Impact of the EU-GDPR on Entry, Exit, and Consumer Welfare*

Evidence from the Online Market of Mobile Applications

Reinhold Kesler†  Michael E. Kummer‡  Joel Waldfogel§

Rebecca Janssen¶

March 19, 2019

Abstract

The enactment of the General Data Protection Regulation by the European Union has been the largest regulatory intervention in the recent history of data-driven (online) markets. We study the effect on the market of mobile smartphone applications. Specifically, we analyze how the enactment of the regulation has affected exit, entry, and consumer welfare in this market. To understand the short-run effects of the regulation, we pursue a difference-in-differences strategy that compares differences between apps affected by the regulation to apps that operate in unaffected markets before and after the regulation. Our first results indicate that the enactment has a drastic effect on entry and exit in the app market. In a second step, we exploit the regulatory shock to quantify the value of data for consumers and app developers, and to shed light on the welfare implications of the GDPR.

Keywords: Data, Privacy, Entry, Exit, Consumer Welfare, GDPR, Apps

JEL Classification Numbers: L17, D4, D29, D85

*Financial support by the State Government of Baden-Württemberg, Germany, through the research program *Strengthening Efficiency and Competitiveness in the European Knowledge Economies* (SEEK) is gratefully acknowledged.

†ZEW - Leibniz Centre for European Economic Research; L7,1; D-68161 Mannheim, Germany; reinhold.kesler@zew.de

‡Georgia Institute of Technology and ZEW; 221, Bobby Dodd Way, Room 222, Atlanta, GA 30332, USA, michael.kummer@econ.gatech.edu

§University of Minnesota and ZEW; 3-177 Carlson School, Minneapolis, MN 55455, USA, jwaldfog@umn.edu

¶ZEW - Leibniz Centre for European Economic Research; L7,1; D-68161 Mannheim, Germany; rebecca.janssen@zew.de
Extended Abstract

Introduction

Out of concerns for consumers’ privacy, the European Union instituted the general data protection regulation (GDPR) in May of 2018, which restricts the data that mobile app developers can collect without disclosure, giving developers a choice between disclosing intrusive measures or undertaking potentially costly updates of their apps. The GDPR may have a substantial effect on the app market, for two reasons. First, the revenue from operating apps comes mainly from two sources: advertising inside of the apps as well as the data the app developer could collect – arguably surreptitiously – from the app users. The GDPR, by mandating disclosure of intrusive data collection, has threatened the second source of app developer revenue. Second, remaining in the market requires potentially costly updates of apps. It is possible that the GDPR will therefore induce app exit and inhibit continued entry.

Whether the GDPR benefits consumers depends on how consumers view privacy. If consumers care about privacy and have only unwittingly gave away their data prior to the GDPR’s disclosure requirements, then possible negative effects of the GDPR on net app entry can be efficient. On the other hand, if the GDPR undermines entry by inhibiting practices that consumers did not object to, then the GDPR may have a negative effect on welfare.

In this paper, we attempt three broad tasks. First, using an extensive data set on the Google Play Store, we document patterns of entry and exit surrounding the announcement and imposition of the GDPR. We find substantial exit and diminished entry, particularly among apps involving substantial intrusions on privacy. Second, we explore consumers’ views about app intrusiveness. Once apps must disclose their intrusiveness, we can see whether consumers react more negatively to apps that collect more information. Third, we plan to build on the insights from the first two tasks to estimate a structural model of app demand that will allow us to characterize the evolution of the value that consumers derive from the app market over time, with the imposition of GDPR.
Background

In the context of the digital single market, the European Union has enacted the general data protection regulation to protect personal data in the age of digitisation. It stipulates particularly the strengthening of rights for every individual with regard to his or her personal data and lays out principles regarding data processing along with certain obligations and security measures that have to be met by entities in control of the data and which are supervised by authorities with the ability to impose fines. The law was passed in April 2016, and the regulation was directly binding from the 25th of May 2018 onwards, for all firms that are based in the EU or process personal data of individuals within the EU, irrespective of size. Concerns have been expressed that the burden associated with the regulation regarding the processing of personal data may slow down the innovation dynamics of new technologies. This has also implications from a European perspective, as leading countries such as the USA or China have less restrictive data protection laws. However, besides the anticipation and enactment of the EU-GDPR, the enforcement by platforms is also to be considered as is displayed in Figure 1 for the case of Google and its Play Store.

Figure 1: Chronology of Events

Notes: The figure shows a timeline of relevant events regarding data-related activities.

Data

For our analysis, we consider apps from the Google Play Store. Figure 2 describes the retrieval of apps by means of web scraping. Subsequently, we scrape the complete Play Store content page of each app to ultimately get product-level information. This has been done on a quarterly basis from October 2015 to January 2019 for almost all apps available
in the Google Play Store, resulting in detailed information about the apps. Specifically, for the purpose of the analysis for the effects of the GDPR enactment, we observe an app’s requested permissions to approximate privacy, as well as entry and updating activity to measure innovation.

Figure 2: Web Scraping of the Google Play Store

**Notes:** The figure shows the process of retrieving apps in the Google Play Store by means of web scraping. Starting with a registry of the most relevant apps, e.g., from AndroidRank, new apps are identified by the similar apps provided on the Play Store page for each app.

<table>
<thead>
<tr>
<th>Cross Section (Jan ’19)</th>
<th>Panel</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Permissions</strong></td>
<td><strong>mean</strong></td>
</tr>
<tr>
<td>7.32</td>
<td>6.00</td>
</tr>
<tr>
<td>1.37</td>
<td>1.00</td>
</tr>
<tr>
<td>5.95</td>
<td>6.00</td>
</tr>
<tr>
<td>0.27</td>
<td>0.00</td>
</tr>
<tr>
<td>224.43</td>
<td>226.00</td>
</tr>
<tr>
<td><strong>App Price (1=Yes)</strong></td>
<td><strong>mean</strong></td>
</tr>
<tr>
<td>0.06</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>In-App Product (1=Yes)</strong></td>
<td><strong>mean</strong></td>
</tr>
<tr>
<td>0.08</td>
<td>0.00</td>
</tr>
<tr>
<td>0.50</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Number of Installations</strong></td>
<td><strong>mean</strong></td>
</tr>
<tr>
<td>3.26e+05</td>
<td>30000.00</td>
</tr>
<tr>
<td><strong>Number of Ratings</strong></td>
<td><strong>mean</strong></td>
</tr>
<tr>
<td>3406.48</td>
<td>17.00</td>
</tr>
<tr>
<td>4.19</td>
<td>4.30</td>
</tr>
<tr>
<td><strong>App Description</strong></td>
<td><strong>mean</strong></td>
</tr>
<tr>
<td>1228.82</td>
<td>824.00</td>
</tr>
<tr>
<td><strong>Number of Screenshots</strong></td>
<td><strong>mean</strong></td>
</tr>
<tr>
<td>7.11</td>
<td>5.00</td>
</tr>
<tr>
<td><strong>Video (1=Yes)</strong></td>
<td><strong>mean</strong></td>
</tr>
<tr>
<td>0.11</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>USK Rating</strong></td>
<td><strong>mean</strong></td>
</tr>
<tr>
<td>4.75</td>
<td>5.00</td>
</tr>
<tr>
<td><strong>Privacy (1=Yes)</strong></td>
<td><strong>mean</strong></td>
</tr>
<tr>
<td>0.73</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Number of Similar Apps</strong></td>
<td><strong>mean</strong></td>
</tr>
<tr>
<td>30.93</td>
<td>49.00</td>
</tr>
</tbody>
</table>

**Notes:** The table shows the summary statistics for the cross section in columns 1-5, and the summary statistics for the panel in columns 6-10. The first four rows show our key measures of permissions, and rows 5-6 show the basic measures of innovation. Subsequently, we show all the control variables. * P-S stands for privacy-sensitive.

In Table 1, we show summary statistics of the main variables of interest for both the cross section from January 2019 and the complete panel. Our dataset contains more than
1.6 million observations in the cross section and more than 25 million observations in the panel dataset. More details can be found in Appendix A.1 which provides a description of the raw data and a list of all available variables.

In Figure 3, we plot how the app market evolved over time both before, during, and after the enactment of the EU-GDPR. We find substantial exit and diminished entry for the period around the enactment of the GRPR. This pattern is particularly strong among apps involving substantial intrusions on privacy.

Figure 3: Amount, Entry, and Exit of Apps Over Time

(a) Amount
(b) Entry
(c) Exit

Notes: The first figure shows the number of apps web-scraped from the Google Play Store. The second figure shows the number of apps entering the Google Play Store based on the app age from AppBrain. This is a preliminary figure, as the app age is only given for ninety percent of the observations. The third figure shows the number of apps web-scraped for the last time, though, the last period is excluded. The enactment of the GDPR is denoted by the vertical line, respectively.

Empirical Strategy

For our empirical strategy we conceptualize the EU-GDPR as an exogenous shock for data-related activities of app developers. In other words, we treat the regulation as a “political experiment.” The varying scope of the regulation can be exploited, by comparing apps that were active on the European market, or relied on European data for their business to “non-European” apps. To understand the short-run effects of the new regulation we can use a difference-in-differences approach that compares differences between apps in the newly regulated markets to apps that operate in unaffected markets before and after the regulation was enacted in the EU.

In a first approach we plan to compare apps from the EU and East Asia. Figure 4 shows the use of permissions by app developers for the two groups. It suggests very similar trends pre-treatment, but that the enactment of GDPR had affected the groups differently. The decrease in the number of privacy-intrusive permissions is higher for EU apps after the enactment of the EU-GDPR, while the number of clean permissions converge between the country groups.
Figure 4: Number of Requested Permissions

(a) Total

(b) Clean

(c) Problematic

Notes: The figure shows the number of total, clean, and problematic permissions requested by the two sets of apps in the Google Play Store. The solid line indicates East Asia apps, respectively. The enactment of the GDPR is denoted by the vertical line, respectively.

Obviously, this first visual analysis needs to be subjected to rigorous scrutiny, before it can be taken for granted. A key challenge is defining a treatment and control group. The scope of regulation requires applying great care when identifying unaffected apps that have no business with personal data within the EU. In addition it is a non-trivial challenge to capture the industry dynamics in light of the massive exit and slowdown of entry, which also makes aggregate trends hard to interpret.

Hence, extending our analysis to alternative control groups and refining our definition of the comparison groups is the goal of our future work. We can tackle the aforementioned issues by exploiting the fact that the app market can be segmented in sub markets. We can exploit the category tree and our information on the app developer’s location, and the app’s geographical scope (by using country-specific app ranks), in order to apply clustering to identify segments of markets that supply relatively specialized and/or local services such as banking or newspapers.
A Appendix

A.1 Raw Data

The raw data include the following app-specific information which we use to construct our privacy and innovation measures as well as our control variables:

Key variables:

- all permissions that apps are requesting (upon installation) and that apps require to perform certain functions (in total more than 200 permissions, including, e.g., ‘network access’, ‘read contents of USB’, ‘read contact data’, ‘read browser data’, ‘read sensitive log data’),
- additional information about these permissions (special flag by Google, considered privacy-sensitive by researchers etc.),
- app age (retrieved from AppBrain),
- information on updates: date, textual information on what is new, and version number.

Control variables:

- price (in Euro),
- existence of in-app purchases and the price range of such items in Euro,
- existence of in-app advertisements,
- the total number of installations of an app,
- number and values of quantitative ratings (from 1 to 5 stars),
- app category (e.g., Racing, Personalization, Traveling, Weather, Social, Health & Fitness, Finance, Communication etc.),
- apps’ description (length, content) and its illustration in the Play Store (video and screenshot availability),
- content rating (USK) and "may contain" warnings,
- code size (in KB),
- availability of interactive elements (e.g., ‘users interact’, ‘digital purchases’ etc.),
• is the app an editor’s choice (yes/no),

• Android version required for installation,

• presence of a privacy policy,

• contact information of the app (including website, e-mail, and address),

• the names and IDs of similar apps.

Further developer-specific information:

• the name of the developer,

• top developer status (yes/no),

• number of its apps,

• the set of its available apps.