each possible action a_{jt} at $t = 0$. We simulate random draws of the value function $S = 500$ times and average them to obtain an approximation of the value function $\tilde{V}_j(a_j|\omega_j,\omega_{-j})$.

With our estimates of value functions in hands, we recover now θ_c parameters. Since ν_{jt} is T1 extreme value distributed, we can write the probability of action a_{jt} as follows:

$$
\tilde{p}(a_j|\omega_j,\omega_{-j},\theta_c) = \frac{exp\left(\tilde{V}_j(a_j|\omega_j,\omega_{-j},\theta_c)\right)}{\sum_{a'_j\in A^k} exp\left(\tilde{V}_j(a'_j|\omega_j,\omega_{-j},\theta_c)\right)}
\tag{16}
$$

Following Hotz *et al.* (1994) we look for θ_c minimizing the squared distance between probabilities of action computed with the empirical policy equations (13) and probabilities estimated using value functions in equation (16):

$$
\hat{\theta}_c = argmin_{\theta_c}\ \left[\hat{p}(a_j|\omega_j,\omega_{-j}) - \tilde{p}(a_j|\omega_j,\omega_{-j},\theta_c)\right]^2
\tag{17}
$$

Table shows the result of cost estimates. The first column shows results without the number of downloads, and the second one with the use of the number of downloads. Results in both specifications are quite similar. Our results indicate that the cost to update the game without data collection is around \$12500. The cost of an update with data collection is equals to \$25000. The number of downloads increases costs as expected. However, the magnitude of the coefficient shows that it's quite insignificant compared to the cost of an update. We also find a large random coefficient indicating that heterogeneity in costs and revenues is large. This result was expected because app revenues are very heterogeneous.

6 Counterfactuals

To answer the main question of the paper, we rely on counterfactual scenarios. Our baseline is the actual situation of the mobile app industry where developers cannot privacy discriminate between countries. Our counterfactual scenarios are situations where developer can set different data collection for different countries. More precisely, a developer can now decide to create two versions of the same game. One is commercialized in a set of countries and is collecting consumer data, the other is commercialized in another set of countries and is not collecting consumer data. However, we let the possibility for developers to commercialize only one version of the game in all countries. For our counterfactual scenarios, we consider two sets of countries $K \equiv [K_1, K_2]$. Each developer chooses actions a_{jt}, one by set of countries, maximizing their value functions. We rewrite equation (8) in the case of the counterfactual scenario:

$$\begin{aligned} V_j(\omega_j, \omega_{-j}|\theta) = &\sum_{\kappa \in K} \max_{a^k_{j\kappa} \in A^k} \Big\{ R_{j\kappa}(\omega_j, \omega_{-j}|\alpha_c, \theta_r) - c_{j\kappa}(a^k_{j\kappa}, Update^{DC}_j, dl_j|\psi_j, \theta_c) \\ &+ \beta E\big[V_{j\kappa}(\omega'_j, \omega'_{-j}|\omega_j, \omega_{-j}, a^k_{j\kappa}\psi_j, \theta)\big] \Big\} \end{aligned} \tag{18}$$

We also change the cost function (14) in our counterfactual scenarios, introducing a ϕ_{jt}, a discount factor on the cost of updates without data collection:

$$c_{jt}(\psi_j, Update_{jt}, dl_{jt,t}, \nu_{jt}|\theta_c) = \theta_{c1} Update^{dc}_{jt} + \psi_{jt}\theta_{c2} Update^{ndc}_{jt} + \theta_{c3} dl_{jt,t} + \theta_{c4}\nu_{jt} \tag{19}$$

where:

$$\psi_{jt} = \begin{cases} \Psi \geqslant 0 & \text{if } Update^{dc}_{jt\kappa} = 1 \ \& \ Update^{ndc}_{jt\kappa'} = 1 \\ 1 & \text{Otherwise} \end{cases}$$

ϕ_{jt} represents the scale economy of updating the game without data collection while you also update the other version of the game with data collection. In fact, with two versions of the same game, developers have incentives to introduce features not requiring consumer data in both versions because he only has to pay once the development of the update. However, there is a potential cost to introduce the new developed feature in each version justifying that Ψ not necessarily equals to 0.

We consider two counterfactual scenarios. First, the developer can create a version of the game for the US and another one for Canada, France and Germany. Even though isolating the US only is not a credible policy recommendation, this scenario is particularly interesting. As shown previously, the US is the country with the highest revenues and revenues per consumer, but is also the country with the lowest privacy sensitivity of consumers. Second, the developer can create a version of the game in the US and Canada and another one in France and Germany. This scenario is probably more credible since the privacy regulations are done by the European Union. For each scenario, we consider 3 cases $\Psi \in [0, 0.5, 1]$. We compute the percentage change, from the baseline to the counterfactual, in consumer surplus, firm surplus, average actual quality of apps and the average number of consumers sharing their data.

Table 5 shows the result of our counterfactual analysis. For the case of $\Psi = 1$, in both scenarios, the consumer surplus (CS), firm surplus (FS) and the average of *actual* quality game (q_{jt}) are similar to the baseline. However, there are changes in the number of consumer disclosing data (DC_{jt}). In the scenario isolating the US, the number of consumer disclosing data is decreasing by 5.6% in Canada, 16.2% in Germany and 9.23% in France. However, we observe an increase in the US by 1.57%. In the scenario where the discrimination is possible between North America and Europe, the number of consumer disclosing data increase by

Table 5: Counterfactual results

		$\Psi = 1$			$\Psi = 0.5$			$\Psi = 0$		
		%ΔCS	%Δq_{jt}	%ΔDC_{jt}	%ΔCS	%Δq_{jt}	%ΔDC_{jt}	%ΔCS	%Δq_{jt}	%ΔDC_{jt}
US vs. FR,DE,CA	CA	0.00	-0.00	-5.6	0.01	0.00	-17.3	0.28	0.09	-88.2
	DE	0.00	-0.00	-16.2	0.03	0.00	-37.1	0.51	0.16	-93.0
	FR	0.00	-0.00	-9.23	0.06	0.00	-40.6	0.88	0.15	-94.6
	US	0.00	0.00	1.57	0.2	0.00	6.5	0.08	0.03	75.0
	%ΔFS		0.00			1.39			2.30	
US, CA vs. FR, DE	CA	-0.00	0.00	2.16	0.07	0.00	-14.9	0.33	0.08	29.7
	DE	0.00	0.00	-3.47	0.03	0.00	-28.5	0.33	0.10	-93.0
	FR	0.00	0.00	-1.28	0.00	0.06	-35.7	0.70	0.09	-95.5
	US	0.00	0.00	1.56	0.09	0.01	-21.9	0.28	0.09	13.3
	%ΔFS		0.00			1.31			2.65	

Baseline scenario is the non-discrimination case.
All results are in % change from the baseline.

2.16% in Canada, and the decrease in France and Germany is lowered compared to the other scenario. These results confirm the hypothesis that because of the non-privacy discrimination, some developers collect consumer data in France, Germany and Canada while they would prefer not. We now consider the case with scale economies on the update without data collection ($\Psi > 0$). Our results show that an increase in the scale economy rises consumer and firm surplus and the average product quality in all our countries. In the extreme case where $\Psi = 0$ and there is a discrimination between North America and Europe, firm surplus increases by 2.65%. The consumer surplus increases by 0.33% in Canada and Germany, 0.7% in France and 0.28% in the US. The average game quality approximately increases by 0.09% in all countries. Finally, the number of consumers disclosing data decreases by 93% in Germany and 95.5% in France. However, it increases by 29.7% in Canada, and 13.3% in the US.

7 Conclusion

In this paper, we show that competitive markets can be used to measure the privacy sensitivity of individuals. Using the substitution between products, and controlling for all potential unobserved characteristics, we show that the data collection decreases the demand for the product. Our results indicate that the privacy sensitivity is heterogeneous across markets. Consumers in France and Germany are more privacy-sensitive than in Canada and the US.

Moreover, we don't find any evidence that collecting consumer data in the US decreases the demand for the game. Such measure of the privacy sensitivity has the advantage to not be biased by the *privacy paradox*. In fact, this estimate is built on consumers' decisions but not on intentions to protect their privacy. We emphasize that this measure can be replicated on any other competitive industry where the use of data is not mandatory.

The second contribution of our paper is about privacy discrimination. Our results show that in case of non-discrimination in privacy, heterogeneous markets can lead to some inefficiency. More specifically, all markets are suffering from a decrease in consumer surplus, firm surplus and average product quality when a leading market captures a high part of profits and their consumer are non-privacy-sensitive. We also find that the non-privacy discrimination in this case decreases consumers' privacy in the smallest countries, but it increases in the leading country. Allowing firms to privacy discriminate between market or groups of markets can lead to benefits in terms of consumer surplus, firm surplus and average product quality. Moreover, privacy in markets where individuals are privacy-sensitive increases with a well-defined privacy discrimination.

Specifically to the mobile app industry, we show evidences that the European Union should regulate the industry to allow the privacy discrimination with other countries. It will necessarily increase the privacy of their consumers without ever harming the consumer surplus, firm surplus and the average product quality. Although this privacy discrimination doesn't harm surplus and product quality in the US and Canada, it decreases privacy of their consumers. One might emphasize that the decrease in privacy is not serious since the consumer surplus does not decrease. However, consumers probably underestimate the potential costs of disclosing data, such as data leakage and risks of identity theft. If such unexpected costs are high, a regulation of the privacy discrimination by the European Union can decrease the consumer surplus in the US and Canada, and could lead to divergent interests between the European Union and countries in North American.

References

Acquisti, Alessandro, Taylor, Curtis, & Wagman, Liad. 2016. The economics of privacy. *Journal of economic Literature*, **54**(2), 442–92.

Aguirregabiria, Victor, & Mira, Pedro. 2007. Sequential estimation of dynamic discrete games. *Econometrica*, **75**(1), 1–53.

Arellano, Manuel, & Bond, Stephen. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The review of economic studies*, **58**(2), 277–297.

Athey, Susan, Catalini, Christian, & Tucker, Catherine. 2017. *The digital privacy paradox: Small money, small costs, small talk*. Tech. rept. National Bureau of Economic Research.

Bajari, Patrick, Benkard, C Lanier, & Levin, Jonathan. 2007. Estimating dynamic models of imperfect competition. *Econometrica*, **75**(5), 1331–1370.

Barth, Susanne, & De Jong, Menno DT. 2017. The privacy paradox–Investigating discrepancies between expressed privacy concerns and actual online behavior–A systematic literature review. *Telematics and informatics*, **34**(7), 1038–1058.

Barth, Susanne, de Jong, Menno D.T., Junger, Marianne, Hartel, Pieter H., & Roppelt, Janina C. 2019. Putting the privacy paradox to the test: Online privacy and security behaviors among users with technical knowledge, privacy awareness, and financial resources. *Telematics and Informatics*, **41**, 55–69.

Berry, Steven, Levinsohn, James, & Pakes, Ariel. 1995. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Bleier, Alexander, & Eisenbeiss, Maik. 2015. Personalized online advertising effectiveness: The interplay of what, when, and where. *Marketing Science*, **34**(5), 669–688.

Campbell, James, Goldfarb, Avi, & Tucker, Catherine. 2015. Privacy regulation and market structure. *Journal of Economics & Management Strategy*, **24**(1), 47–73.

Comino, Stefano, Manenti, Fabio M, & Mariuzzo, Franco. 2019. Updates management in mobile applications: iTunes versus Google Play. *Journal of Economics & Management Strategy*, **28**(3), 392–419.

Farias, Vivek, Saure, Denis, & Weintraub, Gabriel Y. 2012. An approximate dynamic programming approach to solving dynamic oligopoly models. *The RAND Journal of Economics*, **43**(2), 253–282.

Goldfarb, Avi, & Tucker, Catherine. 2011. Online display advertising: Targeting and obtrusiveness. *Marketing Science*, **30**(3), 389–404.

Goldfarb, Avi, & Tucker, Catherine. 2012. Privacy and innovation. *Innovation policy and the economy*, **12**(1), 65–90.

Hann, Il-Horn, Hui, Kai-Lung, Lee, Sang-Yong Tom, & Png, Ivan PL. 2007. Overcoming online information privacy concerns: An information-processing theory approach. *Journal of Management Information Systems*, **24**(2), 13–42.

Hashmi, Aamir Rafique, & Biesebroeck, Johannes Van. 2016. The relationship between market structure and innovation in industry equilibrium: a case study of the global automobile industry. *Review of Economics and Statistics*, **98**(1), 192–208.

Hotz, V Joseph, Miller, Robert A, Sanders, Seth, & Smith, Jeffrey. 1994. A simulation estimator for dynamic models of discrete choice. *The Review of Economic Studies*, **61**(2), 265–289.

Ichihashi, Shota. 2020. Online privacy and information disclosure by consumers. *American Economic Review*, **110**(2), 569–95.

Janssen, Rebecca, Kesler, Reinhold, Kummer, Michael E, & Waldfogel, Joel. 2022. *GDPR and the lost generation of innovative apps.* Tech. rept. National Bureau of Economic Research.

Johnson, Garrett, Shriver, Scott, & Goldberg, Samuel. 2021. Privacy & market concentration: Intended & unintended consequences of the GDPR. *Available at SSRN 3477686.*

Johnson, Garrett A, Lewis, Randall A, & Reiley, David H. 2017. When less is more: Data and power in advertising experiments. *Marketing Science*, **36**(1), 43–53.

Joinson, Adam N, Reips, Ulf-Dietrich, Buchanan, Tom, & Schofield, Carina B Paine. 2010. Privacy, trust, and self-disclosure online. *Human–Computer Interaction*, **25**(1), 1–24.

Leyden, Benjamin T. 2019. *There's an app (update) for that: Understanding product updating under digitization.* Tech. rept. Working paper, Cornell University.

Manant, Matthieu, Pajak, Serge, & Soulié, Nicolas. 2019. Can social media lead to labor market discrimination? Evidence from a field experiment. *Journal of Economics & Management Strategy*, **28**(2), 225–246.

McFadden, Daniel. 1986. The choice theory approach to market research. *Marketing science*, **5**(4), 275–297.

Mikians, Jakub, Gyarmati, László, Erramilli, Vijay, & Laoutaris, Nikolaos. 2012. Detecting price and search discrimination on the internet. *Pages 79–84 of: Proceedings of the 11th ACM workshop on hot topics in networks.*

Mikians, Jakub, Gyarmati, László, Erramilli, Vijay, & Laoutaris, Nikolaos. 2013. Crowd-assisted search for price discrimination in e-commerce: First results. *Pages 1–6 of: Proceedings of the ninth ACM conference on Emerging networking experiments and technologies.*

Miller, Amalia R, & Tucker, Catherine E. 2011. Can health care information technology save babies? *Journal of Political Economy*, **119**(2), 289–324.

Pakes, Ariel, Ostrovsky, Michael, & Berry, Steven. 2007. Simple estimators for the parameters of discrete dynamic games (with entry/exit examples). *the RAND Journal of Economics*, **38**(2), 373–399.

Peukert, Christian, Bechtold, Stefan, Batikas, Michail, & Kretschmer, Tobias. 2022. Regulatory spillovers and data governance: Evidence from the GDPR. *Marketing Science*.

Preibusch, Sören, Kübler, Dorothea, & Beresford, Alastair R. 2013. Price versus privacy: an experiment into the competitive advantage of collecting less personal information. *Electronic Commerce Research*, **13**(4), 423–455.

Savage, Scott J, & Waldman, Donald M. 2014. *The Value of Online Privacy: Evidence from Smartphone Applications*. Tech. rept. Technical Report.

Sweeting, Andrew. 2010. The effects of mergers on product positioning: evidence from the music radio industry. *The RAND Journal of Economics*, **41**(2), 372–397.

Taylor, Curtis R. 2004. Consumer Privacy and the Market for Customer Information. *The RAND Journal of Economics*, **35**(4), 631–650.

Tsai, Janice Y, Egelman, Serge, Cranor, Lorrie, & Acquisti, Alessandro. 2011. The effect of online privacy information on purchasing behavior: An experimental study. *Information systems research*, **22**(2), 254–268.

Valentino-Devries, Jennifer, Singer-Vine, Jeremy, & Soltani, Ashkan. 2012. Websites vary prices, deals based on users' information. *Wall Street Journal*, **10**, 60–68.

Weintraub, Gabriel Y, Benkard, C Lanier, & Van Roy, Benjamin. 2008. Markov perfect industry dynamics with many firms. *Econometrica*, **76**(6), 1375–1411.