GDPR and the Lost Generation of Innovative Apps*

Rebecca Janssen[†], Reinhold Kesler[‡], Michael Kummer[§], Joel Waldfogel[¶]

February 8, 2021

Work in Progress

Abstract

The General Data Protection Regulation (GDPR), enacted with the goal of protecting user privacy, imposed compliance costs on app developers and may have inhibited revenue generation. Using data on 4.1 million apps at the Google Play Store from 2016 to 2019, we document that GDPR induced the exit of about a third of available apps. Moreover, in the quarters following implementation, entry of new apps fell by over half. While the exiting apps had very little usage, the reduction in entry was more consequential for consumers. Because app success is unpredictable at launch, the missing apps would have been nearly as useful, on average, as those that still entered: Post-GDPR entry cohorts, less than half as large as their pre-GDPR counterparts, account for just over half as much usage as average pre-GDPR cohorts at the same ages. After documenting these descriptive facts, we estimate a structural model of demand and entry in the app market. Comparing equilibria with and without GDPR, we find that GDPR reduces consumer surplus by 32 percent and aggregate app usage by 26 percent. We conclude that, whatever the privacy benefits of GDPR, they come at substantial costs to consumers and producers.

JEL Classification: O31, L51, L82

Keywords: welfare benefit of new products, GDPR, apps, privacy

---

*Financial support by the state government of Baden-Wuerttemberg, Germany, through the research program 'Strengthening Efficiency and Competitiveness in the European Knowledge Economies' (SEEK) is gratefully acknowledged. We are grateful to participants of the seminars at ZEW, CLSBE Digital Economy Workshop 2019, EARIE 2019, RSM, CCP, CESifo Area Conference on the Economics of Digitization 2019, MaCCI/EPoS Virtual IO Seminar, CERNA MINES ParisTech, and VIDE BrownBag for their helpful comments.

[†]ZEW Mannheim; P.O. Box 103443, D–68034 Mannheim, Germany. E-Mail: rebecca.janssen@zew.de.

[‡]University of Zurich and ZEW Mannheim; Plattenstrasse 14, CH-8032 Zurich, Switzerland. E-Mail: reinhold.kesler@business.uzh.ch.

[§]University of East Anglia, Georgia Institute of Technology, and ZEW Mannheim; 2.25E Registry And Council House, Norwich NR4 7TJ, United Kingdom. E-Mail: m.kummer@uea.ac.uk.

[¶]University of Minnesota, NBER, and ZEW Mannheim; 3-177 Carlson School, Minneapolis, MN 55455, USA. E-Mail: jwaldfog@umn.edu.

1 Introduction

In an effort to better protect user privacy, the European Union (EU) enacted the General Data Protection Regulation (GDPR) in May of 2018. The regulation restricted the use of personal information, potentially reducing revenue, and required developers of mobile applications ("apps") to engage in potentially costly compliance activities. This raised the possibility that GDPR would cause non-compliant products to exit – and would curb further product entry into – the app market. While the protection of privacy was of course the direct intent of GDPR, the new law could also bring about an unintended consequence: A reduction in the volume of app entry could hamper innovation and undermine the availability of new and potentially valuable apps to consumers, particularly if the quality of apps – like many digital products – were unpredictable at the time of entry.

In many markets, it is difficult to predict which new products will succeed; and unpredictability of new product success can have important consequences for the welfare benefits of entry. When success is unpredictable, an increase in the number of new products, even those with modest ex ante commercial prospects, can deliver products with substantial realized value.[1] For the most part, digitization has delivered reductions in entry costs, inducing substantial additional entry in a variety of media product categories. GDPR may be like the reverse of digitization. By raising developers' costs and reducing their revenue, the regulation may have induced exit and may have prevented the entry of a "lost generation" of valuable apps. This leads us to ask how GDPR has affected the welfare of consumers and producers in the app market.

We use the Google Play Store selling apps as our study context. Our data consist of 4.1 million apps available at the Google Play Store between July 2016 and October 2019, along with measures of their usage based on the volume of user ratings. We ask four descriptive questions, then present structural estimates. First, we document the impact of GDPR on app exit, the flow of new app entry, and the resulting number of apps available. Second, we explore what happened to the privacy-invasiveness of apps. Third, we turn to evidence of the welfare impacts of GDPR, asking whether lost apps would

[1] See Arrow (1969); Bergemann and Hege (2005); Kerr et al. (2014); Manso (2011, 2016); Weitzman (1979). Aguiar and Waldfogel (2018) measure the welfare benefit from increased product entry into recorded music.

have been valuable to consumers. In particular, we ask whether smaller post-GDPR app birth cohorts account for less aggregate usage at each age of their lives than their pre-GDPR counterparts. Fourth, we look for evidence of higher app development costs from changed realized usage, per app, after GDPR's implementation. We then turn to structural welfare estimation. First, we estimate the welfare loss to consumers from the lost generation of apps using a nested logit model of demand. Second, we use the demand model, along with an entry model with ex ante unpredictability of product quality (as in Aguiar and Waldfogel (2018)) to develop estimates of the losses to producers.

We have six broad findings. First, GDPR sharply curtailed the number of available apps, via two mechanisms. When it took effect, GDPR precipitated the exit of over a third of available apps; and following its enactment, the rate of new entry fell by 56 percent, in effect creating a lost generation of apps. Second, apps became less intrusive after GDPR, but the decline in intrusiveness appears to be the continuation of a pre-existing trend. Third, the GDPR-induced baby-bust app vintages account for 41 percent less usage over time than earlier vintages. That a 56 percent reduction in app entry reduced cohort app usage by 41 percent indicates that app quality is highly – albeit not entirely – unpredictable ex ante and, moreover, suggests a negative welfare impact of GDPR. Fourth, average usage per app rose for the vintages launched after the imposition of GDPR, consistent with GDPR raising app development costs. Fifth, using a nested logit demand model, we estimate a long-run 32 percent reduction in consumer surplus from reduced entry. Sixth, using the estimated model, we simulate that GDPR substantially raised development costs for the marginal app and reduced aggregate app usage by 26 percent, and – if revenue per user remained constant in the counterfactual post-GDPR equilibrium – would reduce revenue by 25 percent as well. Whatever the benefits of GDPR's privacy protection, it appears to have been accompanied by substantial costs to consumers, from a diminished choice set, and to producers from depressed revenue and increased costs.

The paper proceeds in six sections after the introduction. Section 2 describes the major provisions of the GDPR, explains how GDPR would be expected to raise costs and reduce revenue, and presents links to relevant literature. Section 3 introduces a theoretical framework describing app entry, exit, and welfare, to guide our measurement exercises. Section 4 describes the data used in the study. Section 5 presents our empirical strategy and descriptive results on exit, entry, usage, and app intrusiveness before and after GDPR. Section 6 presents structural welfare estimates, and Section 7 concludes.

2 Background

2.1 The app market

There are two large distinct mobile app platforms for Apple and Google mobile operating systems, respectively. Apps generate revenue from a combination of user charges (e.g., download prices and in-app purchases) as well as in-app advertising; and the collective market is large. Revenue from the user side to both platforms together grew from $43.6 billion in 2016 to $83.6 billion in 2019, with roughly two thirds of it generated for Apple devices.[2] Aggregate mobile advertisement revenue (across both platforms) grew from $80.7 billion in 2016 to $189.2 billion in 2019.[3] Applying Google's share of user-based revenue to all revenue, the Google-relevant revenue rose from $42.8 billion in 2016 to $95.9 billion in 2019. Over the 2016-19 period, just over two thirds of total app revenue was from ads.

The number of Android users, and therefore potential users of Play Store apps, reached 1 billion in June 2014, 2 billion in May 2017, and 2.5 billion in May 2019.[4]

2.2 GDPR

The EU enacted the GDPR in an effort to protect the personal data of European citizens and harmonize privacy laws across member states. The regulation strengthened users' privacy rights and obliged app developers to take security measures. Under GDPR, app developers must guarantee users their rights of access, rectification, erasure, restriction of processing, data portability, and the right to object; and developers are obliged to protect user data "by design and default."[5] Compliance with these provisions could raise operation costs at both the app and developer levels. The law applies to all firms

[2] See https://sensortower.com/blog/app-revenue-and-downloads-2017 and https://sensortower.com/blog/app-revenue-and-downloads-2019.

[3] See http://www.statista.com/statistics/303817/mobile-internet-advertising-revenue-worldwide/.

[4] See https://www.zdnet.com/article/google-io-android-stands-at-one-billion-active-users-and-counting/, https://www.theverge.com/2017/5/17/15654454/android-reaches-2-billion-monthly-active-users, and https://www.theverge.com/2019/5/7/18528297/google-io-2019-android-devices-play-store-total-number-statistic-keynote.

[5] See the regulation: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679.

processing personal data of EU residents regardless of the firm's headquarter location.[6] GDPR gives the EU powers to investigate, and in the case of violations, to impose fines of up to the larger of 4 percent of annual revenue or 20 million EUR.

Timing: The law was passed on April 27th, 2016 and went into effect on May 25th, 2018. We have three measures showing that awareness of the law grew during the period between passage and the enactment.[7] First, the volume of Google searches on "GDPR," in Figure 1, shows that interest rose slightly from 2016 onwards, then jumped substantially in the quarter it took effect. Second, app developers' online expressions of concern about GDPR – the volume of comments including the term "GDPR" at, for example, Stack Overflow (Android tag) and Reddit (r/androiddev) – jumped similarly. Third, the volume of editing on Wikipedia for the English-speaking article on GDPR moved similarly over time. Each measure of GDPR interest peaks in the quarter GDPR took effect. It is clear that developers were aware of GDPR's arrival.

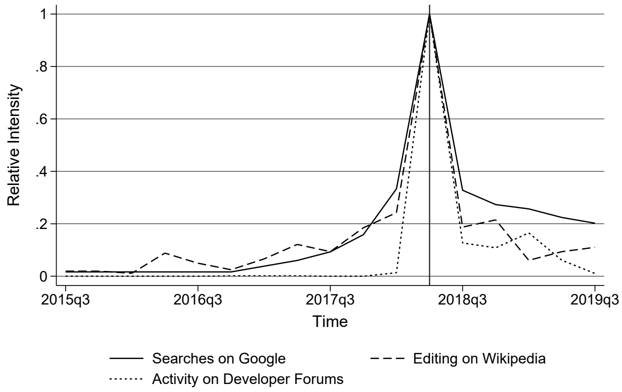


Figure 1: Interest in GDPR on different platforms

Anticipated effects on app developers' costs and revenues: The new regulation, by its nature, could be expected to raise developers' costs and to reduce their revenue. We explored the effects felt by practitioners with a survey of 650 German app developers in October 2019.[8] Asked about challenges of compliance with GDPR, 85 percent listed "administrative burdens," 48 percent noted "additional costs," and 36 percent indicated

[6] This extraterritorial scope of the regulation makes it difficult to find "untreated" apps, a point we return to below.

[7] See Appendix A.1 for more details.

[8] See Appendix A.2 for more details, in particular figures A.1 through A.4.

a "lack of knowledge about the regulation's details." In particular, developers mentioned costs for data protection officers and legal advice, and many reported spending a substantial amount of time on implementing GDPR compliance.[9] One in seven of the developers reported having removed an app from the market due to new requirements and costs, and one in eleven reported choosing not to launch a developed app.[10]

The second major provision of GDPR affects how developers can use the data collected from users. Under GDPR, developers must obtain user consent to continue processing user data. These new rules may restrict developers' deployment of targeted advertisement and may reduce expected revenue (Böhmecke-Schwafert and Niebel, 2018; Goldberg et al., 2020). Several developers in our survey, particularly those generating revenue in data-intensive ways, experienced reduced revenue under GDPR. For example, 38 percent of developers using ads for revenue generation, and almost all apps selling data to third parties, reported a decline in revenue with their post-GDPR monetization strategies.[11]

2.3 Related literature

Our study is related to four strands of literature. First, our study is part of a literature on the welfare benefit of new products when success is unpredictable. While canonical contributions ask how particular products, for example, minivans (Petrin, 2002) or new telecommunications services (Hausman, 1997) raise the value of the choice set to consumers, an alternative approach is to incorporate the stochastic nature of new product quality, asking how changes in the tendency for products to enter affects welfare. This is the approach of Aguiar and Waldfogel (2018), who estimate the welfare benefit of digitization's reduced entry costs in the recorded music market. Valuing entry with unpredictable product quality also echoes the approach of a literature treating entrepreneurship as "experimentation" (Arrow, 1969; Bergemann and Hege, 2005; Ewens et al., 2018; Kerr et al., 2014; Manso, 2011, 2016; Weitzman, 1979).

Second, our study is part of a growing literature examining the impact of GDPR on various outcomes, including the concentration of the market for web technology services

[9] Survey respondents report needing to update privacy policy information for every app not already in compliance as well as Google Analytics settings. Moreover, developers needed to designate a data protection officer, entrusted with guaranteeing users' rights and the security of the data.

[10] One of our survey respondents wrote 'Removed several small apps completely in order to minimize the risk and because of the uncertain as well as non-transparent legal situation.'

[11] Some studies find an increase in the value of remaining customers that offsets the decrease in the number of customers (Aridor et al., 2020; Godinho de Matos and Adjerid, 2020).

(Batikas et al., 2020; Johnson et al., 2020), firms' ability to collect data (Aridor et al., 2020; Godinho de Matos and Adjerid, 2020), the profits of e-commerce firms (Goldberg et al., 2020), interconnection agreements between networks on the Internet (Zhuo et al., 2019), venture investments (Jia et al., 2019), and the ability of web publishers to continue financing content creation (Lefrere et al., 2020). We also document the effects of GDPR on various outcomes, including entry, exit, and product usage in the app market, along with impacts on consequent welfare.

Third, our paper relates to an extensive literature on consumer demand for online privacy and the effects of privacy regulation on service providers in digital markets (Acquisti et al., 2016; Goldfarb and Tucker, 2011; Tucker, 2012, 2014). Momen et al. (2019) document that the enactment of GDPR resulted in moderately reduced usage of privacy-sensitive permissions by app developers, and Sørensen and Kosta (2019) document that fewer third-party libraries were present among affected websites. Batikas et al. (2020) and Johnson et al. (2020) document the same impact, although they find the reduction to be temporary.

Fourth, our paper contributes to the literature about the app market per se. Kummer and Schulte (2019) document the role of app data collection in revenue generation. Leyden (2019) examines how platform choices affect incremental innovation in apps, and Ershov (2020) documents that a rearrangement of the app store that facilitated consumer search also promoted app entry. Other relevant studies of the app market include Carare (2012), who measures the impact of bestseller ranks on sales and Ghose and Han (2014), who estimate welfare effects of apps and the influence of advertising and in-app purchases on demand.

3 Theoretical Framework

Based on the foregoing discussion of GDPR, we expect GDPR to have two distinct but related effects. First, it will raise the cost of operating existing apps, whose developers may or may not be in compliance with GDPR standards for maintaining privacy. Second, it will raise development costs and reduce the revenue available from launching new apps, in relation to costs. Both mechanisms will eliminate low-value apps – those with few users and presumably generating little revenue – but the mechanisms affect entry and exit differently. Moreover, they will have larger effects on the value of the app choice set, the less predictable app success is at the time of development.

3.1 GDPR, entry, and exit

A developer contemplating the creation of an app forms an estimate of the revenue (r) the app will generate. That estimate is true revenue (ρ), plus a random error: $r = \rho + \epsilon$. App development has a fixed (sunk) cost of development which, prior to GDPR, is given by C_0. An app enters if its expected revenue is bigger than development costs C_0. Equilibrium arises when there are no profitable opportunities for entry.

When GDPR goes into effect, C_0 rises to C_1, and revenue falls by Δ, so expected revenue falls from r to $r' = r - \Delta$. These features of the economic environment affect entry and exit differently. First, consider entry. Once a developer knows that GDPR will go into effect, the developer knows that an app's prospective revenue less cost is smaller than it would have been, absent GDPR. Some potential products that seemed promising when GDPR was not on the horizon no longer have expected revenue in excess of costs, so entry falls. Because apps generate revenue over time after installation, the announcement of the new law can depress entry even prior to the law going into effect. In short, facing higher entry costs and lower revenue per user, developers will only launch apps with higher expected usage than before GDPR.

Effects on exit are different. Note that the incremental cost of development under GDPR is $\Delta C = C_1 - C_0$. Apps already in existence prior to GDPR are earning (realized) ρ per period. Absent GDPR, already-existing apps continue operating until they are obsolete. When GDPR goes into effect, the developer now compares the new stream of realized revenue $\rho - \Delta$ against the cost of bringing the app into compliance, which is positive even if less than the full cost of new development. Unlike for entry, where revenue must be predicted, the realized revenue of existing apps is known. Developers compare the realized revenue of existing apps against the cost of coming to compliance. Hence, low-value apps exit. There is no reason for exit to occur prior to the law taking effect, but when the law goes into effect, we expect exit of low-value apps.

As before, equilibrium means no opportunities for profitable entry. Because of higher costs and lower revenues, the equilibrium under GDPR has fewer apps available and higher development costs – and usage – per app. Under GDPR the marginal entering app could need more users, for two reasons. First, development costs are higher. For a given level of revenue per user, an app needs more users to generate sufficient revenue to cover costs. Second, GDPR may reduce the revenue available per user, making it necessary for a marginal entering app to have more users to cover its costs.

3.2 GDPR and welfare

The welfare generated by the app market consists of the consumer surplus enjoyed by app users, plus the profits of developers (their revenues less the costs they incur to create and operate apps). GDPR may also improve privacy in ways that are socially valuable but which consumers do not appreciate (and therefore escape our quantification through consumer surplus). Consequently, our welfare analysis provides a measure of the cost of GDPR to consumers and producers, which policy makers can balance against other benefits. Here, we discuss effects on consumer surplus and on producers.

Reduced consumer welfare from a diminished choice set: Because of the unpredictability of app quality at entry, GDPR's depressing effect on the number of apps available – and, in particular, its depressing effect on entry – can have a substantial effect on the value of the app choice set to consumers. The apps that exit upon GDPR implementation are those with low realized value, so it is not worth the cost of bringing them into compliance. Because these apps are generating little consumer usage, their exit may have little effect on welfare.

The effects operating through depressed entry are potentially quite different. Developers compare expected revenues to costs and enter only if their (potentially lower) expected revenues exceed the new cost threshold. After GDPR, fewer apps have expected revenue in excess of the new cost threshold, which leads to less entry. The apps that would have entered previously but do not enter when facing GDPR's higher costs, make up the "lost generation" of apps.

The missing apps have low *expected* revenue, but if success is not entirely predictable, many of the lost apps would have been valuable to consumers. Extreme examples make this clear. If app success were completely unpredictable, then a shock that reduced entry by x percent would reduce the number of valuable apps by x percent as well. Indeed, if x is the percentage reduction in entry across birth cohorts and y is the percentage reduction in the cohort share of total usage, then the test for complete unpredictability is whether $x = y$. If so, then "nobody knows anything;" and the lost generation of apps would have been as useful, on average as those that entered. If success were entirely predictable, by contrast, then the decline in usage y would be much smaller than the decline in entry x.

If the former characterization is accurate (i.e. if app success is unpredictable at the time of entry), then the birth cohorts that are contracted because of GDPR will account for less subsequent usage. In particular, the amount of usage accounted for by the smaller birth cohorts will be smaller than it would have been if the birth cohorts had been of their traditional size.

Even if GDPR prevents the entry of some apps that would have been widely used, that is not sufficient to demonstrate an effect on consumer welfare. The app environment includes thousands of products. If the missing apps have close substitutes among the apps that continue to be made available, then consumers may not be substantially harmed by the reduction. Consequently, our welfare measurement framework needs to incorporate possible substitutability across apps.

Welfare costs borne by producers: The second possible welfare cost of GDPR arises from the additional costs – and reduced profits – incurred by developers. Figure 2 provides a simple analysis. The downward-sloping curve shows the revenue per app, ordered from those expected to be most revenue-generating to the least, while the horizontal line labelled C_0 shows the cost of developing and operating an additional app.[12] The downward slope of the upper curve in Figure 2 arises from differences across apps in their expected numbers of users. The vertical axis shows money revenues and costs, while the horizontal axis shows the number of apps in the market. In the pre-GDPR equilibrium, developers operate N_0 apps and earn profits of $A + B + C$, while app development costs are $D + E$. When GDPR takes effect, the cost of operating apps rises to C_1. The number of apps the market can now sustain falls to N_1. Producer surplus falls by $B + C$, to A.

Two alternatives to the basic setup merit discussion. First, if app success is completely unpredictable, then all apps have the same expected revenue. There would still be a downward-sloping revenue curve due to the relationship between total revenue and the number of entering apps N. But for any level of entry, all apps have the same expected

[12] Anticipating our empirical implementation, we view the "price" (revenue per user) as fixed, as if app developers were price takers in the ad market.

revenue. Under this interpretation, profits would be zero with or without GDPR. For example, with $N = N_0$, revenue equals $D + E$, as do aggregate costs. Then, PS is zero with or without GDPR, and the only welfare effect of GDPR is felt by consumers.

Second, even with app quality predictability at entry, it is possible that profits are as low as zero if app development costs are higher for inframarginal than for marginal entrants. The region $A + B + C$ is an upper bound on profit, based on the idea that development costs are C_0 per app for all apps. That entry continues to N_0 reflects that the expected revenue of the marginal entrant equals C_0. The expected revenue of inframarginal entrants exceeds C_0. Thus, we can infer that inframarginal apps have costs between their expected revenue and C_0. Hence, with $N = N_0$, profits lie between $A + B + C$ and zero.

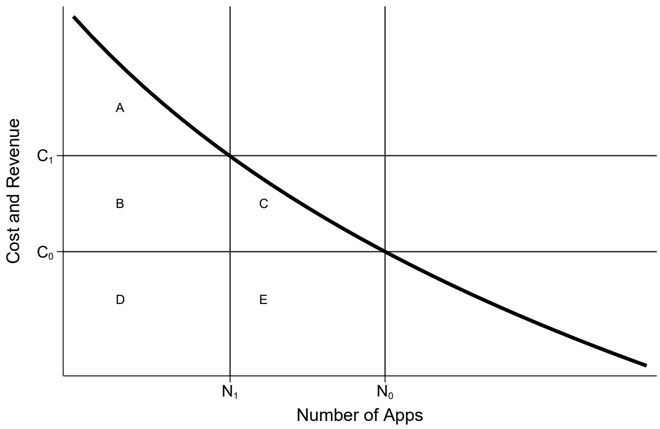


Figure 2: GDPR and producer surplus

Our empirical work below has two parts, following our data description in section 4. First, we document impacts on entry, exit, the app choice set available to consumers, usage of the diminished entry cohorts, and average usage per app (cf. section 5). We then turn to structural analysis aimed at quantifying impacts of GDPR on welfare. We use a structural model of demand that allows for substitutability among apps to quantify the long-run impact of GDPR-depressed entry on consumer surplus (cf. section 6.1). The model also allows us to estimate effects on developer revenues and costs.

4 Data

4.1 Data Collection and Preparation

To measure GDPR's effect on the app market, we need to observe product exit, entry, and usage. We obtain these measures, along with app characteristics, from the Google Play Store. Once we find apps in the Play Store data, we also obtain their entry dates from AppBrain and Google.

Because there is no available catalog of all apps, we created a list – which serves as the backbone of the database – using the following iterative process. We started with a substantial but potentially incomplete initial list of apps from AndroidPIT.[13] To assemble our full list of apps, we queried the Play Store for apps on our list. Each query returned a list of "similar apps;" and we added the suggested apps not already in our data to our list. We repeated this process quarterly between October 2015 and July 2016, until the list stabilized, i.e. the only apps suggested by Google not already in our list were new products. At that point, the number of apps on our list – and therefore in our sample (2.1 million in July 2016) – was within ten percent of the 2.2 million reported by AppBrain.[14]

Using our eventual list, we queried the Play Store quarterly between July 2016 and October 2019 for each app. In each quarter, we added newly-appearing suggestions – new entrants – to our data. As a result, we have quarterly data on each app's availability and cumulative usage, as well as privacy features, price, and for a subsample also the country of origin for a developer.[15] For each app found, we know the app's birth quarter.

Even with our comprehensive app list, the data collection process gives rise to two kinds of missing data. First, once we begin observing an app, we sometimes miss data collection on that app, usually for just a quarter. This occurs in 12.6 percent of quarterly observations. Second, because new apps enter our dataset via Google's related app suggestions, we do not always observe an app in its birth quarter. We first observe 44 percent of the apps in their birth quarters and another 26 percent in their second quarters. All told, we observe 89 percent in their first four quarters of life. Once we observe an app, we can fill in the missing data by imputation: We linearly interpolate between observed

[13]See https://web.archive.org/web/20130819094306/http:/www.androidpit.de/de/android-market/paid-android-apps-BOOKS_AND_REFERENCE.

[14]See http://web.archive.org/web/20160625194024/https://www.appbrain.com/stats/number-of-android-apps.

[15]See Appendix A.3 for more details.

cumulative usage measures, treating cumulative usage at entry as zero.

The delay in first observing apps creates another, more consequential problem. We can only include an app in the sample once we observe it. We are interested in the volume of entry over time, and the sample ends five quarters after the GDPR goes into effect. Unless we account for the problem of delayed first observation, we risk mistaking delayed observation for reduced entry. We deal with this by comparing volumes of entry first seen in, say, their second quarter of life; we discuss this further in section 4.2.

Main Variables: We obtain each app's date of entry into the Play Store from AppBrain or Google.[16] We measure the date of an app's exit as the quarter of its last appearance in our data. We have two variables reflecting the cumulative usage of an app. The first is a categorical measure of cumulative installations.[17] Cumulative installations change over time, but because of the width of the categories, the measure changes little quarter to quarter. Moreover, the categoric measure cannot easily be translated into a continuous measure of quarterly usage. The second variable is a continuous measure of the number of times each app has been rated by a user. The measures are highly correlated: Appendix figure A.5 shows the strong and monotonic relationship between the log of cumulative ratings for each app (along with 90th and 10th percentiles) and the cumulative installation category. The strong relationship between the cumulative numbers of ratings and installations provides evidence that the continuous ratings-based measure is informative about app usage. Hence, the main usage measure we employ is the quarter-to-quarter change in the cumulative number of ratings that users have left for each app, which we impute when missing.

We document the evolution of privacy measures surrounding the enactment of GDPR using the presence of a privacy policy and the number of privacy-sensitive permissions requested by an app upon the installation.[18] We also observe each app's price (usually zero), and the country of origin for a large fraction (40.1 percent) of developers.

[16]We obtain entry dates for nearly 3.8 million apps from an app's page on AppBrain. We obtained an additional 300,000 entry dates from metadata at the Play Store page. If we look at the first appearance of apps with missing age, they are scattered across the whole observation period rather than towards the end.

[17]The categorical measure has the following bins: 5, 10, 50, 100, 500, 1,000, 5,000, 10,000, 50,000, 100,000, 500,000, 1 million (m), 5m, 10m, 50m, 100m, 500m, 1 billion (b), 5b, and 10b.

[18]We follow Kummer and Schulte (2019), and define privacy-sensitive permissions based on their potential to undermine a user's privacy (e.g., phone identity, location, contacts, or messages).

Table 1 summarizes the main variables. The sample contains 4,098,275 apps and 31.4 million app-wave observations. Of these, 12.6 percent contain imputed usage measures. The average usage – based on the change in the number of ratings – is 140. 43.8 percent of observations have positive usage measures, and 7 percent of observations have positive download prices.

Table 1: Summary Statistics

	App-Wave Mean	SD	N
Delta # Ratings	140.03	9406.91	25363602
# P-S Permissions	1.36	1.82	31422427
P-S Permissions Dummy	0.52	0.50	31422427
Privacy Policy Dummy	0.46	0.50	31423119
Paid App Dummy	0.07	0.26	31420206
Age in Quarters	8.40	6.68	30263626
EU Dummy	0.39	0.49	8432088

Notes: This table shows the main variables that we use in the analysis at the app-wave level it. We observe 4,098,275 apps over the entire period. The sample period is 14 quarters, but not all apps are observed every period, so we have 31,422,427 quarterly observations in total. The number of ratings is missing if the app has none or when the measure cannot be imputed as average between two periods because the app is not observed in a second period.

Available Apps as a Starting Point: Before turning to nuances of the delayed observation problem, we document the evolution of the entire market. Figure 3 shows the number of distinct apps available over time. The figure has a vertical line at the quarter just before GDPR took effect; and the pattern is striking, even bearing in mind the possibility that the last few quarters are depressed by delayed observation. After rising from 2.1 million to 2.8 million between the second quarter of 2016 and the fourth quarter of 2017, the number of available apps fell by almost one million – about 32 percent – by the end of 2018.

While Figure 3 shows a clear effect of GDPR, it leaves a number of questions unanswered. First, the drop in available apps is the net result of changed exit and changed entry, which the raw total obscures. Second, the total number of available apps understates the number of recently-entering apps toward the end of the sample. Third, while a drop in the number of available product choices is suggestive of a harm to consumers, drawing such a conclusion requires a few additional steps, including documenting that the missing apps would have been widely used and that their absence would leave consumers worse off in light of potential substitutability. In the remainder of the paper, we explore

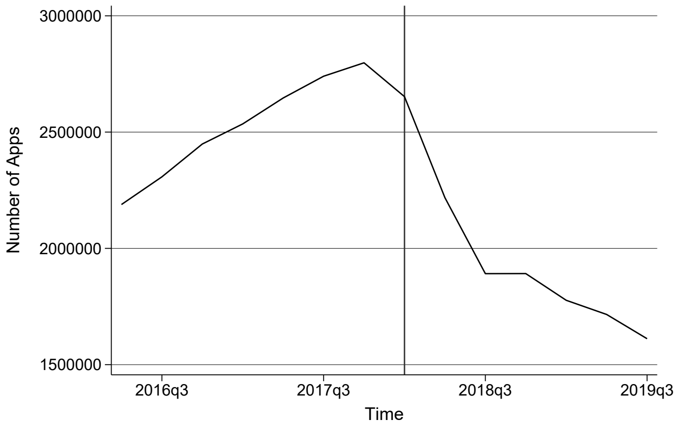


Figure 3: Apps Available and GDPR

these concerns more systematically, beginning in this section with the development of measures of entry and usage that deal with the delayed observation problem.

4.2 Measuring entry and usage patterns

Using vintage data to measure entry patterns: The problem that the delay in first observing apps creates for observing the timing of entry can be addressed using data on the vintage of each app. Because it takes time for apps to be observed in the data, we can make a "fair" comparison of entry volumes across birth quarters by comparing apps first observed at a particular age, for example when the apps are in their first quarter of life. Figure 4 provides an example of this, showing the apps born and observed in the same quarter, over time. This measure averages roughly 100,000 per quarter until the second quarter of 2017. Thereafter, the measure falls to about 50,000 per quarter for the following periods concluded by another drop to about 15,000 on average for the first three quarters in 2019. This is, of course, strong evidence that app entry falls over time with – and even before – the implementation of GDPR.

Some notation facilitates the discussion. Define N_{tv} as the number of apps entering in period v (their "vintage") and first observed in period t. Figure 4 shows the relationship between the number of apps first observed at age 0, or $N_{tv|t-v=0}$ and v. We can track the evolution of entry across vintages from $N_{tv|t-v=k}$ for each age k. If the sample ends at period T, then the series $N_{tv|t-v=k}$ is only available until period $T-k$. Below we apply this idea – of comparing birth cohorts at particular ages – to create an entry index that we can use for measuring the impact of GDPR.

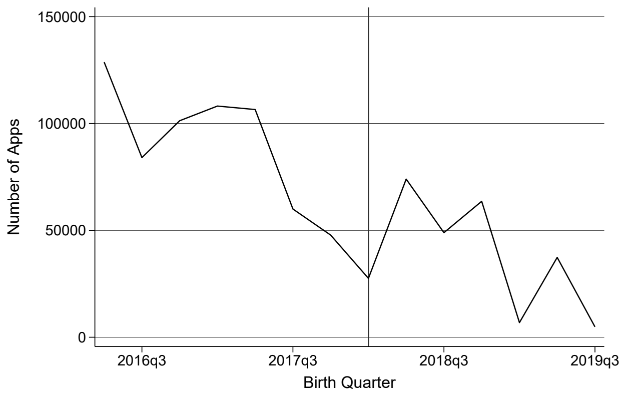


Figure 4: Apps Born this Quarter

Usage of entering app cohorts over time: If we define q_{it} as our raw usage measure (the number of new ratings app i received in quarter t), then we calculate the usage relative to market size as $s_{it} = q_{it}/M_t$, where M_t is the number of Android users, or "market size" in quarter t. The market size is drawn from the number of Android users announced by Google, and we linearly interpolate between these announcement dates (cf. section 2.1).

We would like to measure the "usefulness" of entering app cohorts and whether it changes across cohorts as the number of entrants changes. One simple approach would be to look at the distribution of usage across vintages at a point in time; but such a comparison would be confounded by age: 4-year-old apps may be used less than 3-year-old apps in 2020 because of depreciation even if the 2016 vintage was more useful. However, we can compare the usefulness of 2017 and 2016-vintage apps by comparing their usage when they were the same age. We can add the usage of all apps i born in each vintage v: $s_{tv} = \Sigma_{i \in v} s_{it}$. Then a simple way to measure the relative usage of different birth cohorts is to ask how the usage share for apps born in, say, the previous quarter evolves over time ($s_{t,v|t-v=1}$). Figure 5 illustrates this idea with our usage measures for ages 1 through 3 against the birth quarter. These measures appear to decline after GDPR, indicating that vintages become less useful. As with the entry measures, we will combine usage data from apps of all ages into indices for measuring the impact of GDPR on vintage usefulness.

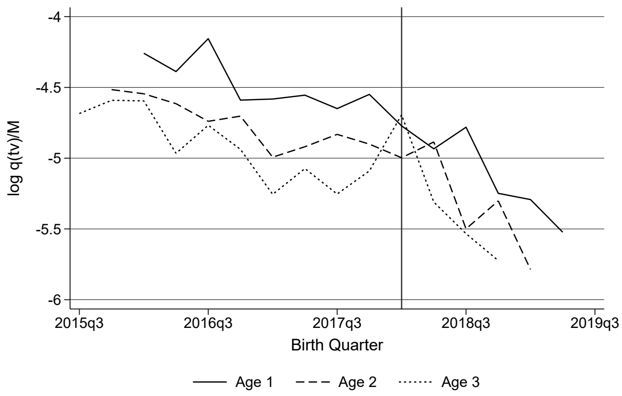


Figure 5: Usage of Age 1-3 Apps by Birth Quarter

5 Empirical Strategy and Results

5.1 Setup

Ideally, we would document the effects of GDPR on various outcomes using a research design with treated and untreated regions of the world. That is, we would compare the market for, say, EU-based apps targeted only to EU consumers with, say, Asian apps targeted exclusively to Asian consumers. We will provide some comparisons along these lines below, but it is worth pointing out at the outset that, as numerous other event studies of GDPR find (Batikas et al., 2020; Johnson et al., 2020), the world lacks untreated regions. GDPR seems to have had substantial extra-territorial effects, including – surprisingly – in places where neither the developer nor the users are protected by GDPR.[19] Lacking a control group, we will appeal to other evidence that the sharp changes in entry and exit around GDPR are its effects.

5.2 Exit

As Figure 6 shows, app exit – which had averaged about 100,000 per quarter up to the third quarter of 2017 – rose sharply to 600,000 apps last observed in the first quarter of 2018 just before GDPR took effect in May 2018. In the year surrounding the arrival of GDPR, 1.4 million apps exited, roughly 1.1 million over the baseline rate of app exit.[20]

[19]While our main dataset reflects apps available at the Play Store accessed by Germans – whom GDPR protects – we verified that apps exiting from the Play Store available to Germans also exited the stores targeting countries outside the EU (see Appendix A.3 for more details).

[20]AppBrain also documents a net exit of more than 1.2 million apps around the GDPR enactment: https://web.archive.org/web/20190117122626/https://www.appbrain.com/stats/number-of-android-apps. We obtain very similar patterns if we infer exit from three, four, or five quarters of absence from the dataset, rather than the remainder of the sample period.

The timing of the apparent impact is consistent with our theoretical prediction that GDPR would bring about exit at the imposition rather than before.

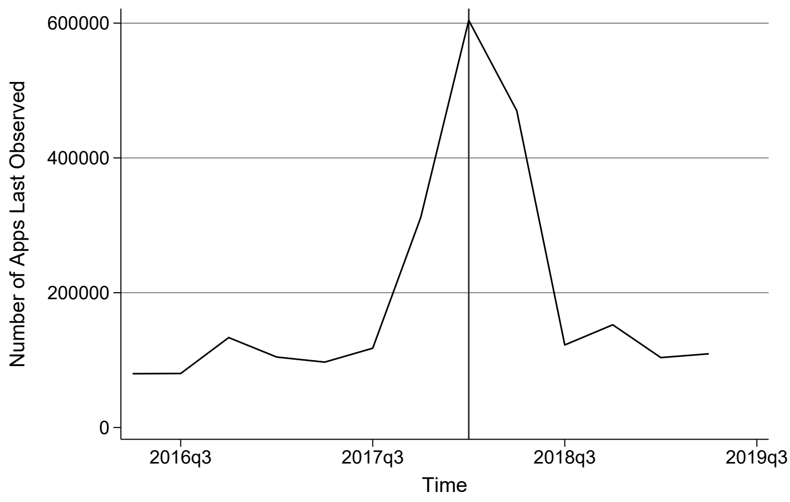


Figure 6: App Exits

The sharp spike in exit in Figure 6 as GDPR takes effect is *prima facie* evidence of a causal impact. Standard practice, however, is to document effects relative to patterns in untreated areas. To attempt this, we examine exit patterns for the subsample of apps whose developers have known locations and are located in the EU, and we compare this to exit patterns for apps whose developers are in six non-EU countries culturally or linguistically distinct from the EU: Israel, India, Japan, Korea, Russia, and Taiwan. While 42.1 percent of EU-developed apps exit in the year following GDPR, the analogous figure averages between 37.7 and 50 percent in the other six countries, confirming the difficulty in finding an untreated part of the world.

Beyond its timing, there is one other clue that the exit spike is GDPR-induced: Apps requesting privacy-sensitive information exit sooner. Of the apps operating in the last pre-GDPR quarter, 29.7 percent of those requesting at least one privacy-sensitive permission exited within one quarter, compared to 15.6 percent of those requesting none.

The absence of distinct "treatment" and "control" contexts makes it important that we attribute to GDPR patterns arising from other causes. We are aware that GDPR is not the only potential influence on the number of apps available. Google itself has also instituted policies that police apps for potential privacy violations. Yet, these policy changes occurred either substantially before or long after GDPR took effect and cannot explain the exit spike in Figure 6.[21]

5.3 Entry

Because of our inability to find an untreated part of the world, and therefore a control group, our basic approach to determine whether GDPR depressed entry is to ask whether entry fell following GDPR. For this, we use N_{tv} as the number of apps which entered in period v (their vintage), and which are *first* observed to be available in period t. We regress $ln(N_{tv})$ on age and vintage dummies to isolate vintage effects, as in Waldfogel (2012).

$$ln(N_{tv}) = \mu_{t-v} + \eta_v + \epsilon_{tv}. \quad (1)$$

The terms μ_{t-v} are age effects, and they vary in accordance with the amount of time between birth and when apps are first observed. The terms η_v show the volume of entry by vintage, controlling for age. We use logarithms for the dependent variable because at different ages, apps account for different absolute numbers, while the information we seek to extract is contained in the proportionate differences in N_{tv} across vintages.

Figure 7 shows the resulting vintage coefficients, normalizing the last pre-GDPR quarter to zero. Prior to the law taking effect, the η_v terms are nearly zero; after GDPR takes effect, the entry index falls. We measure the potentially depressing effect of GDPR with a regression of $ln(N_{tv})$ on age dummies and a post-GDPR indicator, in Table 2, whose coefficient shows how the tendency to enter in the post-GDPR period compares with the average prior to the change. The post-GDPR coefficient is -0.8170 (se = 0.250), indicating that GDPR-depressed entry by 56 percent, on average.[22]

We also create a time series showing the implied absolute level of app entry by multiplying the number of entering apps in the first sample wave by the exponentiated change

[21] In early 2017, Google announced that it would penalize apps without valid privacy policies by "limiting their visibility." Developers were given until March 15, 2017 – a year before GDPR took effect – to "link to a valid privacy policy" or to simply remove privacy-sensitive permissions requests. Developers ignoring the warning were "at risk of being hidden from view in the app store or removed altogether" (Osborne, 2017). A Google Play project manager reported that Google had taken down "more than 700,000 apps that violated the Google Play policies" during 2017 and about 200,000 apps during 2016 (Ahn, 2018). In late 2019, Google announced a tightening of the review process for apps that would cause developers to wait "up to 7 days or longer" for app approval (Siddiqui, 2019). While Google's own actions to protect privacy are potentially important, the timing of these actions does not align with the timing of GDPR.

[22] Based on $1 - e^{-0.8170}$.

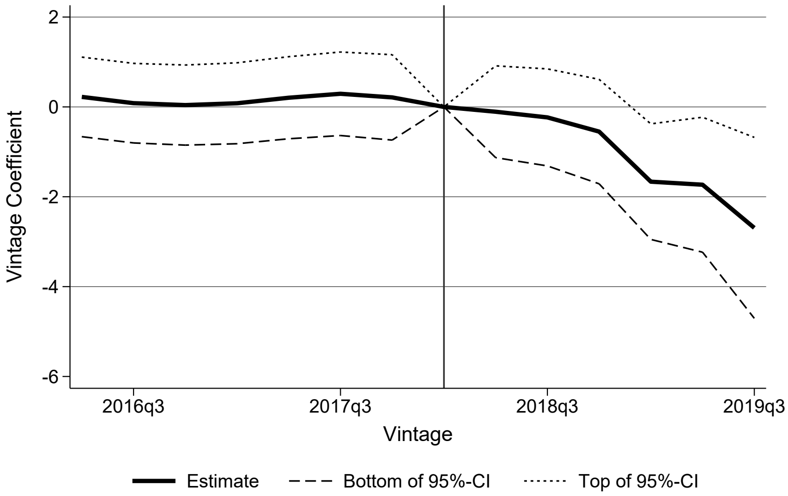


Figure 7: App Entry by Vintage

Table 2: Vintage Regressions for Entry, Usage, Average Usage per App, and Privacy

	$ln(N_{tv})$	$ln(s_{tv})$	q_{it}	$D_{P-S\ Permissions}$
Post-GDPR Dummy	-0.8170***	-0.5316***	41.6420***	-0.0785***
	(0.250)	(0.072)	(6.210)	(0.000)
Constant	11.1159***	-4.4847***	227.3353***	0.5260***
	(0.270)	(0.051)	(6.177)	(0.000)
Age Dummies	Yes	Yes	Yes	No
N	105	105	21,817,964	30,262,965
R^2	0.771	0.776	0.000	0.002

Notes: Column 1 reports a regression of the log of the number of apps born at vintage v first observed in quarter t on age indicators and an indicator for vintages born after GDPR. Column 2 reports a regression of the log of quarter t usage of apps born in vintage v on the same explanatory variables as in column 1. Columns 3 and 4 turn to app-level data. Column 3 reports a regression of app i's usage on age indicators and a post-GDPR indicator. We exclude observations prior to the first quarter in which the app is observed to avoid the delayed observation problem. Column 4 reports a regression of an indicator for whether an app requests privacy-sensitive permissions on an indicator for whether the app's vintage is post-GDPR. All regressions show robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

in the η_v index. If E_0 is the number of apps eventually observed to have entered in the 2nd quarter of 2016, then $E_t = E_0 e^{\eta_v/\eta_0}$. Figure 8 shows the resulting time series. The number of entering apps is roughly 200,000 per quarter in the middle of 2016 and remains around 200,000 until the fourth quarter of 2017, then falls to 100,000 after GDPR and subsequently even further, falling below 50,000 in the first three quarters of 2019. It is worth noting that app entry declines after GDPR is announced but before it took effect, a pattern that is consistent with our theoretical prediction that forward-looking entry is discouraged by a policy anticipated to make entry less profitable.

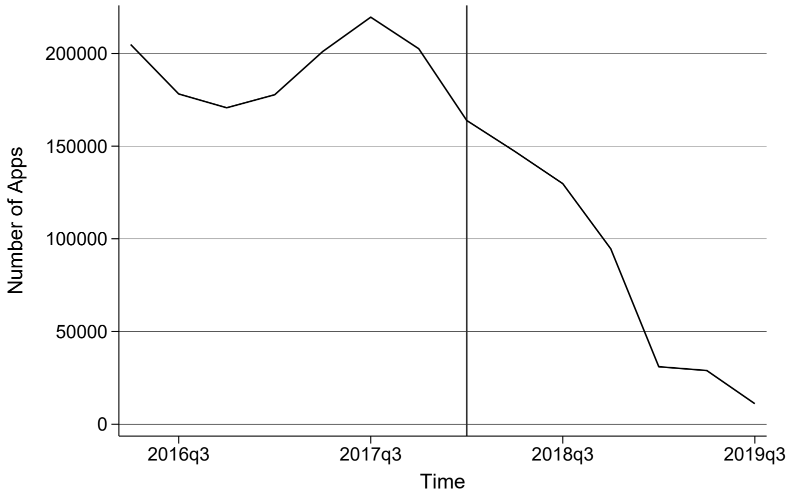


Figure 8: Implied App Entry

The post-GDPR decline in entry in Figure 8 is dramatic, and the decline plays a major role in our results. Hence, it is worthwhile to compare our finding with trends reported in other data sources. AppBrain reports aggregate app entry over time. Our data show an average reduction in entry of 61 percent, from 189,000 per quarter before GDPR to 74,000 per quarter after. AppBrain data show a 50 percent reduction, from an average of 192,000 per quarter before to 95,000 per quarter afterward.[23] The depressed entry in AppBrain data continues beyond our sample period, with app entry averaging about 60,000 per quarter during 2020.[24] The dramatic decline in entry that we document is not peculiar to our data source.

[23] Statistics by AppBrain are available at https://web.archive.org/web/20180109202516/http://www.appbrain.com/stats/number-of-android-apps, https://web.archive.org/web/20190117122626/https://www.appbrain.com/stats/number-of-android-apps, and https://web.archive.org/web/20191223023602/https://www.appbrain.com/stats/number-of-android-apps.

[24] See https://web.archive.org/web/20201027022203/https://www.appbrain.com/stats/number-of-android-apps.

5.4 Usage

Usage data provide insight on two questions. First, we can draw inferences about the utility that users derive from pre- and post-GDPR apps from the total usage of app vintages. Second, if entry is affected by development costs, and developers enter in anticipation of revenue in excess of costs, then we can draw inferences about GDPR's effects on costs in relation to the average usage per app for apps entering before and after GDPR. If costs have risen, then apps require more usage, all else equal, to be viable after GDPR.

5.4.1 Total usage by vintage

To measure the impact of GDPR on app usage, we combine the information on usage by app birth cohort over time. In particular, we regress the period-t usage of vintage-v apps ($ln(s_{tv})$) on age and vintage dummies. Unlike for the entry regressions, where we sought the number of apps born at vintage v first observed at quarter t, here we are interested in the total usage of apps by vintage and observation quarter, with the only caveat that we exclude usage in the birth quarter. Then the coefficients on vintage dummies ρ_v provide an index of app usage by birth cohort.

$$ln(s_{tv}) = \psi_{t-v} + \rho_v + \epsilon_{tv}. \tag{2}$$

Figure 9 shows the resulting vintage effects, normalized to zero in the last pre-GDPR quarter. Vintage usage is stable in the periods before GDPR takes effect and declines sharply thereafter. Table 2 reports measurements of the effect of GDPR. Column (2) includes a post-GDPR dummy, and the average pre- vs post-GDPR difference is -0.5316 (0.072). This coefficient implies that GDPR depressed the usefulness of post-GDPR cohorts by 41 percent. Because the depression in usage is smaller than the decline in entry, we can infer that the apps eliminated by GDPR are less useful than those that remain and, moreover, that app success is not completely unpredictable.

The reduced usage of post-GDPR cohorts is important for two distinct but related reasons. First, it indicates that apps entering under GDPR are collectively less useful than their pre-GDPR forebears. Although it is possible that apps adhering to GDPR's privacy protections would be perceived as more useful by consumers than earlier apps, this does not appear to be the case. The post-GDPR cohorts are smaller, and despite their enhanced privacy protections, they do not attract enough collective usage to offset their cohorts' collective reduced entry.

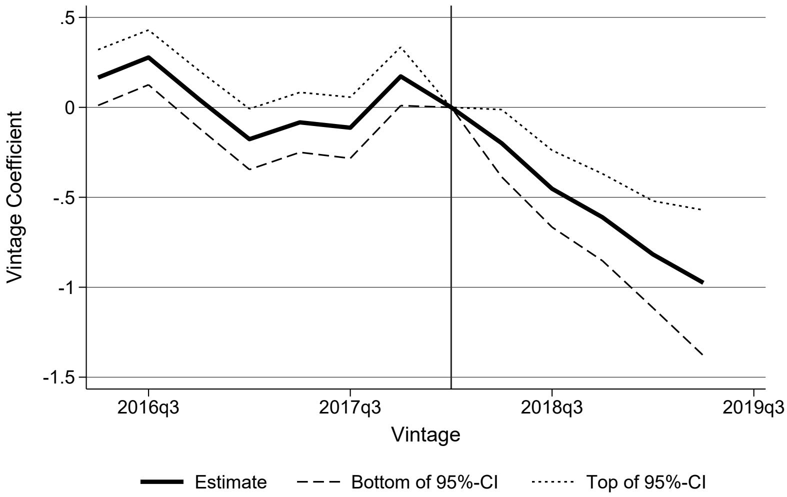


Figure 9: Usage by Vintage

5.4.2 Expected costs and average usage per app

If developers' app launch decisions are undertaken with an eye toward profit, then higher development costs – and potentially lower revenue per user – under GDPR should be reflected in higher measures of usage, per app, for apps born after GDPR.

We can explore this by regressing app-level usage on vintage and age dummies.[25] The vintage coefficients, in Figure 10, show how average usage per app evolves before and after the imposition of GDPR. While there are fluctuations prior to GDPR, average usage rises just before GDPR and sharply toward the end of the sample period. If we replace the vintage variables with a GDPR dummy beginning in the second quarter of 2018, the coefficient from a level specification is 41.64 (se = 6.21), as displayed in column (3) of Table 2, while the coefficient from a log specification is 0.10 (0.003). The raw average usage for pre-GDPR vintages is 227.33 (2.27) in levels (logs). Hence, the average usage per app rose by roughly 10-20 percent.

[25]To deal with the delayed observation problem, we include only observations once an app has been observed. In this way, we avoid mistaking the mix of apps observed at the end of the sample with the effect of GDPR.

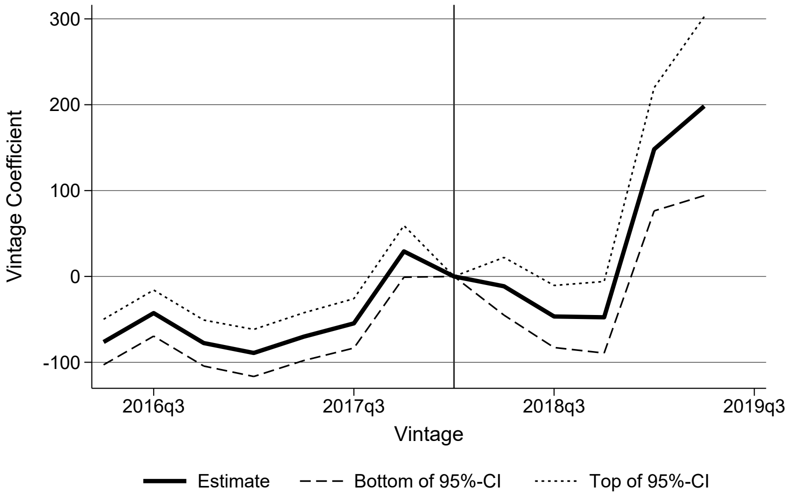
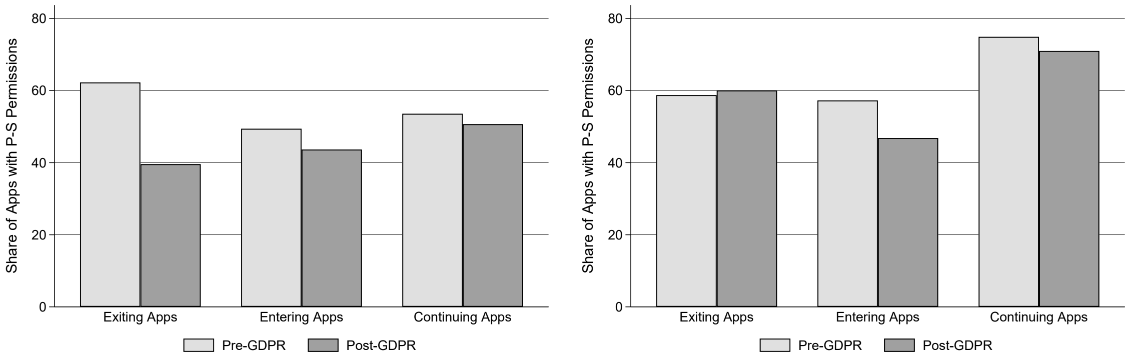


Figure 10: Average Usage per App by Vintage

5.5 Privacy

The main purpose of GDPR is the protection of user privacy, and our data allow us to examine how the privacy characteristics of available apps, or apps used, have changed over time. We focus on two such measures, the presence of privacy-sensitive permission requests by apps and whether the apps have a privacy policy. First, in Figure 11, we note that the apps exiting before the enactment of GDPR were on average more intrusive than the apps that remained. The share of apps requesting privacy-sensitive permissions among exiters was 65 percent, in comparison with 50 percent for the apps that remained. This is not only evidence that GDPR affected privacy conditions but also suggests that the exit wave documented above is induced by GDPR. Second, entering apps grow less intrusive after GDPR than before. The tendency for apps to request privacy-sensitive permissions falls, from 49.4 percent for those born pre-GDPR, to 43.7 percent for those born post-GDPR. This finding also emerges from the regression of an indicator for requesting privacy-sensitive permissions on a post-GDPR vintage indicator, in column (4) of Table 2. Some part of the post-GDPR coefficient reflects pre-existing trends: including a time trend in the regression (not reported) reduces the coefficient to -0.05.

While these changes reflect apparent effects of GDPR on user privacy, two other facts suggest that GDPR's effect is limited. First, usage-weighted measures of privacy intrusiveness changed only slightly over time. The experienced share of apps with privacy-sensitive permissions averaged 75 percent until the third quarter of 2017 and 70 after-



(a) raw

(b) weighted by usage

Figure 11: Apps with P-S Permissions by Continuation Status

wards. Second, while raw and experienced app intrusiveness declined over time, much of the change appears to have begun prior to the imposition of GDPR. In this respect, we note a steady adoption of privacy policies among apps with the share rising from 20 percent to around 80 percent, with the usage-weighted measure suggesting even higher levels.

Summary of descriptive findings: It is clear that the Android app market was substantially reshaped in the wake of GDPR. First, exit rose sharply at GDPR's imposition: The number of apps available fell by a third in the quarters immediately following implementation. Second, after GDPR, app entry fell by 56 percent, and the usage of the entry-depressed cohorts fell by 41 percent. Third, average users per app rose by about a quarter for apps born after the imposition of GDPR. Finally, apps became less intrusive after GDPR, continuing what may be a pre-existing trend. These facts are suggestive of welfare impacts on consumers and firms; and with the addition of some structure, we can estimate explicit welfare impacts.

6 Welfare

The welfare generated by apps has two parts, the surplus to consumers (CS) from the (generally) free apps, as well as the profits that developers earn (PS), largely from ad sales. That is, $W = CS + p\Sigma q_j - NC$, where q_j is the usage for app j, p is an aggregate "price" translating usage to revenue, N is the number of apps operating, and C is the development cost per app. Our estimates of the welfare impacts of GDPR on consumers and producers embed findings above (on the entry and usage impacts of GDPR) within a structural demand model. This allows us to estimate CS, as well as components of developer profits, with and without GDPR.

6.1 Consumer Surplus

Quantifying the impact of GDPR on welfare requires a measurement approach that embodies the possibility of substitution across apps. We use a nested logit demand model to compare consumer surplus in a baseline pre-GDPR period to a counterfactual long-run post-GDPR period in which 56 percent of apps were eliminated in a way that reduced the diminished cohorts' share of usage by 41 percent.

In particular, we estimate a nested logit random utility model following Berry (1994). In each quarter, consumers choose among J+1 choices (J apps and the outside good). The utility that consumer i derives from app j is:

$$u_{ij} = \delta_j + \zeta_i + (1 - \sigma)\epsilon_{ij} \tag{3}$$

In this equation, the mean utility of each product is given by $\delta_j = x_j\beta - \alpha p_j + \xi_j$, where x_j contains characteristics of app j, p_j is the download price of app j, ξ_j is the component of mean utility unobserved to the researcher, and ϵ_{ij} is an i.i.d. extreme value error. For consumer i, the variable ζ_i is common across all apps and has a distribution function that depends on σ. The apps are potentially substitutable for one another, and the degree of substitutability is summarized in the parameter σ. If σ is zero, then the nested logit resolves to the plain logit; when σ is 1, apps are perfect substitutes for one another. The more substitutable apps are for one another, the smaller the effect of a reduction in the number of products on consumer surplus. This gives rise to a closed-form equation that we can use for the estimation:

$$ln(s_{jt}) - ln(s_{0t}) = x_{jt}\beta - \alpha p_{jt} + \sigma ln(\frac{s_{jt}}{1 - s_{0t}}) + \xi_{jt} \tag{4}$$

We calculate app j's share in quarter t (s_{jt}) as the change in the number of ratings divided by the number of Android users. The term s_{0t} is the outside share in quarter t.

The vector x_{jt} contains app category dummies, and p_{jt} is the price of app j in quarter t (which is 0 for most apps).

The parameter σ is particularly important for our exercise, and its estimation requires a plausible source of exogenous variation in the number of apps available, N, arising for reasons related to supply rather than demand for apps. GDPR, by raising the cost of launching and continuing apps, makes the number of apps a reasonable instrument that we can use to identify the σ parameter. Given the model, we can calculate the quantity of each product as:

$$q_j = Ms_j = \frac{e^{\delta_j/(1-\sigma)}}{D}\frac{D^{1-\sigma}}{1+D^{1-\sigma}}, \tag{5}$$

where $\delta_j = ln(s_{jt}) - \sigma ln(\frac{s_{jt}}{1-s_{0t}}) - ln(s_{0t})$, $D = \Sigma e^{\delta_j/(1-\sigma)}$, and M is market size (the number of Android users).

While we do not have a compelling instrument for the price, we note that the proportionate change in consumer surplus arising from the long run effect of GDPR, is invariant with the price parameter. The nested logit formula for consumer surplus is

$$CS = ln\left[1 + (\sum e^{\frac{\delta_j}{1-\sigma}})^{1-\sigma}\right]\left[\frac{M}{\alpha}\right]. \tag{6}$$

In this equation the summation occurs over j apps available in a particular quarter. If CS_0 is the quarterly consumer surplus from the pre-GDPR choice set and CS_1 is the CS from the choice set contracted due to long-run GDPR effects, it is easy to see that the parameter α cancels from the proportionate change in CS: CS_1/CS_0.

Table 3 presents demand estimates, and each reported specification includes indicators for each of the nearly 50 app categories. Column (1) uses OLS, and the resulting estimate of σ is nearly 1, while the price coefficient is negative. Column (2) presents the first-stage regression of the inside share ($ln(\frac{s_j}{1-s_0})$) on the log number of apps, and the instrument works, in the sense that time periods with more apps have significantly smaller average app shares. Instrumenting the inside share, in column (3), delivers a σ estimate of 0.361 (standard error = 0.004). This estimate indicates partial substitutability of apps for one another.

To evaluate the long run effect of GDPR on consumer surplus, we start with a pre-GDPR choice set (corresponding to the second quarter of 2017), and we calculate the

Table 3: Nested Logit Estimation Results

	OLS	1st Stage	IV
$ln(\frac{s_j}{1-s_0})$	0.99573		0.36079
	(0.00003)***		(0.00355)***
App Price	-0.00019		-0.00606
	(0.00002)***		(0.00019)***
$ln(\#Apps)$		-0.52898	
		(0.00324)***	
Category Dummies	Yes	Yes	Yes
R^2	0.99	0.00	0.61
N	13,753,320	13,754,377	13,753,320

Notes: This table shows the results of the demand estimation corresponding to equation 4. Column (3) instruments an app's inside share with the total number of apps in the market. The first stage of this IV-regression is shown in column (2). The table uses only 13,753,320 observations (cf. Table 1), because taking the logarithm of the usage measure leaves only the 43.8 percent of observations with positive values. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

baseline pre-GDPR consumer surplus (CS_0). We model the long run effect of GDPR – with steady-state entry depressed by 56 percent – by removing 56 percent of apps from the baseline choice set. If we remove 56 percent of apps at random (therefore removing apps as useful as those that remain), then we obtain an upper bound on the welfare effect. Taking an average across 500 draws, this reduces the quarter's CS from \$45.0 billion to \$27.5 billion, or by 38.9 percent. In reality, developers have some ability to predict app quality at entry, so that the missing apps are not as good, on average, as those that remain.

To more accurately measure the welfare loss from reduced entry, we need to model entry in a way that delivers the observed relationship between the cohort entry and usage reductions. We do so as follows. Define δ_j' as the predicted quality of app j at entry. Predicted quality is true quality plus a random error. That is, $\delta_j' = \delta_j + \kappa\nu_j$, where ν_j is N(0,1), and κ is a scaling parameter. We then simulate GDPR's effect on the app choice set: Given a scaling parameter κ and a draw on ν_j terms, we order apps by δ_j', from highest to lowest expected revenue, then remove the bottom 56 percent of apps according to δ_j'. If the scaling parameter κ is large, this operates as random removal. As κ gets smaller, the predictability of realized app quality improves (reaching perfect predictability when κ equals 0). We simulate for a range of values of κ, then fit a line between κ and the percent reduction in usage. We then choose the value of κ that delivers the observed reduction in vintage usage documented earlier (41 percent). With our chosen $\kappa^* = 15.988$, the correlation between expected and realized app quality – between δ' and δ – is 0.09. Given κ^*, we average across 1,000 draws, and the resulting change in quarterly consumer surplus, in Table 4, is from \$45.0 billion without GDPR to \$30.5 billion with GDPR, a reduction of 32.2 percent.

Table 4: Welfare Effects of GDPR

	CS (\$ mil)	# Apps (N)	Avg. Usage per App	Expected Usage Marginal App	App FC	Aggreg. App Revenue (\$ mil)	Aggreg. App Costs (\$ mil)	Aggreg. App Profits (\$ mil)
Pre-GDPR	44,990	1,148,445	223.4	84.1	4,868	14,850.0	5,590.4	9,259.6
GDPR	30,500	505,316	373.8	272.7	15,784	10,932.9	7,975.9	2,957.0
Change	-14,490	-643,129	150.4	188.6	10,916	-3,917.1	2,385.6	-6,302.7
% Change	-32.2%	-56.0%	67.3%	224.3%	224.3%	-26.4%	42.7%	-68.1%

Notes: The pre-GDPR row is based on simulations with the Q2 2017 choice set, in which 1,148,445 apps (with nonzero usage) are available. The GDPR calculations are based on the removal of 56 percent of apps in a way that reduces the share of usage accounted for those apps by 41 percent. Usage is translated into revenue using the 2017 Android app revenue estimate (\$59.4 billion annual, divided by aggregate usage, divided by four quarters). The resulting "price" is held constant in the GDPR period. Usage and dollar figures are based on a quarter.

6.2 Developer Profits

The remaining welfare effect is the change in developer profits: $\Delta\pi = [p_1\Sigma q_j^1 - N_1C_1] - [p_0\Sigma q_j^0 - N_0C_0]$, where 0 denotes pre-GDPR and 1 denotes post-GDPR. Calculating developer profits requires four quantities, in each of the two – GDPR and non-GDPR – equilibria. First, we need total app usage (Σq_j^0). The pre-GDPR values are available as data from the baseline quarter. Given estimated σ and a choice of remaining products, we calculate the post-GDPR q_j values for each product and Σq_j using equation 5.

Second, the number of available apps is directly observed prior to GDPR (N_0), while the post-GDPR figure (N_1) is given from the descriptive results above. Because GDPR reduces entry by 56 percent, $N_1 = 0.44N_0$.

Third, we need measures of the fixed costs of app development before and after GDPR, C_0 and C_1. Each of these, in turn, is the product of the expected usage of the marginal app and the revenue per user. We can calculate the expected usage of the marginal entering app from our entry model. On the assumption that entry occurs as long as it is expected to be profitable, the fixed cost of app development (C_0 or C_1) is the revenue per user times the expected usage of the marginal entering app. These estimates of app development costs are based on marginal apps and are, technically, estimates of the fixed costs of the marginal entering app. Inframarginal apps could have higher development costs and still expect profits from entry. Hence, the estimates of per-app and aggregate development costs are lower bounds.

To calculate the expected usage of the marginal entering product, note first that with our demand model, the *realized* quantity of product j is given by: $M \frac{e^{\delta_j/(1-\sigma)}}{D} \frac{D^{1-\sigma}}{1+D^{1-\sigma}}$, where $D = \Sigma e^{\delta_j/(1-\sigma)}$. To calculate the *expected* revenue of the $K+1^{st}$ entrant, we take a draw of the top K products in expected quality, which gives us an estimate of D_K (based on a summation to the K^{th} product). We then calculate the expected revenue for the $K+1^{st}$ entrant, using $ME[\frac{e^{\delta_{K+1}/(1-\sigma)}}{D'} \frac{D'^{1-\sigma}}{1+D'^{1-\sigma}}]$, where $D' = D_K + e^{\delta_{K+1}}$. To speed up the estimation, we estimate the average usage for the $K+1^{st}$ product using the 10,000 products from $K+1$ to $K+10,000$ along with a D_K that includes the top K products in the draw. We take 1,000 draws on the δ_j vector ordered by δ_j' (of the top K products), so our estimate of the expected usage of the marginal entering app is effectively based on 10,000,000 draws.

The expected usage of the marginal entrant rises from 84.1 without GDPR to 272.7 under GDPR, or increases by a factor of over three, while the average usage per app $\Sigma q_j/N$ rises from 223.4 without GDPR to 373.8 under GDPR. Because the total number of apps falls from the baseline 1.148 million to 0.505 million, aggregate usage falls from 256.6 million to 188.9 million, or by 26.4 percent. If revenue per user were constant, then app revenue would fall by 26.4 percent as well. Moreover, with constant prices, the expected cost of developing the marginal app would rise by a factor of more than three.

Translating effects on usage into dollar costs requires prices that translate aggregate usage into revenue. Using data on user and ad revenue for apps discussed in section 2.1, along with Google's market share, we roughly estimate revenue to be \$59.4 billion in 2017 (prior to GDPR), or just under \$15 billion per quarter. Using this total gives us the dollar figures in Table 4.

Drawing inferences about changes in costs from the change in expected usage of marginal apps is complicated by the possibility that GDPR changes both development costs and revenues per user. Despite concerns about GDPR reducing revenue, preliminary journalistic accounts give no indication of reduced revenue.[26] What matters for the proportionate changes in dollar values of per-app development costs, aggregate revenues, or aggregate costs is the growth rate of the revenue per user, not its initial level. For example, if revenue per user fell by 25 percent while expected usage of the marginal app doubled, then we would infer that the expected cost of developing the marginal app had risen by 50 percent (=(1-0.25)2-1 = 0.5). If we define p_1 as revenue per user in the GDPR equilibrium and p_0 as pre-GDPR revenue per users, then $p_1 = (1+\phi)p_0$; and if we define q'_0, for example, as the expected usage of the marginal entering app without GDPR, then the proportionate growth in per-app development costs is $C_1/C_0 = (1+\phi)\frac{q'_1}{q'_0}$. The proportionate growth in revenue is $(1+\phi)\frac{\Sigma q_j^1}{\Sigma q_j^0}$. The proportionate growth in aggregate costs is $(1+\phi)\frac{N_1}{N_0}\frac{\Sigma q_j^1}{\Sigma q_j^0}$. Hence, if revenue per user did not change ($\phi = 0$), then the cost of developing the marginal app would rise by 124 percent. Given the 56 percent in the number of apps operating, aggregate development costs would rise by 43 percent. Reduced usage would decrease aggregate revenue by 26 percent. It is difficult to know what "price" translating our usage measure to total revenue would prevail in the eventual GDPR equilibrium, but we can see that costs would rise, and revenue would fall, for a wide range of possible values of ϕ. Unless revenue per user rose by at least a third ($\phi > 1/3$), aggregate revenue would fall. Unless revenue per user fell by more than 50 percent ($\phi < -0.5$), aggregate costs would rise.

Robustness to σ: Our welfare analysis depends on the estimated parameter σ reflecting the degree of substitutability among apps; and it is of interest to know how our results would change with different values of σ. To this end we calculate alternative results for different values of σ, using the following procedure. We choose a value of σ. Because $\delta_j = ln(s_j) - \sigma ln(\frac{s_j}{1-s_0}) - ln(s_0)$, our mean utilities depend on σ. Given a vector of $\delta(\sigma)_j$, we need to find the scaling parameter κ such that a 56 percent reduction in the number of products delivers a 41 percent reduction in the usage of the depressed cohorts. Given $\kappa^*(\sigma)$, we can calculate pre-GDPR and GDPR CS and the components of PS.

The dependence of κ on σ complicates the effect of alternative σ values on the results. Holding κ constant, raising σ raises the expected usage of a marginal app; and holding

[26] If revenue per user did not change, then a doubling in average usage would imply that the expected cost of creating the marginal app had doubled as well. If revenue per user (p) fell, however, then we would infer that the cost change fell short of a doubling.

σ constant, raising κ raises the expected usage of a marginal app. A higher σ, however, requires a lower value of κ to deliver the target (41 percent) usage reduction. Hence, the effect of alternative σ values on the basic results need not – and turns out not to – be monotonic.

Table 5 reports results for $\sigma = 0.25$ and $\sigma = 0.75$, in addition to the baseline $\sigma = 0.361$. With $\sigma = 0.25$, GDPR would reduce CS by 29 percent, slightly less than the baseline 32 percent reduction. With $\sigma = 0.75$, by contrast, GDPR would reduce CS by 11 percent.

Table 5: Sensitivity of Welfare Loss

		Average usage per app		Expected usage of marginal app	
σ	% change CS	pre-GDPR	GDPR	pre-GDPR	GDPR
0.25	-28.7	223.4	356.3	80.2	264.7
0.361	-32.1	223.4	373.8	84.1	272.7
0.75	-11.4	223.4	458.7	70.5	325.8

Different degrees of substitutability also affect the usage measures. While the baseline pre-GDPR estimate of the expected usage of the marginal app is 84.1, it is 80.2 with $\sigma = 0.25$ and 70.5 with $\sigma = 0.75$. Expected usage of the marginal app changes with σ for the GDPR counterfactual as well. The proportionate change in the expected usage of the marginal app is similarly large across a range of σ parameters. The ratio of the post-GDPR expected usage of the marginal app to the pre-GDPR value – 3.2 in the baseline – is 3.3 when $\sigma = 0.25$ and 4.6 when $\sigma = 0.75$.

For a wide range of σ parameters, GDPR substantially depresses consumer surplus and substantially raises the expected usage – and therefore the expected development cost in relation to revenue – for the marginal app. We conclude that our broad result – that the privacy benefits of GDPR come at substantial costs in consumer and producer surplus – are robust to different degrees of substitutability across apps.

7 Conclusion

GDPR has had substantial effects on Google's app market. In the year following its implementation, about a third of existing apps exited the market; and following GDPR's enactment, the rate of app entry fell by more than half. Moreover, GDPR-diminished entry cohorts account for 41 percent less app usage than their pre-GDPR counterparts, indicating that the missing apps would have been valuable. Finally, apps entering after GDPR have higher average usage per app, suggesting increased development costs. We incorporate these patterns into a structural model of app demand and entry, and we find that GDPR reduces consumer surplus, app usage, and – if revenue per user did not change – developer revenue by about a quarter.

We have two broad conclusions, one about innovation in general and the other about GDPR in particular. First, we conclude that GDPR, whatever its beneficial impacts on privacy protection, also produced the unintended consequence of slowing innovation. It is possible that privacy is valuable to consumers in ways that do not manifest themselves in usage choices. Indeed, this is the "privacy paradox" that others (Acquisti et al., 2016; Norberg et al., 2007) have documented: Citizens clamor for privacy protections in ways that belie their behavior as consumers. We are hesitant to draw policy conclusions about the advisability of GDPR from our results alone. A full evaluation of GDPR requires a tallying of the potential beneficial effects on privacy, along with its various unintended consequences such as increases in market concentration (Batikas et al., 2020; Johnson et al., 2020), undermining revenue models for content production (Lefrere et al., 2020), and – here – reducing beneficial innovation.

Second, we take our findings as additional evidence that when product quality is unpredictable, the ease of entry is an important determinant of the ex post value of the choice set to consumers. Factors reducing entry costs deliver large welfare benefits, while factors hindering entry – such as GDPR – can deliver substantial welfare losses.

References

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman, "The Economics of Privacy," *Journal of Economic Literature*, 2016, *54* (2), 442–92.

Aguiar, Luis and Joel Waldfogel, "Quality predictability and the welfare benefits from new products: Evidence from the digitization of recorded music," *Journal of Political Economy*, 2018, *126* (2), 492–524.

Ahn, Andrew, "How we fought bad apps and malicious developers in 2017," https://android-developers.googleblog.com/2018/01/how-we-fought-bad-apps-and-malicious.html 2018.

Aridor, Guy, Yeon-Koo Che, and Tobias Salz, "The Economic Consequences of Data Privacy Regulation: Empirical Evidence from GDPR," *NBER Working Paper No. 26900*, 2020.

Arrow, Kenneth J, "Classificatory notes on the production and transmission of technological knowledge," *American Economic Review*, 1969, *59* (2), 29–35.

Batikas, Michail, Stefan Bechtold, Tobias Kretschmer, and Christian Peukert, "European Privacy Law and Global Markets for Data," *CEPR Discussion Paper No. DP14475*, 2020.

Bergemann, Dirk and Ulrich Hege, "The financing of innovation: Learning and stopping," *RAND Journal of Economics*, 2005, pp. 719–752.

Böhmecke-Schwafert, Moritz and Crispin Niebel, "The General Data Protection's (GDPR) Impact on Data-Driven Business Models: The Case of the Right to Data Portability and Facebook," *ITU Journal*, 2018, *1* (2), 1–9.

Carare, Octavian, "The Impact of Bestseller Rank on Demand: Evidence from the App Market," *International Economic Review*, 2012, *53* (3), 717–742.

de Matos, Miguel Godinho and Idris Adjerid, "Consumer Consent and Firm Targeting after GDPR: The Case of a Large Telecom Provider," *Working Paper*, 2020.

Ershov, Daniel, "Consumer Product Discovery Costs, Entry, Quality and Congestion in Online Markets," *Working Paper*, 2020.

Ewens, Michael, Ramana Nanda, and Matthew Rhodes-Kropf, "Cost of experimentation and the evolution of venture capital," *Journal of Financial Economics*, 2018, *128* (3), 422–442.

Ghose, Anindya and Sang Pil Han, "Estimating Demand for Mobile Applications in the New Economy," *Management Science*, 2014, *60* (6), 1470–1488.

Goldberg, Samuel, Garrett Johnson, and Scott Shriver, "Regulating Privacy Online: An Economic Evaluation of the GDPR," *Working Paper*, 2020.

Goldfarb, Avi and Catherine E Tucker, "Privacy regulation and online advertising," *Management Science*, 2011, *57* (1), 57–71.

Hausman, Jerry A, "Valuing the effect of regulation on new services in telecommunications," *Brookings Papers on Economic Activity*, 1997, *1997*, 1–54.

Jia, Jian, Ginger Zhe Jin, and Liad Wagman, "The Short-Run Effects of GDPR on Technology Venture Investment," *NBER Working Paper No. 25248*, 2019.

Johnson, Garrett, Scott Shriver, and Samuel Goldberg, "Privacy & market concentration: Intended & unintended consequences of the GDPR," *Working Paper*, 2020.

Kerr, William R, Ramana Nanda, and Matthew Rhodes-Kropf, "Entrepreneurship as experimentation," *Journal of Economic Perspectives*, 2014, *28* (3), 25–48.

Kesler, Reinhold, Michael Kummer, and Patrick Schulte, "Competition and Privacy in Online markets: Evidence from the Mobile App Industry," *ZEW Discussion Paper No. 19-064*, 2019.

Kummer, Michael and Patrick Schulte, "When private information settles the bill: Money and privacy in Google's market for smartphone applications," *Management Science*, 2019, *65* (8), 3470–3494.

Lefrere, Vincent, Logan Warberg, Christobal Cheyre, Veronica Marotta, and Alessandro Acquisti, "The Impact of the GDPR on Content Providers," *Working Paper*, 2020.

Leyden, Benjamin T, "There's an app (update) for that: Understanding product updating under digitization," *Working Paper*, 2019.

Manso, Gustavo, "Motivating innovation," *The Journal of Finance*, 2011, *66* (5), 1823–1860.

_ , "Experimentation and the Returns to Entrepreneurship," *The Review of Financial Studies*, 2016, *29* (9), 2319–2340.

Momen, Nurul, Majid Hatamian, and Lothar Fritsch, "Did App Privacy Improve After the GDPR?," *IEEE Security Privacy*, 2019, *17* (6), 10–20.

Norberg, Patricia A, Daniel R Horne, and David A Horne, "The privacy paradox: Personal information disclosure intentions versus behaviors," *Journal of Consumer Affairs*, 2007, *41* (1), 100–126.

Osborne, Charlie, "Google plans purge of Play Store apps without privacy policies," https://www.zdnet.com/article/google-plans-purge-of-play-store-apps-without-privacy-policies/ 2017.

Petrin, Amil, "Quantifying the benefits of new products: The case of the minivan," *Journal of Political Economy*, 2002, *110* (4), 705–729.

Siddiqui, Aamir, "Google Play Store approval for new apps will now take more time," https://www.xda-developers.com/google-play-store-approval-new-apps-time/ 2019.

Sørensen, Jannick and Sokol Kosta, "Before and After GDPR: The Changes in Third Party Presence at Public and Private European Websites," *Proceedings of the 2019 World Wide Web Conference*, 2019, pp. 1590–1600.

Tucker, Catherine E, "The Economics of Advertising and Privacy," *International Journal of Industrial Organization*, 2012, *30* (3), 326–329.

_ , "Social networks, personalized advertising, and privacy controls," *Journal of Marketing Research*, 2014, *51* (5), 546–562.

Waldfogel, Joel, "Copyright protection, technological change, and the quality of new products: Evidence from recorded music since Napster," *The Journal of Law and Economics*, 2012, *55* (4), 715–740.

Weitzman, Martin L, "Optimal search for the best alternative," *Econometrica*, 1979, *47* (3), 641–654.

Zhuo, Ran, Bradley Huffaker, Shane Greenstein, and KC Claffy, "The Impact of the General Data Protection Regulation on Internet Interconnection," *NBER Working Paper No. 26481*, 2019.

A Appendix

A.1 Additional Information on Timing

The official law of GDPR was passed in April 2016 but came into effect only in May 2018. Hence, most of those affected were able to inform themselves about the regulation and necessary adjustments to be implemented already some time before. If this transition period was used, one could cast doubt on the assumption of GDPR serving as an exogenous shock in May 2018. In order to investigate the timing as to when GDPR affected the app market, we collected different indicators about the awareness of the consequences stemming from the GDPR. Figure 1 shows results on three measures with the maximum of each time-series serving as the benchmark for our observation period.

First, as GDPR is of enormous scope for firms as well as individuals, it is, therefore, of general public interest. This can be approximated by searches for 'GDPR' on Google as the most popular search engine worldwide. The figure shows that although the number of searches rose slightly from 2016 onwards, there is a big jump only around the enactment. While Google searches mainly absorb the demand side for information, Wikipedia can serve as a measure of information provision. In this sense, we analyzed the editing behavior on Wikipedia which may represent the status quo of common knowledge and details of GDPR, where we find the same pattern as before. Lastly, we are interested in app developers' awareness about GDPR in particular. Developers could have adjusted apps already some time before the enactment of GDPR if sufficient information was around and they were willing to do so. To analyse this, we used web-scraped data of large online developer forums (Stack Overflow with 'android' tag, Reddit subreddit 'r/androiddev', and 'android-forum') and analyzed, when posts and comments included 'GDPR' as a keyword. As depicted graphically, the peak around GDPR relative to the pre-period is even more pronounced in this case. Developers have not talked about GDPR until the beginning of 2018 and only started shortly before enactment. Overall, our findings serve as anecdotal evidence that the awareness of GDPR was rather limited during the transition period and, especially for app developers, was of interest only when forced to comply.

A.2 Developer Survey

To shed further light on the experiences of companies who have to comply with GDPR, we conducted a survey among app developers with regard to GDPR. This offers additional information about the benefits and challenges coming with the new privacy regulation. In the following, we will provide further details on the data and results.

A.2.1 Data

We conducted the survey on consequences of GDPR for apps in the Google Play Store in October 2019. The questionnaire was sent to German app developers identified in the Google Play Store during the time period from October 2015 to January 2019. For the analysis, we can use about 650 valid answers. In the survey, we asked for basic information like developer types, the point in time of entering the Google Play Store or the number of apps as well as information regarding their users, data usage, revenue and specific consequences of GDPR. The study was implemented in LimeSurvey, an online-survey tool, and contained 31 different questions.

A.2.2 Developer Types and Apps

To get an impression what types of developers exist in the Google Play Store the respondents had to select themselves into the four categories (i) hobby developer, (ii) self-employed with employees, (iii) self-employed without employees, and (iv) company.

Figure A.1: Distribution of Developer Types

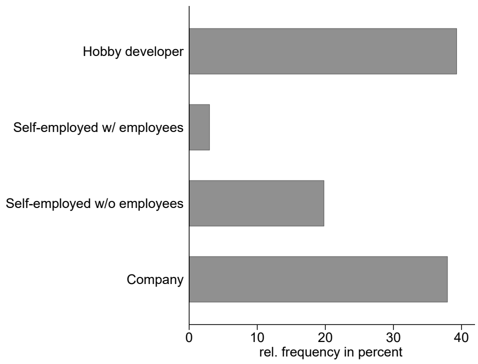
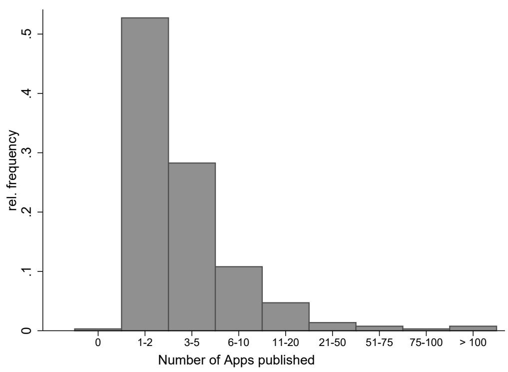


Figure A.1 shows that most of the respondents develop apps in their free-time (37.2%) or within a company (41.4%). Self-employed individuals rather work on their own without any employees (17.74% vs. 3.66%). Furthermore, the majority of self-employed developers without employees (58.18%) develop apps as a sideline job.

The beginning of their activity on the Google Play Store ranges from the opening of the platform in 2008 to the time of our survey, with an increase until the end of 2017 and a decline afterwards.

Figure A.2 shows the share of developers with a given number of published apps from our survey. On average, developers created 5.8 apps with a median of 2 apps and a maximum value of 250 apps. Only 72% of their apps ever published are still available in the Play Store.

Figure A.2: Distribution of Apps Published



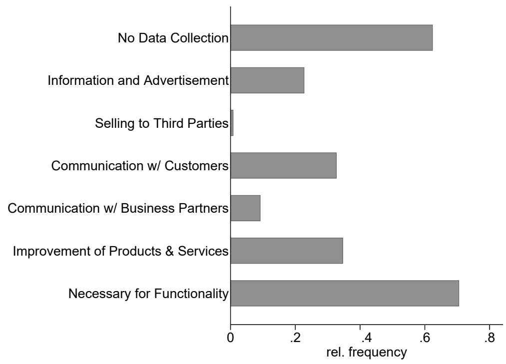
As GDPR has a extraterritorial scope, where all companies have to comply which serve users within the EU, we were interested in the origin of app users. Of course, the results are not representative as we only surveyed developers in Germany. Therefore, it is not surprising that 96.6% of developers have Europeans among their user group. The other regions (North America, South America, Asia, Africa, and Australia) are stated in 20% to 40% of cases. About 6% percent of developers do not know their users' origin.

A.2.3 Data Usage

GDPR dictates stringent regulations regarding data usage and sharing by the controller along with demands for transparent disclosure. This means that developers and apps, which use and share users' data a lot, may be more affected and concerned compared to those which do not. Figure A.3 depicts the collecting and use of personally identifiable information. 61.14% of developers do not collect this kind of data at all. Of those who do, the majority need personal data for the app's functionality (68.18%).

Other reasons for data collection are distribution of information and ads (22.38%), data selling to third parties (0.7%), communication with customers (31.82%), and business partners (8.74%), as well as improvements of products and services (32.87%).

Figure A.3: Data Collection and Usage



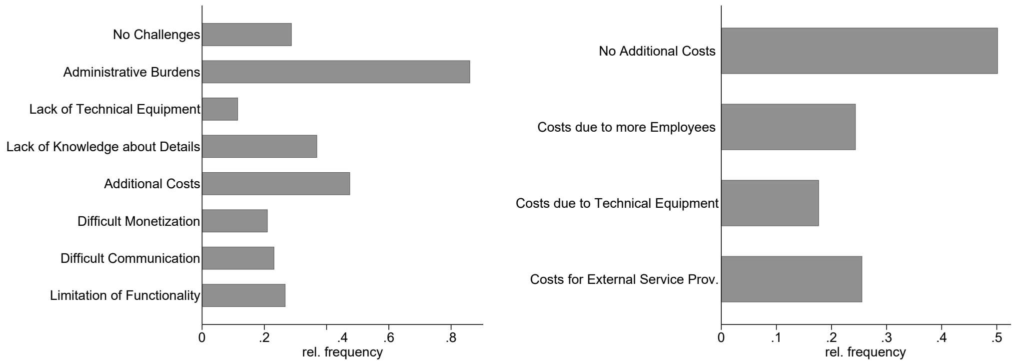
A.2.4 GDPR and its consequences

99.28% of respondents know about the introduction of GDPR, 60.68% of respondents see it as an effective instrument for better data protection. While more than 80% of developers do not see any changes in demand for their apps or the level of data collection, there are still quite some effects on their everyday work. 14.22% of developers have removed at least one of their apps temporarily. This may be due to necessary adaptions in order to comply with the new regulations. For 6.89% of respondents even a complete deletion of at least one of their apps was necessary. Others did not launch prepared apps due to GDPR (8.7%).

Figure A.4 shows the participants' answers to questions regarding challenges and costs associated with GDPR. The three most prevalent challenges are administrative burdens (85.15%), additional costs (47.92%), and a lack of knowledge about the regulation's details (36.44%).

As the right panel of Figure A.4 depicts, costs associated with GDPR may come from additional staff, necessary technical equipment, or external service providers (e.g., acting as data protection officer which has to be in place for many companies). These types of costs have in common that they can be assumed to be fixed costs for companies. Hence, larger companies or developers with several apps may benefit from fixed cost degression compared to small and single-app developers.

Figure A.4: Challenges and Costs associated with GDPR



A.2.5 Revenue

Not all of the respondents make revenue by developing and selling apps, this is only the case for 43.07%. In case of revenue generation, sources are in-app-prices (44.34%), advertisements (42.72%), app prices (41.75%), paid memberships (12.94%), or data selling to third parties (0.97%). The revenue strategies remain constant, even after GDPR. When asking for changes in revenue streams the majority of respondents do not record larger differences. Nevertheless, 40% of respondents with revenues have seen a lower importance of advertisements for revenue generation, while at least 16.67% register increases in paid memberships.

A.3 Additional Information on Data

Data Validation: As we web-scrape the English-speaking version of the Google Play Store from Germany[27], there may be concerns about the external validity of the data and results. First, we look into outside sources like AppBrain confirming our set of available apps to be complete and similar patterns to be present (cf. section 4.1 and 5). Second, we repeat our web-scraping from the US and several other non-European countries to verify that exited apps did not appear elsewhere.

Imputation: As described in section 4.1, we have two kinds of missing data. 'Missing in-between' observations are imputed by considering the app's observations before and after the 'gap.' We observe and impute 6.82 percent of in-between missings. Installations and ratings are interpolated with the average, whereas all the other measures are carry-

[27]Additional details about the data collection routine are provided in Kesler et al. (2019).

forwarded. We impute missing observations that result from apps that are already born but not yet observed, using the information from the first observation and the known date of birth. We imputed 9.69 percent of observations before the app is observed. Installations and ratings are interpolated with the three-month average based on the first (real) observation, while the other measures get the value of the first (real) observation.

Measuring Usage: Figure A.5 shows for the first period of observation the relationship between the log of cumulative ratings for each app (along with 90th and 10th percentiles) and the cumulative installation quantity category. The relationship between the two measures of usage is strong and monotonic, which inspires confidence, that they are informative of usage.

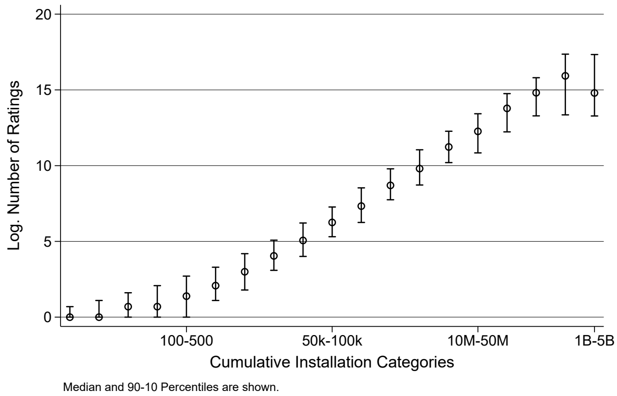


Figure A.5: Ratings and Cumulative Installations in First Quarter

Locating Developers: In order to retrieve the country of origin by a developer, we first look at the contact address given on the app's Play Store page. However, in most of the cases, a developer did not report a contact address as it is not mandatory. In that case, the next best guess is to take the last part of the app's e-mail (some of which might hint to a country such as 'co.uk' or '.de'). Lastly, we look at the first part of the Google ID (starting with 'de.' for example) that can be definitely associated with a specific country. In total, we are able to successfully locate 40.1 percent of developers.