NBER WORKING PAPER SERIES

THE GDPR AND SDK USAGE IN ANDROID MOBILE APPS

Ginger Zhe Jin
Ziqiao Liu
Liad Wagman

Working Paper 33099
http://www.nber.org/papers/w33099

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2024

We are thankful to Apptopia, the Data Catalyst Institute, and the Internet Safety Labs for providing access to the data, and to the George Mason University Research Roundtable on Competition and Consumer Protection Issues Surrounding Information Flows for financial support. James Cooper, Neil Chilson, Anthony Dukes, Daniel Gilman, Il-Horn Hann, Ben Leyden, Geoffery Manne, Dina Mayzlin, Balaji Padmanabhan, Sasha Romanosky, Daniel Sokol, Guofu Tan, and participants at the GMU Research Roundtable, the University of Southern California, and the University of Maryland Smith School of Business have provided constructive comments. All rights reserved. All errors are ours. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The GDPR and SDK Usage In Android Mobile Apps
Ginger Zhe Jin, Ziqiao Liu, and Liad Wagman
NBER Working Paper No. 33099
November 2024
JEL No. D22, D8, L15, L51, L86

ABSTRACT

Using data on Software Development Kits (SDKs), we study SDKs in Android apps vis-à-vis the adoption of the General Data Protection Regulation (GDPR) by the European Union in May 2018. Relative to US-only apps, the number of non-Google SDKs used per app in 5 major European countries (EU5) exhibits a 1.3% decline in EU5-only apps after GDPR, and a 6.3% decline in apps multihoming US and EU5. These effects are mostly driven by lower-ranked apps. Using apps' monthly active users as weights in computing average user exposure to SDKs in app category and country groups, we find no significant change in the average SDK exposure in EU5-only relative to US-only apps. The only significant change in multihoming apps is a 4.2% drop in exposure to SDKs developed by minor SDK providers, although SDKs from Google and major SDK providers were independently assessed as riskier than those from minor SDK providers.

Ginger Zhe Jin
University of Maryland
Department of Economics
College Park, MD 20742-7211
and NBER
ginger@umd.edu

Liad Wagman
Lally School of Management
Rensselaer Polytechnic Institute
110 8th Street
Troy, NY 12180-3590
wagman@rpi.edu

Ziqiao Liu
University of Maryland
zliu26@umd.edu

1 Introduction

To safeguard user data, the European Union's (EU's) General Data Protection Regulation (GDPR) mandates firms to incorporate "privacy by design," yet it provides limited technical guidance on how to achieve this. Although a growing body of research documents the GDPR's effects across various dimensions,[1] little is known about its impact on the back-end layers of digital products, particularly the technical architectures of apps, which govern data flows between apps and the user's device. To address this gap, we examine whether and how Android mobile apps have adjusted their usage of Software Development Kits (SDKs) before and after the GDPR's rollout in May 2018.

SDKs are broadly defined as collections of software tools and libraries that developers can use to create applications for specific platforms or hardware. While developers can build similar tools in-house, third-party SDKs are widely adopted in mobile app development due to their pre-built nature, low cost (often free), and ease of integration. These SDKs significantly reduce development time and resources. For instance, a developer may use a payment processing SDK for in-app purchases, another SDK for user analytics and app crash reporting, and a third SDK to deliver advertisements to the app's users.

The GDPR, as a landmark data regulation, can influence app developers' incentives to use third-party SDKs in multiple ways. On the benefit side, the GDPR allows European citizens to decide whether to share their personally identifiable information (PII) with app developers. This may result in less data available for monetization, reduced benefits from data collection, and subsequently, fewer incentives to adopt SDKs. On the cost side, the GDPR empowers regulators to fine firms found to have mishandled or misused EU users' PII. Since app developers have greater control over their own code than over third-party SDKs, the GDPR might encourage them to exert stricter control over data practices, potentially leading to reduced SDK usage. These effects — both on the benefit and cost sides — are likely to be more pronounced for SDKs that process sensitive PII-related data (e.g., biometrics), as EU users may become more hesitant to share such data post-GDPR. Consequently, the GDPR may have a more significant negative effect on developers' use of SDKs that handle sensitive PII-related data.

Several factors may further complicate the benefit-cost analysis of SDK usage under the GDPR. First, some SDK providers—particularly those associated with larger digital platforms—may have more resources and expertise to produce GDPR-compliant SDKs than the average app developer. This creates a trade-off for app developers between compliance and control: using third-party SDKs might help them avoid the costs of developing in-house functionalities to comply with the GDPR, but it also means relinquishing some control over the app's data practices to third-party SDK providers.

[1] Johnson, 2024 provides a recent survey.

Second, significant legal ambiguity remains regarding liability for GDPR violations in cases where an EU citizen's PII is misused. While mobile apps serve as the primary interface for user consent to data collection, the PII collected via a third-party SDK may be accessible to both the app developer and the SDK provider. It is unclear whether users fully understand who has access to and processes their data when they consent to data collection. Given this ambiguity, app developers may prefer using third-party SDKs to shift or share the risk of future GDPR enforcement with SDK providers, whereas in-house development would likely make the developer the clearly responsible entity in case of GDPR violations. These complexities may incentivize app developers to adopt third-party SDKs—particularly those from larger platforms—if they believe these SDKs can help them achieve GDPR compliance or mitigate enforcement risks.

The GDPR may thus influence app developers to either increase or decrease their use of third-party SDKs, depending on how they weigh the reduced benefits of data sharing and monetization, the costs of ensuring GDPR compliance through either in-house development or third-party SDKs, and the potential risk of GDPR liability.

To examine these considerations, we assess the impact of the GDPR on SDK usage by Android apps, using a large dataset from Apptopia. Since the GDPR applies broadly to products and services offered to EU users, our main identification strategy compares mobile apps that serve only US users (hereafter US-only apps) with those that serve users in any of five major European countries (UK, France, Germany, Spain, and Italy) but not in the US (hereafter EU5-only apps), and apps that serve both US and EU5 markets (henceforth multihoming apps).

In the raw data, we observe that, on average, both US-only and EU5-only Android apps have increased SDK usage post-GDPR, despite the GDPR's emphasis on data minimization. Multihoming apps exhibit varied trends, with SDK usage increasing or decreasing post-GDPR, depending on the type of SDK. Before the GDPR's rollout, US-only, EU5-only, and multihoming apps exhibited different trends in SDK usage, motivating the use of a synthetic difference-in-differences (SDID) framework to ensure comparability in our analysis.

Our SDID estimates find that, relative to US-only apps, the number of non-Google SDKs used per app decreased by 1.3% in EU5-only apps and by 6.3% in multihoming apps after the GDPR's rollout. However, these declines are mostly driven by lower-ranked apps tracked in our data, and are *not* more pronounced for SDKs handling more sensitive PII, as might be expected.

We further classify a non-Google SDK provider as "major" if it offers at least two separate SDKs in our data and had a market capitalization above 10 billion US dollars as of January 2016 (before the enactment of the GDPR). This definition identifies 23 major SDK providers (such as Verizon, Adobe, Amazon, Meta, Microsoft, Alibaba, Salesforce, IBM, Oracle, Tencent, etc), most of which have some business verticals that can be characterized as

digital platforms. We classify the other 1,250 non-Google SDK providers as "minor." We refer to the SDKs provided by major or minor SDK providers as major or minor SDKs, respectively. Although the count of unique major SDKs is smaller than that of minor SDKs, on average, a major SDK is more widely used than a minor SDK. Additionally, according to independent audits conducted by the Internet Safety Labs (ISL), the average SDK-specific safety risk score is significantly higher for Google and major SDKs than for minor SDKs, and the average developer-specific risk score is significantly higher for Google and major SDK providers than for minor SDK providers.

Using each app's monthly active users as weights in computing the average user exposure to SDK usage in each app category and country group, we do not find any significant change in the average user's SDK exposure in EU5-only apps. The only significant change in multihoming apps is a 4.2% drop in exposure to minor SDKs, all relative to US-only apps. More specifically, no significant change is found in the average user exposure to Google and major SDKs, although independent, external audits deem them as riskier than minor SDKs. This leads us to conclude that the GDPR has limited effects in curbing SDK usage and user exposure to SDK usage in Android mobile apps.

Our work contributes to the fast-growing literature on digital regulations in three main ways. First, we are among the first to call attention to SDK usage as an important mechanism in response to a landmark data regulation such as the GDPR, and offer one of the few studies that explores the role of SDKs in consumer privacy more broadly. Our emphasis on SDKs as part of the back-end data architecture complements studies on the many effects of the GDPR that researchers have identified across other dimensions. Our focus on the GDPR also supplements studies that explore the effects of market-driven mechanisms such as Apple's privacy labels or its App Tracking Transparency (ATT) initiative. Second, to our knowledge, we are among the first to link mobile apps to ISL's independent audits of users' data safety vis-à-vis SDKs and SDK providers. Although the audits only cover 19% of the SDKs in our app data, the nuanced characterization of SDK risk, including for Google and major SDKs, presents a unique opportunity for analyzing users' data exposure.[2] Furthermore, the lack of a discernible effect in the average user exposure to Google and major SDKs questions GDPR's effectiveness in reshaping the data infrastructure behind Android mobile apps. Third, we distinguish between EU5-only and multihoming apps, although both are subject to the GDPR. In doing so, we introduce another dimension to assess the effects of the GDPR.

The remainder of the paper is organized as follows. Section 2 provides some background

[2]See Forbes article by Jedidiah Yueh titled "GDPR will make big tech even bigger" (6/26/2018), available at `https://www.forbes.com/sites/forbestechcouncil/2018/06/26/gdpr-will-make-big-tech-even-bigger/`. It quotes Snap CEO Evan Spiegel: "There are times in history when regulation has actually entrenched big companies because they're the most capable of complying. I think that's a huge mistake, because I think that that would inhibit innovation."

context regarding SDKs and the GDPR. Section 3 describes the potential effects of the GDPR on SDK usage in mobile apps, and articulates how our work relates to the relevant literatures in economics, law, and computer science. Section 4 describes the data, and Section 5 presents the empirical results. Section 6 discusses the policy implications of our findings.

2 Background

2.1 SDK

Software Development Kits (SDKs) provide tools, libraries, and documentation that developers use to create software applications for specific platforms or systems. In the context of mobile apps, SDKs are commonly used to streamline the development process and provide access to various features and functionalities. While app developers can develop their own SDKs for repeated in-house use, most SDKs adopted by mobile apps are provided by third-party SDK providers. Unless otherwise specified, the SDK usage discussed in this paper refers to third-party SDKs.

Third-party SDKs can be particularly useful to app developers. They may include tools for performance monitoring, profiling, and optimization, enabling developers to quickly identify and resolve performance bottlenecks in their apps. Additionally, they allow app developers to integrate third-party services into mobile apps, such as analytics, advertising, social media, payment gateways, push notifications, and navigation. For example, a developer might integrate the Google Maps SDK to add mapping functionalities to their apps; use the Facebook SDK to enable users to log in with their Facebook credentials; incorporate the Amazon Web Services (AWS) Mobile SDK to enable AWS cloud features; or use the X (formerly Twitter) SDK to embed X content and features.

Furthermore, SDKs developed by the operating system platforms that host mobile app stores provide access to platform-specific features and functionalities. For instance, iOS and Android SDKs provide application programming interfaces (APIs) for accessing device hardware such as the camera, microphone, GPS, and gyroscopic sensors. Developers can leverage these SDKs to enhance app capabilities and improve the user experience. Some SDKs also allow developers to build cross-platform mobile apps using a single codebase, providing frameworks and libraries for writing code once and deploying it across multiple platforms, including iOS and Android. Additionally, some SDKs offer security features such as encryption, authentication, and authorization mechanisms, which help protect user data and prevent unauthorized access to sensitive information. Thus, greater usage of third-party SDKs does not necessarily imply a higher risk of privacy violations or data misuse.

SDK providers vary in size. Besides Google (Alphabet) and Apple, some large SDK providers (such as Amazon, Meta, and X) encompass digital platforms of their own, serving users on multiple sides (e.g., shoppers, gamers, content providers, developers, advertisers,

merchants) and facilitate their interactions in and across mobile apps. We define an SDK provider as "major" if it offers at least two separate SDKs in our data and has a market capitalization above 10 billion US dollars as of January 2016 (before the enactment of the GDPR).[3] We refer to the SDKs availed by these "major" SDK providers as "major SDKs" and those offered by other SDK providers as "minor SDKs."

It is common to observe SDK providers competing, offering the same or similar functionality. Take advertising SDKs as an example: Google AdMob enables developers to display various ad formats and offers tools for ad mediation to maximize revenue; the Facebook Audience Network enables developers to display Facebook's ad formats in their apps and thus make use of Facebook's (Meta's) extensive ad network and targeting capabilities; Chartboost specializes in mobile gaming advertising, enabling developers to promote their apps or integrate ads from other developers and advertisers; Unity Ads provides a way to monetize mobile games through different forms of advertising; IronSource similarly supports a variety of ad formats, including rewarded videos (ads that users can choose to view in exchange for an in-app reward), interstitials (full-screen ads that cover the interface of their host app, typically during a transition), and offerwalls (ads that help developers reward user engagement); and MoPub (by Twitter or X) enables developers to manage multiple ad networks and demand-side sources for optimal ad monetization.

Depending on the data sharing agreement between app developers and SDK providers, the SDK usage of a mobile app may enable the SDK provider to access and process the data related to or generated by the app user. This can enhance the SDK provider's ability to better profile individual users, to expand their own network of direct and indirect users, to improve targeted advertising, and to better monetize their expanded database.[4] These gains to app developers explain why SDK providers are willing to offer SDKs and SDK updates often free of charge, especially if they are digital platforms themselves that heavily depend on the volume, variety, velocity, and accuracy of big data. At times, app developers may adjust the extent to which data is shared by changing some parameters when they incorporate an SDK into their app, but what parameters can be changed and to what extent are determined by the SDK provider. The extent of such data sharing may also influence app developers' choice of which SDKs to utilize.

Third-party SDKs also enable data flows from SDK providers to app developers. For instance, when an individual utilizes Facebook login on an app, that user's name, Facebook

[3]The full list of major SDK providers includes Adobe, Verizon, Meta, Microsoft, Alibaba, Amazon, Salesforce, IBM, Baidu, Tencent, Oracle, X (formerly Twitter), eBay, Samsung, Visa, Nielsen, Rakuten, Kakao Corp, SAP, Softbank, Intel, VMware, and Nokia.

[4]For example, a blog on Meta's official website, dated April 16, 2018, with the title 'Hard Questions: What Data Does Facebook Collect When I'm Not Using Facebook, and Why?" states that "Apps and websites that use our services, such as the Like button or Facebook Analytics, send us information to make their content and ads better" and "When you visit a site or app that uses our services, we receive information even if you're logged out or don't have a Facebook account." Source: `https://about.fb.com/news/2018/04/data-off-facebook/`.

username, user ID (account number), Facebook profile pictures, networks she has joined on Facebook, and any other information she has made public on her Facebook profile are potentially visible to that app's developers.[5] The data sharing from Google sign-in to apps is more limited to the user's name, email address, and profile picture.[6]

In short, the extent to which app developers and SDK providers share data with each other and how they utilize such data depends on their data sharing agreements and each entity's privacy policies and data practices. Data sharing agreements are hardly observable to individual app users; privacy policies, in theory, are observable, but it is unclear how many users understand these details and what controls they have on these data flows when they click yes on a consent form. Some platforms (such as Facebook) do allow users to edit the types of information that apps on their platform can observe. Little is known as to how many users are aware of these options and how they use them. This relative opaqueness suggests that it is ultimately important to understand SDK usage by mobile apps, especially in light of landmark data regulations such as the GDPR.

2.2 Data Regulations in the EU and US

The General Data Protection Regulation (GDPR) is a comprehensive data protection law enacted by the European Union (EU) in April 2016 to regulate the processing of personal data. It applies to organizations, regardless of their location, that collect or process personal data of individuals within the EU. It took effect on May 25, 2018 and is enforced by each of the 27 member states of the EU.

We can summarize the key components ('articles') of the GDPR by front-end and back-end mandates. On the front end, facing individual users, the GDPR, with some limited exceptions, requires clear and explicit consent from individuals for the processing of their personal data. It grants individuals various rights over their personal data, including the right to access, rectify, erase, and restrict processing of their data. Users' 'opt-in' consent must be a freely given, specific, informed, and unambiguously chosen, meaning that silence, pre-ticked boxes, or inactivity should not constitute valid consent.

On the back end, Article 25 of the GDPR emphasizes privacy-by-design. It requires any organization that processes personal data from EU individuals to implement appropriate technical and organizational measures to integrate data protection principles into the design of their processing activities. This includes measures such as pseudonymization, data minimization, and ensuring that only necessary personal data are processed for each specific purpose. It also requires organizations to take proactive approaches to data protection, considering privacy implications at every stage of the development life cycle. Organizations are required to implement appropriate security measures to protect personal data against unauthorized

[5]Source: Facebook Help Center, `https://www.facebook.com/help/187333441316612`.
[6]Source: Google Account Help, `https://support.google.com/accounts/answer/12921417`.

access, disclosure, alteration, and destruction, and conduct privacy impact assessments to identify and mitigate privacy risks associated with their processing activities.

While the GDPR specifies the principles for data and privacy protection, it does not provide technical details on how to implement them at the back end of an organization's data processing infrastructure. The only exception is that organizations involving regular and systematic monitoring of data subjects on a large scale must appoint a Data Protection Officer (DPO), but how to define the responsibility of a DPO and how to integrate the role of a DPO with product development and other data related functions is still up to each organization.

Other GDPR articles include that (i) organizations must report data breaches to relevant authorities within 72 hours of becoming aware of them, (ii) transferring personal data outside the EU is prohibited unless adequate safeguards are in place, and (iii) non-compliance with the GDPR can result in significant fines of up to €20 million or 4% of global annual turnover, whichever is higher.

In contrast to the GDPR's comprehensive approach in Europe, privacy and data regulations are much more fragmented in the US. There is no federal law devoted to consumer privacy and data protection across all industries, though the Federal Trade Commission (FTC) has pursued privacy or data security cases through deceptive or unfair practices under the jurisdiction of the 1915 FTC Act. That being said, some industry-specific laws do exist; for example, the Health Insurance Portability and Accountability Act (HIPAA) applies to organizations that process individual health data, and the Gramm-Leach-Bliley Act (GLB) applies to financial institutions that handle individuals' financial data. There are also laws specific to demographic groups at the federal level. For example, the Children's Online Privacy Protection Act (COPPA) applies to organizations that process data of children under the age of 13, enforced by the FTC. There are also state-level privacy and/or data breach notification laws specific to organizations that serve users in the state.

A notable state law for privacy and data protection in the US is the California Consumer Privacy Act (CCPA), which has some parallels to the GDPR but also includes significant curve-outs in at least two dimensions. First, the CCPA requires data-processing organizations to provide individual users an option to 'opt-out' of personal data collection. This means the organization could set the default as allowing the broadest definition of data sharing and then give individuals an option to change this default. A vast literature has shown that, in retirement account savings, insurance plans and many other settings, individual consumers are reluctant to change the default setting, regardless of what the default is, partly because they are inattentive or lack the knowledge and expertise to fully understand all choices. This implies that the 'opt-out' choice under the CCPA would be less effective than the 'opt-in' choice under the GDPR in ensuring consumer consent to data sharing is an unambiguously informed action. Furthermore, the CCPA only applies to businesses that

serve California residents and meet some criteria in size or their intensity of data-related activities. These criteria include having a gross annual revenue above 25 million US dollars, collecting/buying/sharing the data of more than 50,000 users, or having at least 50% of their revenues generated from selling user data. Most app developers are small startup companies, and therefore are unlikely to be bound by the CCPA even if they serve California residents.

Timewise, the CCPA went into effect on January 1, 2020, and its subsequent amendment (the California Privacy Rights Act, or CPRA) took effect on January 1, 2023. As detailed below, our data analysis ranges from March 2017, 14 months before the rollout of the GDPR, to July 2019, 14 months after the GDPR's rollout but prior to the effective dates of the CCPA and CPRA.

In short, during our sample period, the EU experienced the rollout of the omnibus GDPR, while the landscape of privacy and data regulations in the US remained fragmented. This background motivates us to compare SDK usage by mobile apps in the EU and US. More specifically, the cleanest comparison is between the apps that serve the EU5 only (treated) and the apps that serve the US only (control), because the former group is bounded by the GDPR by definition but the latter group is not. We can also compare the apps that multihome in both the EU5 and US with the US-only apps. In theory, the GDPR applies to the multihoming apps as well, but it is unclear whether multihoming apps change both the EU and US versions of the app simultaneously for GDPR compliance, or only make changes in the EU version of the app. As detailed in Section 4, our data cannot distinguish among different SDK usages in concurrent different versions of an app in a particular month; hence, the data may reflect a mixture of simultaneous and EU5-only changes.

2.3 Legal Complications

We are aware of at least three legal complications regarding mobile app developers and/or SDK providers.

First, until mere days before the GDPR rolled out, the extent to which mobile app developers and mobile operating system providers (primarily Google or Apple) would carry liability associated with GDPR non-compliance by mobile apps was unclear. This is particularly evident in the days and weeks immediately before GDPR's effective date (May 25, 2018), as major platforms, including Google and Apple, began to reveal the ways in which they were tightening their platforms and app stores with new data sharing, data portability, and data liability rules. For example, on January 25, 2018, SafeDK documented that more than half of mobile applications were not GDPR ready.[7] Apple reportedly removed apps that share location data on May 9, 2018[8] and updated its privacy terms on May 23, 2018.[9]

[7]See `https://www.mobilemarketer.com/news/study-55-of-apps-may-not-meet-gdpr-privacy-standards/515546/`.

[8]See `https://www.idownloadblog.com/2018/05/09/apple-removing-apps-location-data/`.

[9]See `https://techcrunch.com/2018/05/23/apple-introduces-new-privacy-portal-to-comply-with-g`

Google released new consent requirements to developers on May 24, 2018, leaving app developers about 9 hours to decide whether to stay on Google's ad platform in light of the GDPR rollout.[10] These last-minute notices made it clear that app developers are responsible for their own compliance with the GDPR, despite their heavy dependence on the app-hosting platforms. They also imply that app compliance with the GDPR is unlikely to occur months before the GDPR's rollout in May 2018, due to uncertainty with regard to platform policies concerning the GDPR.

The second complication is likely specific to multihoming apps, as they may entail some transfer of users' personal data from the EU to the US, which is subject to the GDPR's and other restrictions regarding cross-border data transfers. In July of 2016, the European Commission formally adopted the Privacy Shield Program, which provides a mechanism for US companies to self-certify their compliance with certain privacy principles when processing users' personal data from the EU. This program was invalidated by the Court of Justice of the European Union (CJEU) in July 2020, due to a case filed in 2016 that questioned the adequacy of Privacy Shield in protecting the privacy rights of EU residents. A replacement mechanism — the so-called Data Privacy Framework (DPF) — was adopted by the EU and US in July 2023. Like its predecessor, the DPF enables US companies to voluntarily self-certify for EU-US data transfers in compliance with the DPF (and, by implication, the GDPR). It is worth noting that participating in the DPF is only one of multiple ways of satisfying the GDPR's restrictions on cross-border data transfers. This implies that using SDKs developed by DPF participants may aid multihoming app developers in asserting their compliance with the GDPR in terms of EU-US data transfers, although using SDKs developed by non-participants of the DPF does not necessarily imply non-compliance with the GDPR.

Third, in both the EU and US, there remains significant legal ambiguity regarding which party is responsible for what when data collection, processing, and usage involve multiple entities. The GDPR distinguishes between the entity that determines the purposes and means of processing personal data (the so-called data controller) and the entity that processes personal data on behalf of the data controller (the so-called data processor). When the two are not the same entity, the data controller is responsible for ensuring that the processing of personal data complies with the GDPR's principles and requirements. This includes obtaining a legal basis for processing, providing transparent information to data subjects about how their data will be processed, and respecting data subjects' rights. In contrast, data processors have specific obligations under the GDPR, including implementing appropriate security measures to protect personal data, only processing personal data as instructed by the data controller, and assisting the data controller in meeting its GDPR obligations (such as data subject rights requests and data breach notifications). In the context of mobile apps

dpr/.

[10] See https://bit.ly/2ziUgJA.

utilizing third-party SDKs, it is conceivable that the mobile app serves as the data controller and the SDK provider as the data processor.

While this distinction appears clear in theory, it is more complicated in practice. For example, SDK providers may access and process individual data through mobile apps in ways that may not be clearly communicated to app users. It can be difficult to determine whether such practices occur at the instruction of the app developer or are adopted by the SDK provider independently of the developer.

We could not find any specific GDPR enforcement cases regarding mobile apps and SDK providers.[11] One case enforced by the Italian Data Protection Authority (DPA) in June 2023 involves an app developer as a data processor.[12] In particular, the Italian DPA fined Autostrade per l'Italia spa ('ASPI,' an operator of highway networks) €1 million for unlawfully processing the data of approximately 100,000 registered users of the toll reimbursement app 'Free to X.' The DPA found that ASPI held the position of the data controller rather than the data processor and violated the GDPR by not designating Free to X as a processor. This incorrect assignment of privacy roles resulted in the consent notice to users being incorrect.

In the US, while there is no omnibus federal data regulation that applies across all industries and consumer demographics, the FTC has pursued several cases against app developers and/or SDK providers for privacy violations and data misuse under existing regulations.

In June 2016, for example, the FTC settled with InMobi, a Singapore-based mobile advertising platform for app developers and advertisers.[13] The FTC alleged that InMobi deceptively tracked the locations of hundreds of millions of consumers, including children, without their knowledge or consent, for the purpose of serving them geo-targeted advertising. According to the complaint, InMobi misrepresented that its advertising software would only track consumers' locations when they opted in and in a manner consistent with their device's privacy settings; in reality, the complaint alleges, InMobi tracked locations regardless of whether the apps using its software obtained or requested users' consent.

In two other recent examples, the FTC settled in January 2024 with X-Mode/Outlogic and InMarket, both of which are SDK and app developers that collect, sell, and/or share location data from their own apps or third-party apps that utilize their SDKs.[14] The FTC alleges that X-Mode did not have any policies in place to remove sensitive locations from the raw location data it sold until May 2023 and did not implement reasonable or appropriate

[11] Using the keywords 'mobile app,' we found 15 cases on `https://www.enforcementtracker.com/`, some of which concern data collection, data sharing, and data security pertaining to mobile apps, although none of the case descriptions mention SDKs.

[12] See `https://www.enforcementtracker.com/ETid-2023`.

[13] See the FTC's press release at `https://www.ftc.gov/news-events/news/press-releases/2016/06/mobile-advertising-network-inmobi-settles-ftc-charges-it-tracked-hundreds-millions-consumers`.

[14] See the FTC's press releases at `https://www.ftc.gov/news-events/news/press-releases/2024/01/ftc-order-prohibits-data-broker-x-mode-social-outlogic-selling-sensitive-location-data` and `https://www.ftc.gov/news-events/news/press-releases/2024/05/ftc-finalizes-order-inmarket-prohibiting-it-selling-or-sharing-precise-location-data`.

safeguards against downstream uses of the precise location data it sold, placing consumers' sensitive PII at risk. The FTC also alleges that both X-Mode and InMarket failed to inform users regarding how their own apps as well as third-party apps that utilized their SDKs would use those users' location data.

In some cases, the FTC has targeted app developers rather than SDK providers for deceptive data practices. In May 2023, the FTC charged the developer of the fertility app Premom with deceiving users by sharing their sensitive personal information with third parties, including AppsFlyer and Google, without proper disclosure.[15] The Premom app, which is free to download and used by hundreds of thousands of people, allows users to track ovulation, periods, and other health information. It encourages users to provide details about their menstrual cycles, fertility, and pregnancy, and to import data from other apps such as Apple Health. The FTC alleged that Easy Healthcare, the operator of the Premom app, had repeatedly and deceptively promised users in its privacy policies that it would not share their health information with third parties without consent, and that any data collected would be anonymized and used only for analytics or advertising. However, the FTC claimed that Easy Healthcare failed to take adequate measures to address privacy risks associated with its use of third-party SDKs, and shared health information for advertising purposes without obtaining users' express consent.

In addition to government enforcement, there have also been legal actions taken by private parties. For example, in the class-action lawsuit Greenley vs. Kochava, SDK provider Kochava was accused of using its SDK to secretly collect user data and later selling it to third parties.[16]

Given that both app developers and SDK providers may have access to users' PII and may have data agreements unknown to the users, and given that the roles of data controller and data processor are not always straightforward to delineate, legal ambiguity persists regarding each entity's obligations. This ambiguity is arguably greater in the US, where the applicable regulation is not always clear. As we discuss in the proceeding section, these ambiguities may influence the economic incentives of mobile app developers concerning SDK usage.

3 Conceptual Framework and Related Literature

3.1 Basic Setup

Suppose an app developer is deciding whether to introduce a new functionality to the app. If the decision is affirmative, the functionality can either be developed in-house or incorporated through a third-party SDK that provides the functionality at no monetary cost.

[15] See the FTC's press release at `https://www.ftc.gov/news-events/news/press-releases/2023/05/ovulation-tracking-app-premom-will-be-barred-sharing-health-data-advertising-under-proposed-ftc`.

[16] The case is still ongoing; see, for example, `https://casetext.com/case/greenley-v-kochava-inc-1`.

Let us begin with the app's profits if it does not adopt the new functionality. In this baseline situation, we write the app's profits π_0 as the product of the user base size (N_0) and the extent to which the app monetize users on average (α_0):[17]

$$\pi_0 = \alpha_0 \cdot N_0. \tag{1}$$

We assume that if the app developer introduces the new functionality, the app's user base increases to N_1 and the app may have an extra way to monetize the user base, but only if app users are willing to share data. Denoting the probability of user consent to data sharing as ρ_{data} and the extra way to monetize each data-sharing user as α_1, we can write the app's updated profits as:

$$\pi_1 = (\alpha_0 + \alpha_1 \cdot \rho_{data}) \cdot N_1 - C_{develop}. \tag{2}$$

The app developer is willing to adopt the new functionality if the cost of incorporating it ($C_{develop}$) is lower than the benefits from the user base increase and the new ability to monetize based on user data ($C_{develop} < \alpha_0 \cdot (N_1 - N_0) + \alpha_1 \cdot \rho_{data} \cdot N_1$). Since using the third-party SDK can lower $C_{develop}$ to almost zero and even increase ρ_{data} and α_1 as compared to in-house development (for example, the Facebook login SDK can provide users with the convenience of more easily logging into the focal app, which may make it easier to share their public Facebook profile and facilitate ad targeting based on user data),[18] it is plausible that the app developer may prefer using the third-party SDK to in-house development.

Suppose now that the GDPR applies, which may lower the degree to which app users are willing to share data (from ρ_{data} to ρ_{data}^{Aft}), change the degree to which data can be legally monetized (from α_1 to α_1^{Aft}), impose a new cost of compliance (C_{comply}), and introduce a possibility of fines F should a GDPR violation occur and the app developer is found liable for the violation. Denoting the probability of GDPR violation as $\rho_{violate}$ and the probability of being liable as ρ_{liable}, we can rewrite the profit of adopting the new functionality after the GDPR as:

$$\pi_1^{Aft} = (\alpha_0 + \alpha_1^{Aft} \cdot \rho_{data}^{Aft}) \cdot N_1 - C_{develop} - C_{comply} - \rho_{violate} \cdot \rho_{liable} \cdot F. \tag{3}$$

3.2 Potential Effects of the GDPR on SDK Usage

Comparing Equation 3 with Equations 1 and 2, we observe that the GDPR may affect the app developer's SDK adoption decision in four ways:

First, ρ_{data}^{Aft} is likely lower than ρ_{data}, because the GDPR heightens awareness regarding

[17] We shall interpret α_0 as average profit per user in the baseline situation, including any network effect the app may benefit from by having N_0 users.

[18] Here we only focus on the benefits from these changes that would accrue to the app developer; in practice, these changes may also generate benefits to the SDK provider, e.g., depending on the data sharing agreement between the app developer and the SDK provider.

and the salience of privacy choices and/or it enables users to decline data sharing when using an app.[19] Alternatively, ρ_{data}^{Aft} may be lower because the GDPR inhibits developers' and SDK providers' abilities to legally process and use data to which they may have access. In either case, a lower ρ_{data}^{Aft} will reduce the benefits of the app incorporating the new functionality. If the GDPR makes incorporating the new functionality non-profitable, there is no need to adopt the SDK.

Second, since the SDK has no pecuniary cost, the cost of GDPR compliance (C_{comply}) is only applicable if the app developer prefers in-house development to the SDK. This extra cost post the GDPR's rollout should generate a negative incentive for in-house development and further encourage the app developer to utilize the SDK if she still prefers to adopt the new functionality.

Third, the GDPR introduces a possibility of being fined should the app developer be found liable for the GDPR violation related to the data collected through the new functionality. The presence of this term alone should discourage the app developer from adopting the new functionality, and therefore reduce the use of the SDK.

Fourth, in addition, the expected fine from GDPR enforcement may differ between in-house development of the functionality and using the SDK. More specifically, using the SDK may increase the risk of GDPR violation ($\rho_{violate}$) because the app developer does not have full control of the content of the SDK, but it may reduce the risk of being held liable (ρ_{liable}), as some liability may be borne by the SDK provider. As the two effects counteract each other, it is difficult to predict the net effect.

To summarize, the GDPR may discourage or encourage mobile apps to use third-party SDKs, depending on whether the disincentives originating from less data sharing, less data monetization, and GDPR compliance costs (lower ρ_{data}^{Aft}, lower α_1^{Aft}, and C_{comply}) dominate the potential benefits that SDKs may provide in terms of better compliance and less liability exposure from GDPR violations (lower C_{comply}, $\rho_{violate}$ and ρ_{liable}).

We highlight three variations that may shed more light on the mechanisms behind the GDPR's impact on SDK usage. The first is the extent to which the focal app functionality involves sensitive data from individual users. Government agencies, including the US Department of Health and Human Services[20] and the FTC,[21] have stated that individually identifiable health, location, and financial data, among others, are considered sensitive and require additional protections, because such data in the wrong hands could bring more severe harms to the data subjects.

To the extent that the GDPR has educated consumers to be more wary of sharing sensitive

[19] In theory, ρ_{data}^{Aft} can be higher than ρ_{data} if the GDPR increases user trust in how their data would be used; however, empirical studies have shown that at least some EU residents decline to opt in to data sharing post the GDPR.

[20] https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html

[21] https://www.ftc.gov/business-guidance/blog/2022/07/location-health-and-other-sensitive-information-ftc-committed-fully-enforcing-law-against-illegal

data, an app functionality involving sensitive data is likely to face a greater decline in data sharing (ρ_{data}^{Aft}), which would discourage the app developer from adopting the functionality via an SDK or in-house development. At the same time, sensitive functionalities may have greater monetization potential (α_1^{Aft}) than non-sensitive ones, but enforcement agencies may pay closer attention to violations related to sensitive data, implying a higher $\rho_{violate}$ or ρ_{liable} for sensitive functionalities. Some of these forces may offset each other; thus, whether the GDPR has a greater impact on the usage of SDKs that support sensitive functionalities remains an empirical question.

The second variation important for SDK usage is whether the SDK provider is a large entity. Arguably, larger SDK providers — in terms of market value, number of employees, revenues, and R&D expenditures — are often closely watched by regulatory agencies, and may thus have more resources and stronger incentives to comply with the GDPR (though they may also have greater resources to assert that their legal interpretation of the GDPR is compliant). From the perspective of app developers, they may have already been reliant on the largest SDK providers for core functionalities prior to the GDPR. The GDPR could motivate them to rely even more on these large SDK providers as opposed to smaller alternatives, possibly to limit and contain their liability exposure by using other functionalities from those same large SDK providers. This could result in a substantial decrease in ρ_{liable}, even if $\rho_{violate}$ is similar or somewhat higher for larger SDK providers. In that case, the substantial decrease in liability could result in a net increase in the usage of SDKs offered by the largest providers. Conversely, insofar as SDKs involving sensitive data, larger SDK providers could draw more attention from enforcement agencies, implying a higher $\rho_{violate}$ and not necessarily a lower ρ_{liable} for the app developer, which, in turn, could motivate them to shy away from larger SDK providers. As before, which scenario is more likely to hold is an empirical question.

The third variation is whether the app is US-only, EU5-only or multihoming between US and EU5. During our analysis period (March 2017 to July 2019), no federal or statewide data protection laws in the US are comparable to the GDPR in terms of industry and geographic coverage. But apps serving Europe—including both EU5-only and multihoming apps—were subject to some European privacy and data protection laws preceding the GDPR, and thus the effects we identify for the GDPR is the marginal change above and beyond these preceding laws in Europe. To the extent that multihoming apps are more aware of the regulatory difference between US and EU, and may devote more resources to compliance in anticipation of GDPR, the marginal effect of the GDPR could be smaller for multihoming apps than for EU-onlly apps. At the same time, it is also possible that, because multihoming apps are often larger in user base, are subject to additional regulations concerning cross-border data transfers, and may anticipate more scrutiny from GDPR enforcers, they may experience a larger marginal effect from the GDPR than EU5-only apps.

Table 1 summarizes the above considerations concerning the GDPR's potential effects on

Table 1: Potential Effects of GDPR on EU5-only and Multihoming Apps

Effects of GDPR	**Arguments or Predictions**
Potential negative effects	(a) Less data to monetize after GDPR (b) GDPR imposes fines on the party found responsible for violations (c) GDPR gives app developers a stronger incentive to tightly control the app's data practice
Potential positive effects	(a) More costly to develop in-house SDKs after GDPR (b) Third-party SDKs (often free) can better comply with GDPR (c) Using third-party SDKs allows app developers to shift or share risk with SDK providers
Differential effects	(a) Stronger negative effects on sensitive SDKs because GDPR raises consumer attention to PII-sensitive data, reduce app's ability to monetize PII, and would likely impose greater liability and fines upon mis-use of sensitive data (b) Weaker negative effects on sensitive SDKs because it is more effective to monetize sensitive data than non-sensitive data and GDPR highlights the necessity to monetize data in order to cover the heightened compliance costs (c) Stronger positive effect on Google and major SDKs because they have more resources and ability to comply with GDPR and can help app developers to share/shift GDPR-related risks (d) Stronger negative effect on Google and major SDKs because they attract more regulator attention and may face a higher risk of GDPR scrutiny (e) Stronger effects (positive or negative) on multihoming apps than on EU5-only apps because multihoming apps are larger, are subject to extra regulation on inter-continental data transfer, and face a higher risk of GDPR scrutiny (f) Weaker effects (positive or negative) on multihoming apps than on EU5-only apps because multihoming apps may anticipate GDPR and comply before GDPR's official rollout

EU5-only and multihoming apps relative to US-only apps.

3.3 Related Literature

We contribute to three strands of literature. First, back-end data processes and frameworks for mobile apps, including the functionalities proffered by SDKs, have been studied by computer scientists, economists, and researchers in other disciplines. Outside economics, two studies are particularly related. Kollnig et al. (2021) compare third-party trackers in nearly two million Android apps between 2017 and 2020. Using URLs in apps' '.dex' files rather than SDKs to define third-party tracking in mobile apps, they find little change in the presence of third-party tracking in apps, and a persistent concentration of tracking capabilities among a few large so-called 'gatekeeper' companies.

In two reports (ISL, 2022; ISL, 2023), Internet Safety Labs studies 1,722 educational apps recommended or required by a random sample of 663 schools in 50 states. They find that nearly all apps in their sample (96%) share children's personal information with third parties, 78% of the time with advertising and monetization entities. More specifically, 74.9%

of all apps in their sample included one or more 'Very High Risk SDKs,' likely to share data with 'High Risk entities.'[22] They also find that as far as their sample shows, Google dominates K-12 edtech in the US: 70% of all apps in their sample include Google SDKs and they observe 68% of apps in their sample sending data to Google; in comparison, 38% of apps included Apple SDKs and 36% of apps were observed sending data to Apple. Furthermore, the schools they study do not systematically provide technology notices, allow user consent, or ensure technology safety: only 45% provide a notice clearly listing all technology used by students; only 14% provide the ability for parents or students 18 years or older to consent to technology use; only 29% appeared to be vetting all technology used by students, and those that vet do not necessarily earn a higher safety score (defined by ISL) on the apps they use than schools without vetting, partly because schools with vetting recommend or require more apps on average.

To our best knowledge, we are among the first to use ISL's independent evaluation of SDKs and SDK providers for academic research. Though the ISL audits only cover 19% of the SDKs in our app data, the relative risk between Google/major and minor SDKs we observe runs against the argument that large digital platforms have more resources to comply with the GDPR and therefore may provide safer and more GDPR-compliant tools than other software providers. The lack of a discernible effect of the GDPR on the average user exposure to Google and major SDKs also casts doubts on the GDPR's effectiveness in reshaping the data infrastructure behind Android mobile apps.

The second literature strand relates to market-driven mechanisms that attempt to address data and privacy concerns in mobile apps. Cheyre et al. (2023) offer a study of SDKs with the purpose of assessing the impact of Apple's App Tracking Transparency (ATT) framework on the creation, updating and withdrawal of mobile apps. In doing so, they track SDK usage to examine how developers adapt specific functionalities in their products in response to ATT. Taking Android as the comparison group, they find that the number of available apps in the Apple App Store ecosystem quickly recovers after an initial drop following the introduction of ATT. Additionally, they find that the use of monetization and ad-mediation SDKs decreases after ATT while the use of authentication and payments SDKs increases.

Leveraging the ATT policy change by Apple, though not using SDKs, Li and Tsai (2022) find that the restriction that prohibits apps from using identifiers to track users across other apps leads to a 1.9% decrease in the average number of new downloads for incumbent apps and a 9.5% reduction in new app entries. Relatedly, Bian et al. (2023) study how consumers and investors react to the standardized disclosure of privacy labels that Apple required all iOS apps to provide since December 2020. They find that after the release of privacy labels, an average iOS app experiences 14% and 15% drops in weekly downloads and revenues, respectively, in comparison to its Android counterpart, and the effect is stronger for more

[22]These levels of risk are defined by Internet Safety Labs.

privacy-invasive and substitutable apps.

A more recent study by Aridor et al. (2024) shows that ATT has made targeted advertising difficult for Meta advertisers because conversion-optimized Meta advertisements depends critically depend on third-party data for targeting and measurement. In particular, those with a higher baseline dependence on Meta are found to suffer a 37% reduction in firm-wide revenue, and this is primarily driven primarily steeper declines at smaller firms in the acquisition of new customers versus retaining existing customers. Another study by Deisenroth et al. (2024) studies Meta advertisers as well, finding that the reduced digital ad effectiveness in the wake of ATT led to less investment in advertising, greater market concentration, and higher product prices.

Our focus on the GDPR supplements the above-mentioned studies of market-driven mechanisms. Unlike GDPR, Apple's privacy labels or ATT are not directly applicable to Android mobile apps. To the extent that Android mobile apps are subject to different and potentially less stringent quality checks than iOS apps before being made available in the respective app store, government regulations like the GDPR are more crucial and more relevant for the users and developers of Android mobile apps.

The third related literature strand finds various effects associated with the GDPR, though it does not cover SDK usage. On the one hand, studies show that the GDPR has reduced venture investment in technology startups that are based in the EU (Jia et al., 2021), reduced web traffic, e-commerce revenue and persistent cookies in Europe (Goldberg et al., 2024; Libert et al., 2018; Dabrowski et al., 2019; Congiu et al., 2022; Aridor et al., 2023), increased market concentration among web technology vendors (Johnson et al., 2023; Peukert et al., 2022), raised the search costs of GDPR-covered users (Zhao et al., 2023), and reduced consumer surplus from mobile apps due to app exits and fewer apps entering post the GDPR (Janßen et al., 2022). These studies suggest that EU member states may suffer from less entry, less innovation, and less market competition as a result of the GDPR.

On the other hand, some studies find the GDPR to have zero or offsetting effects. For example, Zhuo et al. (2021) find that the GDPR has had no effect on interconnection agreements between EU and out-of-EU network providers. Aridor et al. (2023) show that, although the opt-in requirement of the GDPR reduced the number of identifiable consumers, an online travel intermediary is able to better track remaining consumers and increase their average advertising value. Following 900+ news and media websites in the US and EU, Lefrere et al. (2022) find that, despite an initial reduction, visitor tracking among EU websites bounced back after a few months, and GDPR has had no significant effect on EU websites' traffic rankings, new content provision, or social media engagement with new content. They also find no difference in the monetization strategy or survival rate of content providers across the EU and the US. These findings suggest that the initial, negative effects of the GDPR may not harm the survival and development of digital innovations in the longer term to the

same extent.

This paper advances this literature strand by (i) highlighting the important role of SDKs in back-end data practices within Android mobile apps; (ii) documenting how the GDPR affects SDK usage across the EU and US, across major and minor SDKs; (iii) examining how the impact of the GDPR on the usage of major SDKs is realized through different types of SDK functions; and (iv) investigating how the GDPR may have differentially affected multihoming apps that serve both the US and the EU as compared to EU5-only apps.

4 Data Description

4.1 Data Sources and Sample Definition

We use two data sources. Our primary data source is Apptopia, a third-party data intelligence company that tracks mobile apps on the iOS and Google Play stores, through which we track mobile apps from the beginning of 2015 through June 1, 2021.

We focus on Android mobile apps in the Google Play store for several reasons: first, in terms of operating systems for mobile apps, Android has a much wider international reach than iOS. According to Statista,[23] Android's worldwide market share has grown from 1.49% at the beginning of 2009 to 70.11% by the end of 2023, whereas the market share of iOS has fluctuated downwards from 37.45% in 2009 to 29.19% in 2023. Android's penetration in the EU resembles this worldwide trend,[24] although iOS still holds a larger market share in the US as of 2023.[25] Second, the growth of Android apps has exceeded that of iOS apps worldwide. According to App Annie's State of Mobile Report (2022), Google Play accounted for 77% of all apps and games releases in 2021. Third, unlike iOS, Android is a more open system and requires app developers to develop their Android apps to be compatible with different Android device manufacturers. This complexity could make Android app developers rely more on third-party SDKs than on in-house SDKs. Related to this reason, Apple's iOS has a reputation as the more secure option due to its stated focus on privacy, security and the end-user experience, though recently Google has also implemented stronger security measures for Android. For example, Apple tends to spend a longer time than Google in reviewing new and updated apps before making them available in app stores, and has introduced quality and data control mechanisms (such as privacy labels and ATT) earlier than Google. Thus, on average, the iOS store is often perceived to have better quality in terms of technical performance (e.g. load speed and smoothness), privacy protection, and security against malware than Google

[23] `https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/.`

[24] `https://www.statista.com/statistics/639928/market-share-mobile-operating-systems-eu/.`

[25] The within-US market shares are 57.93% for iOS and 41.64% for Android according to explodingtopics.com. Source: `https://explodingtopics.com/blog/iphone-android-users.`

Play.[26] To the extent that the GDPR sets a minimum quality standard on the data and privacy features of mobile apps, it could be more important for Android mobile apps than for iOS apps. Because of these reasons, we believe SDK usage in the Android ecosystem is more relevant as far as the impact of the GDPR in the EU relative to the US.

Our analysis focuses on the data period from March 2017 to July 2019, since the Google Play Store had some major changes afterwards, which motivated Apptopia to change its collection of some variables and thus makes it more difficult to compare the data before and after July 2019. California did not roll out the CCPA until January 2020; a sample period that ends before January 2020 helps to keep the US-only apps as a valid control group. The COVID pandemic beginning in early 2020 has also created a significant regime shift in terms of mobile usage and app development (Jin, Liu, and Wagman, 2022), which is another reason to end the sample period before 2020. Given that the GDPR was rolled out on May 25, 2018, the sample period provides 14 months before and 14 months after the GDPR's rollout.

Apptopia tracks high-performance apps in the Android ecosystem. Once an app appears on a top ranking list of Google Play, Apptopia tracks its downloads, # of monthly active users, # of daily active users, and app star ratings by the user's country and app store, until the app's downloads drop to zero for at least six months. If an app is tracked, Apptopia also estimates the app's download revenue, ad revenue, and in-app purchase revenue by user country and app store.

Importantly, Apptopia also examines the back-end coding of Android mobile apps and tracks individual SDKs adopted by each mobile app in Apptopia's data, including the date on which the SDK was adopted and discontinued (if applicable) by each app.

We define an Android app i as using SDK j during a certain month m if the activation date of SDK j on app i is prior to m and the deactivation date is after m. We impute the activation date with the install date of the SDK if the activation date is missing, and we impute the deactivation date with the uninstall date, assuming that for SDKs that are missing an activation or deactivation date, the SDK was activated on the day it was installed and remains active on the app ever since activation until a deactivation date is specified. For apps that do not have any information on SDKs, we assume they do not use any SDKs.

Since EU membership has changed over time, we focus on five major European countries (UK, France, Germany, Spain and Italy, referred to as EU5). Apptopia tracks app performance by month and country; hence, an app is counted as operating in the EU at month m if it has any non-zero performance data in any EU5 countries at m, and as operating in the US if it has any non-zero performance data in the US at m. By this definition, an app may serve EU5 and US concurrently.

As the GDPR applies to any digital products and services that collect PII from EU

[26] See industry comparisons between Android and iOS apps by Androidauthority.com (4/20/2024, https://www.androidauthority.com/ios-vs-android-1068950/) and getastra.com (6/30/2024, https://www.getastra.com/blog/mobile/ios-vs-android-security/).

citizens, the apps that 'multihome' in the US and EU can be subject to the GDPR as well, though in theory they may use different back-end codes for different countries. To incorporate this complication, our analysis sample distinguishes US-only, EU5-only and multihoming apps.

In particular, we start with the apps that have ever been ranked by the Google Play Store as top 500 in the US or top 100 in any EU5 country from January 2015 to March 2017 (in any app category). We use top 100 rather than top 500 in EU5 because Google's ranking is by country-category-month and each EU5 country is much smaller than the US. Since the 100th app in the EU5 ranking is similar in downloads and monthly active users to the 500th app in the US ranking, this sample criterion helps to assure that apps serving US and EU5 are similar on the margin in our sample. By definition, these apps served the US or EU5 at the beginning of our sample period; hence, we can follow their SDK usage and performance data over time. One caveat to this sampling criterion is that it omits apps that entered the Android app market after March 2017; hence, our work does not speak as to how the GDPR affects the entry of new apps.[27]

Among the apps that satisfy the above criteria, we define an app as 'US-only' if it has served the US but never served the EU5 throughout our entire data sample (January 2015 to June 2021); an app as 'EU5-only' if it has served any part of EU5 but never served the US throughout our data sample (January 2015 to June 2021); and an app as 'multihoming' if the app has appeared in the top 500 US ranking and top 100 EU5 ranking in at least one month between January 2015 and March 2017. We drop all of the other apps that do not belong to any of these three groups.

In total, our final sample contains 8,527 US-only apps, 12,043 EU5-only apps, and 38,443 multihoming apps. As the GDPR applies to EU5-only and multihoming apps, we take them as two separate treatment groups and compare them with the control group of US-only apps. For all three groups, we follow their SDK usage and performance data monthly from March 2017 to July 2019. The full sample is unbalanced because some apps may have exited the mobile app market before July 2019[28], or because Apptopia's performance data has missing values in some of an app's monthly records from March 2017 to July 2019. The apps that have valid performance data throughout July 2019 constitute the 'balanced' subsample, which includes 1,643 US-only apps, 3,790 EU5-only apps, and 11,692 multihoming apps. Henceforth, we refer to the unbalanced sample as 'full sample (unbalanced)' and the balanced subsample as 'analysis sample (balanced).' We present summary statistics of these two samples in Section 4.3.

The counts of unique apps in the analysis sample (balanced) are significantly smaller than those in the full sample (unbalanced), mostly because mobile apps are susceptible to high

[27] See Janßen, Kesler, Kummer & Waldfogel (2022) for the impact of the GDPR on entry and exit of mobile apps.

[28] Apptopia stops tracking the performance of an app if the app demonstrates zero downloads and zero monthly active users for at least six months. We assume an app exited the market if Apptopia stops tracking its performance.

turnovers: Appotopia stops tracking 62.5% of the multihoming apps, 74.5% of the US-only apps, and 64.5% of the EU5-only apps in our full sample (unbalanced) before July 2019. The remaining gaps in app counts between the two samples are driven by the fact that Apptopia sometimes has missing values in the middle of two valid performance records of the same app-country. These missing values seem random across different app features and rankings, but we do not impute the missing entries to retain these apps in the analysis sample, out of the concerns that imputation by any arbitrary rules may undermine the quality of the balanced analysis sample.

Our second data source is Internet Safety Labs (ISL), an independent nonprofit 501(c)(3) product safety testing lab and research organization that focuses on software safety. On its homepage,[29] the ISL states that it is committed to not accepting any funding that would impact its ability to remain impartial and objective or compromise its ethical foundation.

In particular, the ISL Safe Software Specification (Version 1.1) measures SDK safety in two layers: first, the ISL Category Taxonomy classifies SDKs into various categories and subcategories. For example, the ISL defines Audio Video as an SDK category and "Audio Video – Analytics" and "Audio Video – Media Player" as two subcategories. Each category-subcategory tuple is assigned a *risk impact* score due to the sensitivity and/or volume of personal information that the SDKs in that tuple utilize. Each category-subcategory-tag triple is assigned a *risk likelihood* based on the type of data sharing performed by the SDKs in that triple (with a particular focus on the likelihood of data monetization as the highest likelihood). The product of the risk impact score and risk likelihood — after normalization — defines *Category Risk Score*, which ranges from -1 to -4, with -4 indicating the highest risk. In the second layer, the ISL's in-house company risk scoring rubric assesses the risk profile of the company or organization that develops the SDK, based on various privacy and risk factors associated with the SDK provider such as privacy practices, historical data breaches, and data handling practices. The resulting *Company Risk Score* ranges from -1 to -4, with more negative scores indicating higher risk. The final risk score for an SDK is computed by adding its Category Risk Score and Company Risk Score. If an SDK belongs to multiple combinations of categories and subcategories, the ISL computes the SDK's risk score in each combination and takes the riskiest of these combinations as the final SDK risk score.

For each SDK evaluated by the ISL, we observe a final SDK risk score ranging from -1 to -8; for each SDK provider included in the ISL dataset, we observe an SDK company risk score ranging from -1 to -4.

The ISL risk score data is subject to a few caveats. It is a cross section as of June 25, 2024, and hence, we cannot observe how risk scores change over time within a particular SDK or a particular SDK provider. Aiming to measure current SDK safety, the ISL data is more likely to cover up-to-date SDKs than SDKs that were developed in the past and have

[29] `https://internetsafetylabs.org/`

less active use as of 2024. For this reason, only 18% of SDKs and 19% of the SDK providers that we observe in the raw Apptopia data (January 2015 to June 2021) can be matched with the ISL risk score data. In Section 4.2, we report the average risk score of these matched SDKs and SDK providers.

4.2 Classification of SDKs

As previously indicated, app developers may have different incentives for adopting various types of SDKs, depending on factors such as the size of the SDK provider and whether the SDK targets a functionality that involves sensitive data.

Since Google provides the Android operating system (OS), and every mobile app on the Google Play Store must integrate with that OS, we categorize all 69 Google-developed SDKs as a separate group. It is challenging to determine whether an app uses Google SDKs out of necessity or convenience, but because Apptopia tracks apps and SDKs in both the Google Play and iOS App Stores, we find that 32 Google SDKs are used exclusively by Android apps, while the remaining 37 SDKs are used by both Android and iOS apps. Arguably, Google's Android-only SDKs are more likely to be associated with Android-specific functionalities compared to SDKs that Google provides for both Android and iOS systems. In our regression analysis, we will differentiate between Google Android-only and Google Android+iOS SDKs.

Among non-Google SDKs, we distinguish between major and minor SDK providers based on size, measured by both the number of unique SDKs developed by the provider and the overall company size. One complication in the data is that some SDK providers listed by Apptopia may belong to the same parent company, even though they appear with distinct names and URLs in the dataset. To account for such parent-child relationships, we utilize the Crunchbase database on venture investments, which identifies the parent company when a venture is or becomes a subsidiary. We use the cross-sectional parent-child relationships as of July 31st, 2019, to avoid shifts in the definition of major or minor SDK providers over time. For example, Google acquired the mobile app developer platform Fabric in 2017; thus, we treat Fabric as part of Google throughout the sample period and count every SDK provided by Fabric as a Google SDK.

After identifying and accounting for the parent companies of all SDK providers in our sample, we observe that the distribution of unique SDKs by provider is highly skewed: only 5 SDK (parent) companies provide 10 or more SDKs, several dozen provide 2–9 SDKs, and each of the remaining SDK providers contributes only a single SDK. We define major SDK providers as those offering 2 or more SDKs and having a market capitalization exceeding 10 billion US dollars as of January 2016. Under this definition, we identify 23 major SDK providers.

Table 2 provides the names of each major SDK provider along with the number of SDKs they provide in parentheses next to their name. In total, these 23 major SDK providers

Table 2: Definition of SDK Groups

Google (69)	32 Google SDKs for Anroid only (referred to as Google Android-Only) 37 Google SDKs available for both Android and iOS (referred to as Google Android+iOS)
Major (151)	Total 23 SDK providers with >= 2 SDKs and market value >= $10bn USD as of Jan 2016: Adobe (19), Verizon (17), Meta (15), Microsoft (15), Alibaba (9), Amazon (9), Salesforce (9), IBM (7), Baidu (6), Tencent (6), Oracle (5), X (formerly Twitter, 5), eBay (5), Samsung (4), Visa (4), Nielsen (2), Rakuten (2), Kakao Corp (2), SAP (2), SoftBank (2), Intel (2), VMware (2), Nokia (2)
Minor (1,411)	Total 1,250 SDK providers that are non-Google and non-Major: e.g. Block (15), PayU (9), Yandex (7), RhythmOne (6), Akamai (5), Mobvista (5), Snap (5), ...
Sensitive (677) exclude Google SDKs	SDK function categories as defined by Apptopia: social, ad_network, analytics, mobile_marketing, monetization, security, attribution, push_messaging, geo_location, mediation, video_ad
Non-sensitive (885) exclude Google SDKs	SDK function categories as defined by Apptopia: support, speech_recognition, multi_media, payment, recording_sound, dev_tool, push_notifications, media_player, game_dev_platform, subscription_analytics, user_acquisition, audio_recognition, ecommerce, messaging, video, support_reviews, CDN, dev_platform, multiplayer_game_dev, photo,cryptocurrency, reviews, ebook_reader, crash_reporting, speech_synthesis, game_dev, data_recovery, translator

Note: In parentheses is the number of SDKs that ever appear in the Apptopia SDK data and are developed by the company, by the group of SDK providers, or for the group of SDK functions. We classify push messaging as sensitive and push notification as non-sensitive based on the advice of industry experts, asserting that push messaging is displayed inside the app and tailored toward the user whereas push notification is displayed outside the app and is less user-specific.

offer 151 SDKs in the Appotopia dataset. In comparison, the other 1,250 non-Google and non-major SDK providers are grouped as 'Minor.' They contribute a total of 1,411 SDKs in the Apptopia dataset. Table 2 lists a few examples of minor SDK providers, some of which do provide multiple SDKs and the majority of them provide only one SDK. Among the 23 major SDK providers, we find that 11 (47.8%) appear on the list that the US Department of Commerce has maintained for voluntary self-certification of compliance with the Data Privacy Framework, a mechanism to satisfy the GDPR's restrictions on cross-border data transfer. In comparison, among the 1,250 minor SDK providers in our sample, only 126 (10.1%) appear on the Data Privacy Framework list. These patterns suggest that using SDKs developed by major SDK providers, relative to those developed by minor SDK providers, may offer an app more compliance coverage with respect to cross-border data transfers. That being said, major SDK providers are larger than minor SDK providers by definition and tend to operate as digital platforms that connect buyers, sellers, content providers, and advertisers in their main businesses. Thus, they may have incentives to collect granular user data from various channels outside their own platform, and one way to do so is encouraging mobile apps to adopt their SDKs.

Because the GDPR, with limited exceptions, entails that mobile apps generally must obtain user opt-in consent for data collection, users may be more reluctant to give away their PII if they believe an app may share or even sell such information to third parties for advertising or other revenue. After consulting with a seasoned app developer, we label

an SDK as 'sensitive' if its main function is to track, analyze, or link the user to some external parties for potential revenue. In comparison, non-sensitive SDK functions may include development tools, payment tools, and all others. Technically, SDK functions in some of the non-sensitive categories may also contain important PII that may raise user privacy concerns, but these categories may be perceived as necessary for an app's normal operation. For example, a user has to share some PII in order to complete a transaction inside an app, even if she is unsure how the app will store and use her PII beyond the focal transaction.

The bottom two rows of Table 2 list the SDK function categories (as defined by Apptopia) corresponding to the sensitive and non-sensitive SDKs. Excluding Google SDKs, we define in total 677 as sensitive SDKs and 885 as non-sensitive SDKs. Out of the 151 major SDKs, 73 (48%) are sensitive. This percentage is lower among minor SDKs (604 out of 1,411 or 42%).

Table 3 summarizes the average risk score of SDKs and SDK providers. Because the ISL risk score data is a cross section as of June 2024 and our Apptopia data covers January 2015 to June 2021, only 298 (18%) SDKs and 242 (19%) SDK providers included in Table 2 can be matched with the ISL data. More specifically, we match 15 (21.7%) of Google SDKs, 32 (21.2%) of major SDKs, 250 (17.8% of minor SDKs), 182 (26.9%) of sensitive SDKs, and 100 (11.3%) non-sensitive SDKs. Among these matched SDKs, the average SDK risk score is worse (more negative) for Google and major SDKs than for minor SDKs, and worse for sensitive SDKs than for non-sensitive SDKs. Interestingly, the average SDK risk score is even slightly worse for major SDKs than for Google SDKs. As shown in the bottom two rows of Table 3, the matching rate between Apptopia and ISL data is higher when conducted by the identity of SDK providers: obviously, both data cover Google as an SDK provider; among non-Google SDK providers, we match 15 (65.2%) of major developers, 226 (18.1%) of minor developers, 165 (28.9%) of those that develop any sensitive SDKs, and 82 (11.2%) of those that develop any non-sensitive SDKs. Like SDK risk scores, the average SDK provider risk score is worse for Google and major developers than for minor developers, and worse for those that develop sensitive SDKs than for those that develop non-sensitive SDKs.

One explanation for the worse risk scores of Google and major SDK providers is that they are more likely to provide SDKs in sensitive functionalities than minor SDK developers. Their wider coverage of SDK functionalities and higher market valuations may also attract more regulatory attention, which may generate more public information about their privacy practices, historical data breaches, and data handling practices that the ISL incorporates in evaluating company risk scores.

Our primary focus is on app utilization of third-party SDKs; however, a complication is that the dataset does not distinguish between in-house and third-party SDKs. Given the fact that most names of app developers and SDK providers in the dataset are simply URLs

Table 3: Summary Statistics of ISL Risk Score

	Google			Non-Google			
Dep. Var.	All Google (1)	Android -only (2)	Android +iOS (3)	Major (4)	Minor (5)	Sensitive (6)	Non-sensitive (7)
# matched SDKs	15	2	13	32	250	182	100
Avg SDK risk score	-6.20	-7.00	-6.07	-6.28	-5.96	-6.53	-5.02
# matched SDK providers	1	1	1	15	226	165	82
Avg developer risk score	-3.00	-3.00	-3.00	-3.05	-2.50	-2.66	-2.28

Note: This table reports summary statistics of all SDKs used in the full sample that are matched with the ISL risk score data.

without further details, it is difficult to systematically separate in-house and third-party SDKs. Instead, we look into the apps that were developed by Google or our 23 major SDK providers, and check the extent to which they use their own SDKs. In the raw data, we observe more than 30,000 apps under the name of a major SDK provider or Google. About 77% of Google apps use a Google developed SDKs, which is unsurprising given that Google runs the Android operating system. Among the 23 major SDK providers, Kakao Corp has the highest percent of apps that use their own SDKs (20.36%). This percentage is around 15% for another three major SDK providers (Tencent, Baidu and Meta), between 2% and 9% for 8 major SDK providers, above 0 but below 2% for 5 major SDK providers, and zero for 4 major SDK providers. This distribution suggests that the SDK usage we observe by other app developers would most likely be third-party SDKs rather than in-house developed SDKs. To minimize the coverage of in-house SDK usage in our analysis sample, we exclude all of the 253 apps offered by Google or any of our major SDK providers that would satisfy our sampling criteria as mentioned above.

4.3 Summary of SDK Usage

Table 4 summarizes our full sample (unbalanced) and analysis sample (balanced analysis) by SDK types. In total, the apps in the full sample have used 64 unique Google SDKs and 1,183 non-Google SDKs. These numbers are smaller than those reported in Table 2, because Table 4 focuses on SDK usage we can observed in the full sample from March 2017 to July 2019 rather than the raw Appotopia data from January 2015 to June 2021.

An average Android app uses 6.82 Google SDKs and 4.37 non-Google SDKs. Within Google SDKs, an average app uses 2.2 Google Android-Only SDKs and 4.42 Google Android+iOS SDKs. Among non-Google SDKs, the count of unique minor SDKs outnumbers major SDKs by eight times, and the average number of minor SDKs used by each app in the sample outnumbers that of major SDKs by more than two times. This suggests that, on average, the utilization rate of each major SDK is significantly higher than that of each minor SDK, though there are many more minor SDKs. This is not surprising since ma-

Table 4: Summary Statistics by SDK Type

		Non-Google SDKs			
	Google SDKs	Major	Minor	Sensitive	Non-sensitive
Full sample (unbalanced):					
# (%) SDKs	64	112	1071	512	671
	(5.13)	(8.98)	(85.87)	(43.28)	(56.72)
# SDKs used by each app	6.82	1.23	3.14	2.23	2.14
% apps using any SDK (0-100)	63.45	38.66	61.78	52.67	58.55
# SDKs used by game app	8.32	1.55	3.92	3.21	2.26
% game apps using any SDK (0-100)	67.90	46.87	66.79	51.66	57.68
# SDKs used by non-game app	6.11	1.08	2.77	1.76	2.09
% non-game apps using any SDK (0-100)	61.33	34.75	59.40	55.95	66.21
Analysis sample (balanced):					
# (%) SDKs	62	108	948	476	580
	(5.55)	(9.66)	(84.79)	(45.08)	(54.92)
# SDKs used by each app	7.69	1.48	3.60	2.55	2.54
% apps using SDK (0-100)	72.93	45.88	70.53	59.60	66.81
# SDKs used by game app	8.81	1.83	4.13	3.49	2.47
% game apps using any SDK (0-100)	74.38	53.17	72.61	63.85	64.42
# SDKs used by non-game app	7.15	1.31	3.34	2.08	2.57
% non-game apps using any SDK (0-100)	72.22	42.31	69.51	65.99	78.15

Summary statistics presented in the table include US-only, EU5-only and multihoming apps. The definitions of major/minor and sensitive/non-sensitive SDKs exclude all Google SDKs.

jor SDK providers are more reputable and resourceful, and their SDKs may be regarded as higher-quality in terms of ease of use, ease of integration, security, and regulatory compliance. In comparison, when we break down non-Google SDKs by sensitive and non-sensitive, an average app utilizes 2.23 sensitive SDKs and 2.14 non-sensitive SDKs. On the extensive margin, about 63.5% of apps use any Google SDKs, 38.7% use any major SDKs, 61.8% use any minor SDKs, 52.7% use any sensitive non-Google SDKs, and 58.6% use any non-sensitive non-Google SDKs. These numbers indicate widespread utilization of SDKs across all of the SDK types we enumerate.

Similar patterns apply to the analysis (balanced) sample. In fact, apps in this subsample use slightly more SDKs in every category than apps in the full sample. This is understandable because the apps in this subsample are more established in app age and remain active throughout our sample period.

Table 5 summarizes SDK usage for US-only, EU5-only, and multihoming apps separately. In the full (unbalanced) sample, most apps are multihoming. Within single-homing apps, we have more EU5-only apps than US-only apps, because many EU apps may be more specific to one or few EU5 countries due to language and cultural barriers while the US is a more integrated market. Likely for the same reason, an average US-only app has more monthly active users, enjoys a greater number of downloads, and uses slightly more SDKs than EU5-only and multihoming apps.

Only 29% of the apps in the full sample (unbalanced) enter the analysis sample (balanced),

Table 5: Summary Statistics by App Groups

	(1) EU5 only	(2) Multihoming	(3) US only
Panel A. Full sample (unbalanced)			
# apps	12,043	38,443	8,527
Monthly Active Users	24,060.03	39,510.29	83,646.28
Downloads	5,620.84	8,863.34	17,889.46
Age (in days as of May 2017)	780.97	821.40	809.86
# SDKs per App	11.01	10.93	13.10
# Game apps	1,256	16,331	748
Panel B. Analysis sample (balanced)			
# apps	3,790	11,692	1,643
Monthly Active Users (MAU)	38,357.77	67,104.67	177,349.39
Downloads	9,121.96	15,301.05	37,890.80
Age (in days as of May 2017)	796.98	838.05	865.48
# SDKs per App	12.64	12.39	15.90
# Game apps	321	5,258	48

as many apps may not survive the whole sample period. The attrition rate in US-only apps is greater than EU5-only and multihoming apps, likely because the US market is more dynamic and has more turnovers than the EU5 market. Because the balanced sample focuses on surviving apps, they tend to be slightly older in age, have significantly more monthly active users and downloads, and use more SDKs than the apps in the full sample (unbalanced) on average.

One important distinction between multihoming and singlehoming apps is that in our full sample, 42.48% of multihoming apps are in the Games category but only 10.43% of EU5-only apps and 8.77% of US-only apps are games. As shown in Table 4, game apps use significantly more SDKs in every SDK type (Google, major, minor, sensitive, non-sensitive), and this pattern stands out in both the full unbalanced sample and the analysis balanced sample. These differences are reasonable because game apps often involve players around the world, players that do not speak the same language can play together in the same game, and game players often have high demand for audio, video, and minimal time delay in screen reaction and concurrent updates. Subscription and in-app purchases are also common among popular game apps, which could make their monetization models different from those that rely on free access and third-party advertising.

Figure 1 plots the average number of SDKs used by US-only, EU5-only and multihoming apps from March 2017 to July 2019, for Google, major, minor, sensitive, and non-sensitive SDKs separately. The dotted vertical line represents the GDPR's rollout in May 2018. All plots are based on the full sample (unbalanced).

Three patterns stand out in Figure 1. First, for US-only and EU5-only apps, the average SDK usage kept growing for all five types of SDKs after the GDPR rollout, despite the GDPR's heuristics concerning data minimization. For multihoming apps, their SDK usage grew over time for major, minor and non-sensitive non-Google SDKs, but their use

of Google SDKs declined over time and their use of sensitive non-Google SDKs fluctuated. Second, the number of SDKs per US-only app grew faster than that of EU5-only and multihoming apps before the GDPR, which presents an empirical challenge to tease out the true effect of the GDPR. It is difficult to tell whether the seemingly growing gap between US-only and EU5-only or multihoming apps is driven by the GDPR's rollout or by their differential pre-treatment trends. As shown in Appendix Figure A1, limiting the sample to balanced data helps to mitigate the differential pre-treatment trends, but does not make them completely parallel. As detailed in Section 5.1, our regression analyses use synthetic difference-in-differences (SDID) to address this issue. Third, multihoming apps seem to follow different dynamic patterns from EU5-only apps, although both groups are subject to the GDPR. In particular, while the number of major SDKs used by multihoming apps demonstrates similar growth to EU5-only apps, the growth in multihoming apps' usage of other types of SDKs is much slower, especially for minor SDKs.

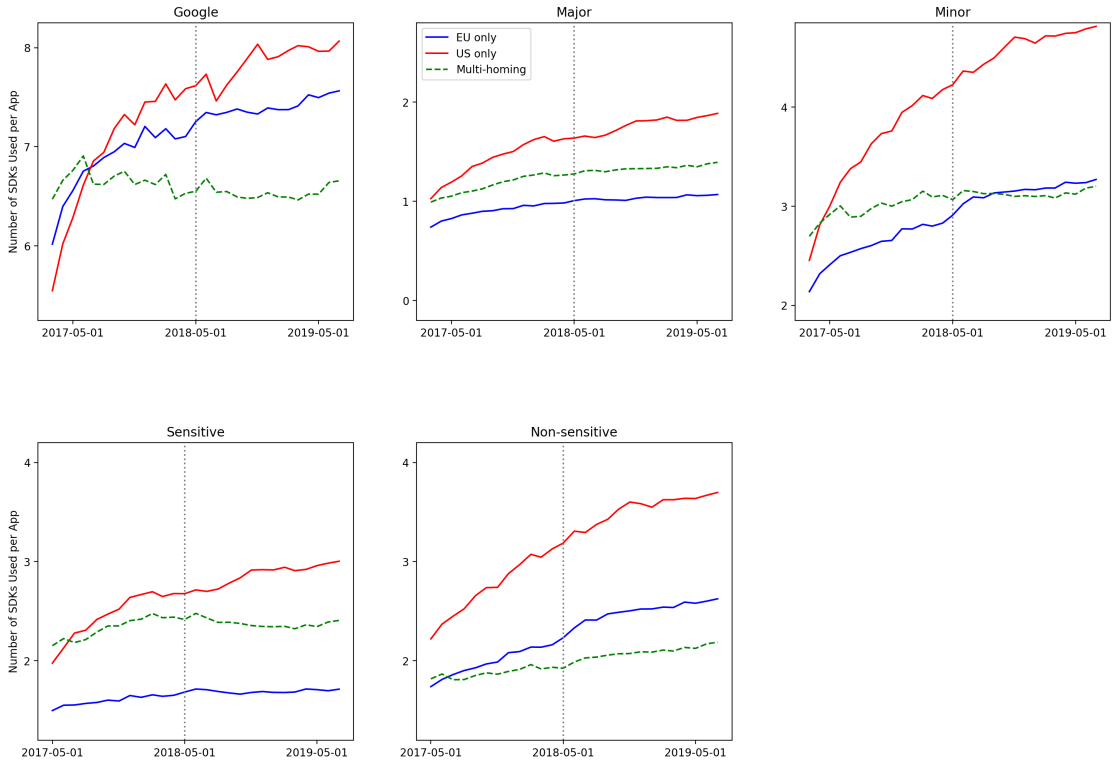
Lastly, in all app groups, we note that the cross-time variations in SDK usage were greater before GDPR than after GDPR. This can be explained by a natural, concave evolution path of mobile apps: while new SDKs enable mobile apps to introduce new functionalities, it gets harder and harder to develop new functionalities on top of what is already available in the market. When we track SDK usage within individual apps, we find that most cross-time variations in an app's SDK usage are driven by the app adding new SDKs rather than dropping old ones. These patterns suggest that it is important to have a control group that captures the general evolution of SDK usage, and, ideally, control for the evolution of different types of SDKs separately. In Section 5, we attempt to do so in the SDID regressions.

4.4 Correlation of SDK Usage and App Performance

Before we dive into the causal impact of the GDPR on SDK usage, one natural question is what kinds of apps are more likely to use third-party SDKs. As shown in Table 4, game and non-game apps tend to have different SDK usage, likely because they call for different sets of app functionality. We can address such cross-app heterogeneity by controlling for app fixed effects and focusing on SDK usage changes within each app.

In this subsection, we highlight the fact that, even within the same app, app performance is closely related to the app's SDK usage. According to our conceptual framework in Section 3, using third-party SDKs can help a mobile app enhance app functionality, attract users, and save on development costs. Though we do not observe app development costs, it is possible to detect the benefits of SDK usage in app performance data. Following this logic, we regress the performance of app i in month t ($Log(Perf_{it})$, defined as log of monthly active users or

Figure 1: Number of SDKs Per App



or log of downloads) on the number of SDKs that app i uses as of t ($NSDK_{it}$):

$$Log(Perf_{it}) = \delta_i + \delta_t + \sum_k \beta_k \cdot NSDK_{ikt} + \epsilon_{it}. \tag{4}$$

The regression controls for app fixed effects (δ_i) and year-month fixed effects (δ_t). We can interpret the coefficient β_k as the extent to which the performance changes within app i when i uses one more SDK of type k over time. Admittedly, β_k does not represent a causal impact of SDK usage on performance because there could be time-varying app attributes that correlate with both and it is difficult to rule out reverse causality (e.g. an app with a larger user base would have a stronger incentive to use SDKs to better monetize user data). Independent of the direction of causality, a positive and significant β_k would at least confirm a positive correlation between type-k SDK usage and performance improvement within the app.

Table 6 reports the estimated β_k for Google, major, minor, sensitive, and non-sensitive SDKs separately. The dependent variable is log(MAU+1) in the first row and log(downloads+1) in the second row. All regressions in this table use the full sample (unbalanced) throughout the entire sample period (March 2017 to July 2019). All but one estimate suggests a significant positive correlation between more SDK usage and better app performance. The only exception is the close-to-zero β_k for sensitive SDK usage and total downloads. Since

Table 6: Regression Results of App Performance on NSDK

		Non-Google			
Dep. Var.	Google (1)	Major (2)	Minor (3)	Sensitive (4)	Non-sensitive (5)
log(MAU+1)	0.012***	0.028***	0.025***	0.011***	0.051***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
N	992,458	992,458	992,458	992,458	992,458
log(downloads+1)	0.007***	0.006**	0.007***	-0.001	0.022***
	(0.001)	(0.003)	(0.001)	(0.001)	(0.002)
N	992,458	992,458	992,458	992,458	992,458

This table reports the coefficient of # of SDKs ($NSDK$ following Equation 4 (with app fixed effects and year-month fixed effects) and based on the full sample (unbalanced). In parentheses are standard errors. *** $p < 0.01$; ** $p < 0.05$, * $p < 0.1$.

we define sensitive SDKs as those that track, analyze, and/or link a user to some external entities for potential revenue generation, it is unsurprising that the usage of sensitive SDKs is strongly correlated with monthly active users but not total downloads.

Table 7 Panel A repeats the same exercise for the subsamples before and after the GDPR separately. The magnitude of the estimated β_k is more than doubled before the GDPR than after the GDPR, confirming the notion that the GDPR may discourage user willingness to share PII-related data (corresponding to $\rho_{data}^{Aft} < \rho_{data}$ in our conceptual framework), and thus reducing apps' ability to legally monetize user data ($\alpha_1^{Aft} < \alpha_1$).

Panel B of Table 7 reports the results for US-only apps, EU5-only apps, and multihoming apps separately. Every cell reports the estimate of β_k with standard errors in the parentheses. Across the three groups of mobile apps, β_k tends to be comparable between US-only and EU5-only apps, both of which are higher than the corresponding β_k of multihoming apps. The lesser sensitivity of app performance to SDK usage may suggest that multihoming apps expect less benefits from SDKs and are therefore more comfortable reducing their SDK utilization after the GDPR's rollout.

Table 7: Regression Results of App Performance on NSDK, by subsamples

	Dep. Var. = $log(MAU+1)$				
		Non-Google			
Subsample	Google (1)	Major (2)	Minor (3)	Sensitive (4)	Non-sensitive (5)
		Panel A: By Time			
Before GDPR	0.013***	0.038***	0.026***	0.019***	0.047***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
After GDPR	0.003***	0.015***	0.008***	0.004	0.021***
	(0.001)	(0.005)	(0.002)	(0.002)	(0.003)
		Panel B: By App Groups			
US only	0.009***	0.053***	0.035***	0.047***	0.037***
	(0.001)	(0.005)	(0.002)	(0.003)	(0.003)
EU only	0.013***	0.051***	0.039***	0.024***	0.059***
	(0.001)	(0.004)	(0.002)	(0.003)	(0.003)
Multihoming	0.009***	0.014***	0.011***	0.004***	0.030***
	(0.001)	(0.003)	(0.001)	(0.001)	(0.002)

This table reports the coefficient of # of SDKs ($NSDK$) following Equation 4 (with app fixed effects and year-month fixed effects). Each row corresponds to a subsample of the full sample (unbalanced). In parentheses are standard errors. *** $p < 0.01$; ** $p < 0.05$, * $p < 0.1$.

5 The Effects of GDPR Rollout on SDK Usage

We begin by presenting the econometric specification. We then report the results for EU5-only vs. US-only apps, followed by the results for multihoming vs. US-only apps.

5.1 Econometric Specification

To detect whether app i changes type-k SDK usage post the GDPR's rollout, we adopt a difference-in-differences (DID) specification:

$$SDKUSE_{ikt} = \gamma_{ik} + \gamma_{kt} + \theta_k \cdot GDPR_t \cdot Treated_i + \theta_x \cdot X_{it} + \epsilon_{ikt}, \tag{5}$$

where the dummy $GDPR_t$ equals 1 if the month of the observation is on or after May 2018, and the dummy $Treated_i$ equals 1 if app i is EU5-only or multihoming. The dependent variable $SDKUSE_{ikt}$ can be (a) the log of the number of type-k SDKs that app i uses in month t ($log(1+NSDK_{ikt})$), (b) a dummy of whether the app-month uses any type-k SDK in month t, and (c) the log of the number of unique providers of type-k SDKs used that app i uses in month t.

To reiterate, we separate Google and non-Google SDKs because Google is the host of the Android operating system. While it is difficult to measure whether an app uses Google

SDKs out of necessity, convenience, or something else, we distinguish Google Android-Only and Google Android+iOS SDKs depending on whether the SDK is Android only or available for iOS as well. For non-Google SDKs, we distinguish major, minor, sensitive, and non-sensitive SDKs, as our conceptual framework has laid out reasons that treated apps may change the usage of different types of SDKs differently after the GDPR.

We apply (5) to SDK usage per app in each SDK type separately. All regressions control for app fixed effects, year-month fixed effects, unemployment rates by country group and month, as well as GDP per capita by country group and quarter in case these time-varying macro conditions affect app development or SDK provision in general.[30]

An econometric challenge we face in estimating (5) is how to ensure the treated groups (EU5-only and multihoming apps) are comparable to the control group (US-only apps). As shown in Figure 1, not only do US-only apps use more SDKs than EU5-only apps, but they also experience faster growth in their number of SDKs prior to the GDPR's rollout. This implies that simply running (5) as a classical DID with two-way fixed effects may not properly identify the impact of the GDPR. To address this, we use the new method of Synthetic DID as proposed by Arkhangelsky et al. (2021). This method utilizes subjects in the control group to construct synthetic controls for subjects in the treated group, so that the synthetic controls (defined as a weighted average of control units) has an average outcome that is approximately parallel to the average outcome for the treated units before the treatment. Under the assumption that the treated and synthetic control units should continue to have parallel outcomes had the treatment not occurred, SDID can identify the causal effect of the GDPR's rollout on $SDKUSE$. Since SDID requires a balanced sample in the panel data, we run the SDID regressions for (5) with only the analysis sample (balanced). Another shortcoming of SDID is that it only applies to linear regressions. Thus, our SDID regressions focus on linear specifications, although some of the dependent variables may be a dummy variable (e.g., whether to use any SDKs) or count data (# of SDKs or # of SDK providers).

Applying (5) to the panel data at the *app* level allows us to look into each app's SDK adoption decisions, but it treats big and small apps equally. By definition, bigger apps involve more individual app users; even if two apps adopt exactly the same set of SDKs, the bigger app exposes more users to the SDK usage than the smaller app. To account for this difference, we construct a weighted average of SDK exposure per monthly active user in each app category c, month t, and SDK type k for each app group (EU5-only, US-only, multihoming) separately, while the weight of app i in category c is the app's monthly active users (MAU) of month t in the corresponding group of countries:

$$SDK\_Exposure_{ckt} = \sum_{i|ct} \frac{MAU_{ict}}{\sum_{i|ct} MAU_{ict}} \cdot SDKUSE_{ickt}. \tag{6}$$

[30]For EU5-only apps, unemployment rates and GDP per capita are averaged across the 5 EU countries for each month/quarter because we only have one observation per app-month for EU5.

We use the full (unblanced) sample to construct *SDK_Exposure*, because an app that has monthly active users but does not use SDK type k should still contribute to the total pool of monthly active users and an app that only appears in part of our sample period should still be included in the months during which it was active. We then apply the following specification to the data at the *app category* level:

$$SDK\_Exposure_{ckt} = \gamma_{ck} + \gamma_{ct} + \theta_k \cdot GDPR_t \cdot Treated_c + \theta_x \cdot X_{ct} + \epsilon_{ckt}, \tag{7}$$

The right-hand side variables are defined similarly to (5) except that we now control for app category fixed effects instead of app fixed effects. Again, we use SDID to ensure pre-treatment parallel trends between the treated groups (EU5-only, Multihoming) and the control group (US-only).

5.2 Results for EU5-only vs. US-only Apps

Following (5), Table 8 reports the SDID estimates of θ_k for all seven types of SDKs (Google Android+iOS, Google Android-Only, All non-Google, major, minor, sensitive and non-sensitive), where the dependent variable is $log(1+NSDK)$. As shown in Appendix Figure A2, the SDID framework ensures parallel pre-treatment trends between the treated (EU5-only) and control (US-only) groups, despite the two groups showing different pre-treatment trends in the raw data.

Since SDID requires a balanced sample, Panel A reports the results based on the analysis sample (balanced). Overall, EU5-only apps have no significant effect on the number of Google-SDKs but a marginal decrease in the number of non-Google SDKs after the GDPR's rollout relative to US-only apps. This small and marginally significant effect[31] (1.3%) is driven more by a statistically significant decline in the number of major SDKs than by the insignificant decline in the number of minor SDKs. Between sensitive and non-sensitive SDKs (all non-Google), the decline is significant for both and is of a similar magnitude.

In short, the small magnitude of θ_k on non-Google SDKs suggests that EU5-only apps decrease the number of non-Google SDKs by 1.3-1.6%, which corresponds to only 0.06-0.07 fewer SDKs. As the average number of non-Google SDKs per app is 4.37 in the sample, this implies that the GDPR's emphasis on data minimization — via consumers' ability to more readily not share PII-related data, apps' diminished ability to legally monetize user data, and compliance costs — may have slightly dominated the benefits of continuing to use as many non-Google SDKs as before the GDPR's rollout for an average EU5-app. The more significant results on major SDKs (as compared to minor SDKs) suggest a different tradeoff between the potential negative effects of the GDPR and the potential positive effects of

[31] Since the dependent variable is log (# SDKs+1), we compute the marginal effect of GDPR as $exp(= 0.013) - 1 = 1.3\%$.

leveraging major SDK providers for better cost-savings, risk sharing and GDPR compliance.

We further note that the similar effects on sensitive and non-sensitive SDKs is against the prediction that the GDPR would raise more consumer awareness on sharing sensitive PII and would have thus reduced more of the potential to collect, share, and monetize sensitive data. We do not observe such predicted differential effects, likely because sensitive SDKs are more crucial for the survival and financial performance of mobile apps than non-sensitive SDKs.

Since the SDID results in Table 8 capture the average treatment effect throughout the entire 14-month sample period post the GDPR's rollout, one may wonder whether the treatment effects vary over time. In an unreported table, we rerun the SDID regressions with the sample period ending in December 2018 and March 2019, as compared to the full sample period ending in July 2019. We find the estimates of θ_k only differ in the third decimal, suggesting that the GDPR's effect on the number of SDKs had already taken place within the initial seven months post the regulation's rollout. This is understandable given that Google had announced its GDPR compliance policy shortly before the regulation's official rollout and given that mobile apps can be updated within a few months (particularly with the aid of third-party SDKs).

To understand why the decline in the number of major SDKs is larger than the decline in the number of minor SDKs, whereas the decline in the number of sensitive and non-sensitive SDKs is similar, Table 9 reports the estimates of θ_k for a more detailed grouping of SDKs (major-sensitive, major-nonsensitive, minor-sensitive, minor-nonsensitive). These results suggest that the decline in non-Google SDK usage among EU5-only apps relative to US-only apps is largely driven by declines in SDK usage by major-sensitive apps and minor-nonsensitive apps. One potential explanation is that GDPR enforcers' potential keener interest in major-sensitive SDKs may make them less attractive to app developers, though app developers still put more trust in the quality of non-sensitive SDKs availed by major SDK providers.

The second and third panels of Table 8 further break the analysis sample (balanced) into a sub-panel comprised of US-only apps that have ever been ranked in the top 200 in the US (in any app category) or EU5-only apps that have ever been ranked top 50 in any of EU5 countries (in any app category; henceforth referred to as the 'top' panel) and the rest of the analysis sample (henceforth referred to as the 'bottom' panel). Appendix Table A1 provides sample statistics for the top and bottom panels. As expected, apps in the top panel are more established in age, monthly active users (MAU), number of downloads, and tend to use more SDKs than apps in the bottom panel on average.

The performance difference is especially stark: average MAU and downloads of apps in the top panel more than quadruple those in the bottom panel. Within the top panel, the average performance of US-only apps more than triple that of EU-only apps, although an

Table 8: Synthetic DID Results of log(NSDK) on GDPR: EU5-only vs. US-only

	Google SDKs		Non-Google SDKs				
Dep. Var.	Android+iOS (1)	Android-Only (2)	All (3)	Major (4)	Minor (5)	Sensitive (6)	Non-sensitive (7)
			Panel A. Analysis sample (balanced)				
log(#SDKs+1)	0.010	0.004	-0.013*	-0.016***	-0.011	-0.014**	-0.016**
	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)	(0.006)	(0.008)
N	157,209	157,209	157,209	157,209	157,209	157,209	157,209
			Panel B. Top: ever top 200 US vs. ever top 50 EU5				
log(#SDKs+1)	0.016**	0.007	-0.004	-0.011*	-0.001	-0.007	-0.008
	(0.007)	(0.006)	(0.009)	(0.005)	(0.010)	(0.008)	(0.008)
N	104,023	104,023	104,023	104,023	104,023	104,023	104,023
			Panel C. Bottom: the rest of top apps				
log(#SDKs +1)	-0.002	-0.003	-0.039***	-0.027**	-0.039***	-0.033***	-0.041***
	(0.013)	(0.014)	(0.014)	(0.011)	(0.014)	(0.012)	(0.014)
N	53,186	53,186	53,186	53,186	53,186	53,186	53,186

Notes: All synthetic DID regressions control for unemployment rate by country group and month, and GDP per capita by country group and quarter, in addition to app fixed effects and year-month fixed effects. Bootstrap standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 9: Synthetic DID Results of log(NSDK) on GDPR by detailed SDK groups, EU5-only vs. US-only

	Non-Google SDKs			
Dep. Var.	Major & sensitive (1)	Major & non-sensitive (2)	Minor & sensitive (3)	Minor & non-sensitive (4)
log(#SDKs +1)	-0.018***	0.009**	-0.000	-0.022**
	(0.004)	(0.004)	(0.005)	(0.009)
N	157209	157209	157209	157209

Notes: All regressions control for app fixed effects, unemployment rate by country group and month, and GDP per capita by country group and quarter. Bootstrap standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

app ranked the 200th in the US are closer to an app ranked the 50th in any EU5 countries in as far as total number of downloads and monthly active users. For example, in the balanced analysis sample, the average MAU of US-only apps ranking $180 \sim 200$ is 48,257.66, and average MAU of EU-only apps ranking $30 \sim 50$ is 22,497.14. While the MAU at the cutoff of US-only apps is still twice as large as that of EU-only apps, we choose top50 as the EU-only cutoff because we also want a large enough sample for EU-only top panel. Similar patterns hold for the bottom subsample of US-only and EU5-only apps.

When we apply (5) to the top and bottom panels separately, we find that the significant decline in overall non-Google SDK usage by EU5-only apps is completely driven by the apps in the bottom panel. In fact, EU5-only apps in the bottom panel demonstrate a nearly 4% drop in their number of non-Google SDKs, and this decline is reflected in major, minor, sensitive, and non-sensitive non-Google SDKs, but not in Google SDKs. In contrast, EU5-only apps in the top panel exhibit no decline in SDK usage significant at the 95% confidence across any of these non-Google SDK types. The only discernible changes are a 1.1% decline in major SDKs (significant at the 90% confidence) and a 1.6% increase in Google Android+iOS SDKs (significant at the 95% confidence). Though we do not have enough evidence to attribute these changes as apps in the top panel substituting away from major SDKs towards Google Android+iOS SDKs, such substitution is possible given the similarity between these two types of SDKs. Above all, Table 8 suggests that the negative effect of the GDPR's rollout on non-Google SDK usage is limited to lower-ranked EU5-only apps and does not apply to a core part of the user distribution of EU5-only apps.

Could the results in Table 8 be driven by the extensive margin because apps shy away from using *any* type-k SDKs? To answer this question, Table 10 reports the results analogous to Table 8 but redefines the dependent variable as a dummy equal to 1 if an app in our sample uses any SDKs in a specific SDK group. The results suggest that the extensive margin is positive (1.1%) for Google Android+iOS SDKs, but close to zero and insignificant for all non-Google SDKs as a group. Breaking non-Google SDKs into different types, we find that top-ranked EU5-only apps become less likely to use any major SDKs after the GDPR's rollout, and lower-ranked EU5-only apps become less likely to use any sensitive SDKs. Comparing the estimates in Table 8 and Table 10, we conclude that the small decline in non-Google SDK usage by EU5-only apps relative to US-only apps post the GDPR's rollout is less likely driven by the extensive margin (whether to use any SDK of type k) than the intensive margin (how many type-k SDKs to use if using one at all).

Another robustness check we apply is to focus on the number of unique SDK providers that a mobile app utilizes rather than the number of SDKs used in the app. One way to achieve GDPR's emphasis on data minimization is to reduce data flows and related security exposures and misuse risks by reducing the number of unique entities involved in processing an app's data. Following Equation 5, Table 11 reports the estimated θ_k when the dependent

Table 10: Synthetic DID Results of Any SDK on GDPR: EU5-only vs. US-only

	Google SDKs		Non-Google SDKs				
Dep. Var.	Android+iOS (1)	Android-Only (2)	All (3)	Major (4)	Minor (5)	Sensitive (6)	Non-sensitive (7)
			Panel A. Analysis sample (balanced)				
Using any SDKs	0.011***	0.003	0.001	-0.016***	0.003	-0.009*	0.005*
	(0.003)	(0.003)	(0.004)	(0.005)	(0.004)	(0.005)	(0.003)
N	157,209	157,209	157,209	157,209	157,209	157,209	157,209
			Panel B. Top: ever top 200 US vs. ever top 50 EU				
Using any SDKs	0.009**	0.003	0.003	-0.014***	0.003	-0.004	0.006
	(0.004)	(0.003)	(0.003)	(0.005)	(0.004)	(0.006)	(0.005)
N	104,023	104,023	104,023	104,023	104,023	104,023	104,023
			Panel C. Bottom: the rest of top apps				
Using any SDKs	0.014**	0.002	-0.005	-0.019*	-0.002	-0.025***	0.001
	(0.007)	(0.007)	(0.007)	(0.011)	(0.007)	(0.009)	(0.009)
N	53,186	53,186	53,186	53,186	53,186	53,186	53,186

Notes: All regressions control for app fixed effects, year-month fixed effects, unemployment rate by country group and month, and GDP per capita by country group and quarter. Bootstrap standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

variable is log (# of unique SDK providers +1). Since Google is the host of the Android operating system, this table focuses on non-Google SDKs only. The results are consistent with Table 8 where the dependent variable is $log(1 + NSDK)$: overall, the GDPR's rollout has a negative effect[32] (-5.8%) on the number of unique SDK providers used by EU5-only apps relative to US-only apps, and the effect is only marginally significant with 90% confidence. Broken down into major and minor SDKs, this negative effect is slightly smaller but with more statistical significance in the number of unique major SDK providers than in the number of unique minor SDK providers, likely because our classification includes many more minor SDK providers than major ones. Between sensitive and non-sensitive SDKs, we observe a significant negative effect in the number of non-sensitive SDK providers but close-to-zero effect in the number of sensitive SDK providers. Again, this is against the prediction that the GDPR should have a more negative effect on sensitive SDKs, probably because sensitive SDKs are more closely tied to the the financial health of mobile apps and thus more difficult to give up.

Following (7), we calculate the weighted average user exposure to SDK usage ($SDK\_Exposure$) at the app category level and run the SDID regression for each SDK type. As before, we have three measures of SDK usage — $log(1+NSDK)$, a dummy of whether using any SDK, and $log(1 + \#\ SDK\ providers)$ — for a particular type of SDK. Within app category c, the weights used in $SDK\_Exposure$ are each app's monthly active users in month t and the corresponding country group (EU5 or US).

[32]Because our dependent variable is log(# SDK providers+1), we compute the marginal effect of the GDPR as $exp(-0.06) - 1 = 5.8\%$.

Table 11: Synthetic DID Results of Number of SDK Providers on GDPR: EU5-only vs. US-only

	Non-Google SDKs				
Dep. Var.	All (1)	Major (2)	Minor (3)	Sensitive (4)	Non-sensitive (5)
log(# unique SDK providers +1)	-0.060*	-0.031***	-0.048*	-0.012	-0.059**
	(0.034)	(0.010)	(0.029)	(0.016)	(0.024)
N	157,209	157,209	157,209	157,209	157,209

Notes: All regressions control for app fixed effects, year-month fixed effects, unemployment rate by country group and month, and GDP per capita by country group and quarter. Bootstrap standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 12: Synthetic DID Results at the App-Category level (Weighted by MAU per App in the Full Sample): EU-only vs. US-only

Dep. Var.	All (1)	Google (2)	Major (3)	Minor (4)
MAU-Weighted log(#SDKs+1)	0.069	-0.026	0.033	0.039
	(0.074)	(0.056)	(0.031)	(0.029)
N	1612	1612	1612	1612
MAU-Weighted any SDK	-0.010	-0.013	0.007	-0.004
	(0.042)	(0.019)	(0.018)	(0.017)
N	1612	1612	1612	1612
MAU-Weighted log(# SDK providers + 1)	0.033	-0.009	0.013	0.026
	(0.047)	(0.012)	(0.026)	(0.022)
N	1612	1612	1612	1612

Notes: All regressions control for app-category fixed effects, year-month fixed effects, unemployment rate by country group and month, and GDP per capita by country group and quarter. Bootstrap standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 12 reports the results for all SDKs, Google SDKs, major SDKs, and minor SDKs in four columns. The three rows represent the three different SDK usage measures. Every number in the table is an estimated effect of the GDPR (θ_k) for the corresponding SDK type and SDK usage measure. None of the coefficients are statistically different from zero at the 90% confidence. This is consistent with our previous findings at the app level: because most of the negative effects on non-Google SDK usage are driven by lower-ranked EU5-only apps but top-ranked apps account for disproportionately more monthly active users, the overall effect of the GDPR on SDK_Exposure per active monthly user in an average app category is negligible.

5.3 Results for Multihoming vs. US-only Apps

In this subsection, we report the estimated effects of the GDPR on multihoming apps relative to US-only apps. We treat multihoming and EU5-only apps separately because they are fundamentally different: multihoming apps tend to be older, have much more active users and downloads, and are more likely to be game apps (Table 5). Although game apps tend to use more SDKs than non-game apps in general, multihoming apps on average use slightly fewer SDKs than US-only and EU5-only apps, mostly because the non-game multihoming apps use significantly fewer SDKs than the non-game singlehoming apps. For example, an average multihoming non-game app in analysis sample (balanced) uses 3.43 non-Google SDKs on average, whereas an EU-only non-game app uses 4.23. Moreover, by definition, multihoming apps have to address multiple languages or cultures across the US and Europe, and the GDPR's restrictions on cross-border data transfers are more readily applicable to multihoming apps than to EU5-only apps.

Following (5), at the app level, Table 13 reports the SDID-estimated θ_k for multihoming apps relative to US-only apps in our analysis sample (balanced). The results in Panel A suggest that multihoming apps reduced the number of all non-Google SDKs by 6.3%[33], the probability of using any non-Google SDK by 1.2%, and the number of unique # of SDK providers by 26.5%.[34] These effects is larger and more significant than that of EU5-only in Table 8, suggesting that multihoming and EU5-only apps face different benefit-cost calculations upon the GDPR's rollout. On the benefit side, as shown in Table 6, multihoming apps have a weaker correlation between SDK usage and app performance than EU5-only and US-only apps, which implies less benefits from SDK usage for multihoming apps. On the cost side, multihoming apps may have been more careful in data practice even before the GDPR rollout and are more likely to face an explicit restriction from the GDPR on cross-border data transfers. Given the on-and-off agreements between the US and EU on cross-border data transfers, multihoming apps may face greater compliance costs post the GDPR's rollout, which may further discourage their SDK usage.

Columns 2-5 of Table 13 further indicate that the significant decline in SDK usage by multihoming apps appear in all four types of non-Google SDKs. The decline in minor SDKs is larger than that in major SDKs, possibly because major SDKs are more likely to self-certify compliance with the Data Privacy Framework (DPR) and thus provide greater protections for app developers. The decline in SDK usage is smaller for sensitive SDKs than for non-sensitive SDKs, although app performance appears to correlate more with non-sensitive SDKs than with sensitive SDKs for multihoming apps (Table 6).

In panels B and C, we further break down the analysis sample into the top-ranked ones

[33]Because our dependent variable is $log(1+NSDK)$, we compute the GDPR's marginal effect as $exp(-0.065)-1=6.3\%$.

[34]We similarly calculate $exp(-0.308)-1=-0.265$.

(ever top 50 in any EU5 vs. ever top 200 in US) and the rest of the top apps. To save space, Panels B and C only report the SDID-estimated coefficient for the GDPR effect when the dependent variable is $log(1 + NSDK)$. Similar to the EU5-only vs. US-only comparison, we find the effects of the GDPR rollout are stronger for the lower-ranked apps than for the top-ranked apps, especially for minor SDKs, sensitive SDKs, and non-sensitive SDKs. Between sensitive and non-sensitive SDKs, the effects are either similar in magnitude or smaller for sensitive SDKs. Again, this is at odds with the prediction that GDPR should have more negative impact on sensitive SDKs.

Because multihoming apps are much more likely to have games apps than single-homing apps, we also run the analysis for games apps and non-games apps separately. It turns out that no SDID-estimated GDPR effect is significant for game apps. To save space, we do not report these results in Table 13. For non-game apps, we report the overall effects and the effects on the subsample of top-ranked apps and the rest of top apps in Panels D, E and F of Table 13. They confirm the previous finding that the effects of the GDPR rollout are more significant for lower-ranked apps than for top-ranked apps, especially for minor SDKs, sensitive SDKs and non-sensitive SDKs.

At the app category level, Table 14 reports the effects of the GDPR rollout on the MAU-weighted user exposure to SDK usage for multihoming apps versus US-only apps. Similar to that of EU5-only vs. US-only, we find most of the SDID estimated coefficients of the GDPR to be statistically close to zero. The only exception is a 4.1% decline on $log(1 + NSDK)$ on minor SDKs and a 1.7% decline on the dummy of using any minor SDKs. Overall, we conclude that, although the GDPR rollout has a significant negative effect on the SDK usage of multihoming apps at the app level, its impact on average user exposure to SDK usage is quite limited.

Table 13: Synthetic DID Results: Multihoming vs. US-only

	Google SDKs		Non-Google SDKs				
Dep. Var.	Android+iOS (1)	Android-Only (2)	All (3)	Major (4)	Minor (5)	Sensitive (6)	Non-sensitive (7)
			Panel A. Analysis Sample (balanced)				
log(#SDKs+1)	0.002	-0.006	-0.065***	-0.015***	-0.073***	-0.044***	-0.067***
	(0.006)	(0.006)	(0.007)	(0.005)	(0.006)	(0.005)	(0.007)
N	385,903	385,903	385,903	385,903	385,903	385,903	385,903
Using any SDKs	0.001	0.000	-0.012***	-0.020***	-0.014***	-0.025***	-0.019***
	(0.003)	(0.003)	(0.003)	(0.005)	(0.004)	(0.004)	(0.004)
N	385,903	385,903	385,903	385,903	385,903	385,903	385,903
log(# SDK providers +1)	0.001	0.000	-0.308***	-0.030***	-0.299***	-0.120***	-0.186***
	(0.003)	(0.003)	(0.038)	(0.008)	(0.032)	(0.016)	(0.024)
N	385,903	385,903	385,903	385,903	385,903	385,903	385,903
			Panel B. Top: ever top 200 US vs. ever top 50 EU				
log(#SDKs+1)	0.005	-0.002	-0.054***	-0.013**	-0.066***	-0.037***	-0.064***
	(0.007)	(0.005)	(0.007)	(0.006)	(0.009)	(0.005)	(0.009)
N	255,954	255,954	255,954	255,954	255,954	255,954	255,954
			Panel C. Bottom: the rest of top apps				
log(#SDKs+1)	-0.003	-0.015	-0.091***	-0.021**	-0.096***	-0.060***	-0.081***
	(0.012)	(0.012)	(0.015)	(0.010)	(0.016)	(0.009)	(0.018)
N	129,949	129,949	129,949	129,949	129,949	129,949	129,949
			Panel D. Analysis Sample of non-game Apps (balanced)				
log(#SDKs+1)	0.005	0.006	-0.059***	-0.020***	-0.062***	-0.050***	-0.050***
	(0.006)	(0.007)	(0.010)	(0.005)	(0.008)	(0.006)	(0.008)
N	232,029	232,029	232,029	232,029	232,029	232,029	232,029
			Panel E. Top non-game: ever top 200 US vs. ever top 50 EU				
log(#SDKs+1)	0.004	0.010*	-0.055***	-0.021***	-0.058***	-0.047***	-0.050***
	(0.007)	(0.006)	(0.008)	(0.007)	(0.008)	(0.006)	(0.009)
N	144,362	144,362	144,362	144,362	144,362	144,362	144362
			Panel F. Bottom non-game Apps: the rest of top non-game apps				
log(#SDKs+1)	0.005	-0.002	-0.078***	-0.024**	-0.083***	-0.062***	-0.064***
	(0.013)	(0.014)	(0.014)	(0.010)	(0.014)	(0.010)	(0.018)
N	87,667	87,667	87,667	87,667	87,667	87,667	87,667

Notes: All regressions control for app fixed effects, year-month fixed effects, unemployment rate by country group and month, and GDP per capita by country group and quarter. Bootstrap standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 14: Synthetic DID Results at the App-Category level (Weighted by MAU per App in the Full Sample): Multi-homing vs. US-only

Dep. Var.	All (1)	Google (2)	Major (3)	Minor (4)
MAU-weighted log(#SDKs+1)	-0.048	-0.015	-0.015	-0.042*
	(0.054)	(0.019)	(0.015)	(0.022)
N	1638	1638	1638	1638
MAU-weighted any SDK	-0.023	0.002	-0.013	-0.017**
	(0.021)	(0.008)	(0.010)	(0.008)
N	1638	1638	1638	1638
MAU-weighted log(# SDK providers + 1)	0.033	0.002	-0.000	0.031
	(0.028)	(0.002)	(0.020)	(0.023)
N	1664	1664	1664	1664

Notes: All regressions control for app-category fixed effects, year-month fixed effects, unemployment rate by country group and month, and GDP per capita by country group and quarter. Bootstrap standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

6 Conclusion

We demonstrate that SDK utilization is an important market layer where some of the effects of the GDPR manifest, in terms of the back-end design of Android mobile apps offered on the Google Play Store.

By comparing apps that are offered to US and EU5 users, we find that EU5-only apps use fewer non-Google SDKs relative to US-only apps after the GDPR's rollout, as the omnibus regulation may have intended in its advocacy for privacy and data minimization. This relative decline applies to sensitive and non-sensitive non-Google SDKs similarly, which is surprising as one may expect the GDPR to have raised more consumer awareness on sensitive PII and thus reduced more of the potential to collect, share, and monetize sensitive data. One potential explanation is that sensitive SDKs are more closely tied to the financial health of mobile apps, making them harder to forgo.

However, the relative decline of SDK usage among EU5-only apps is of a fairly small magnitude (1.3%) and is limited to lower-ranked apps. In contrast, multihoming apps that serve both the US and EU5 tend to reduce their non-Google SDK usage by 6.3% relative to US-only apps, a reduction that is highly statistically significant on average and across sensitive/non-sensitive or major/minor SDKs. Potential explanations include: (a) we observe a lower correlation between SDK usage and app performance for multihoming apps, which suggests less benefits of SDK usage; (b) the larger user base of multihoming apps may have attracted more regulatory scrutiny; and (c) the renewed focus on cross-border data transfer restrictions, amplified by components of the GDPR, may have increased compliance costs and therefore further discourage SDK utilization in multihoming apps. Like EU5-only apps,

the effect of the GDPR on multihoming apps is more driven by lower-ranked apps than top-ranked apps.

Given the substantial differences in the user bases of top-ranked and lower-ranked apps, we use apps' monthly active users as weights in computing average user exposure to SDKs in app category and country groups. At the app-category level, we find no significant change of the average SDK exposure in EU5-only apps relative to US-only apps. The only significant change of user exposure in multihoming apps is a 4.2% drop in exposure to SDKs developed by minor SDK providers. This is despite SDKs offered by Google and major SDK providers having been independently assessed as riskier than those from minor SDK providers. Overall, these patterns suggest that the GDPR has limited effects in curbing SDK usage in Android apps.

Our work is subject to several caveats. First, we are limited to the apps tracked by Apptopia from March 2017 to July 2019. By definition, these apps represent the 'head' of the app distribution, excluding those too small to be listed in Google Play's top 500 by March 2017 or apps that entered Google Play after that date. Since SDID requires a balanced sample, our app-level analysis also omits apps that were popular as of March 2017 but either exited the market or lacked valid performance data through July 2019. These sample restrictions limit our conclusions to existing and surviving apps. Second, without detailed knowledge of the specific data shared between app developers and SDK providers, or the extent to which such sharing complies with the GDPR, we cannot assess the welfare effects of SDK usage per app or SDK exposure per user. Nor can we articulate how the GDPR affects the concentration of SDK usage among providers, as the types and amounts of data shared through SDKs vary greatly across apps, app functionalities, and SDKs.

Future work can take on a number of directions. One such direction is examining the heterogeneous effects of the GDPR on the utilization of SDKs by apps belonging to different categories (e.g., gaming, education, health, finance). Another direction may associate SDK usage with the entry and exit of mobile apps in the wake of the GDPR, or with more explicit data or privacy-related outcomes, such as data breach incidents or lawsuits concerning privacy. An additional direction is a deeper dive into the pairwise connections between individual app developers and SDK providers, after linking their identity to the lines of business in which they engage, beyond mobile apps.

References

Aridor, Guy, Yeon-Koo Che, Brett Hollenbeck, Daniel McCarthy, and Maximilian Kaiser (2024). "Evaluating The Impact of Privacy Regulation on E-Commerce Firms: Evidence from Apple's App Tracking Transparency". In: *SSRN Working Paper*.

Aridor, Guy, Yeon-Koo Che, and Tobias Salz (2023). "The Effect of Privacy Regulation on the Data Industry: Empirical Evidence from GDPR". In: *RAND Journal of Economics* 54.4, pp. 695–730.

Bian, Bo, Xinchen Ma, and Huan Tang (2023). "The Supply and Demand for Data Privacy: Evidence from Mobile Apps". In: *SSRN Working Paper*. URL: https://drive.google.com/file/d/1oTJxM_vJOqgulu1wveFObsIQbLzlnx36/view.

Cheyre, Cristobal, Benjamin T. Leyden, Sagar Baviskar, and Alessandro Acquisti (May 2023). "The Impact of Apple's App Tracking Transparency Framework on the App Ecosystem". In: DOI: 10.2139/ssrn.4453463. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4453463 (visited on 11/07/2023).

Congiu, Raffaele, Lorien Sabatino, and Geza Sapi (2022). "The Impact of Privacy Regulation on Web Traffic: Evidence From the GDPR". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.4025033.

Dabrowski, Adrian, Georg Merzdovnik, Johanna Ullrich, Gerald Sendera, and Edgar Weippl (2019). "Measuring Cookies and Web Privacy in a Post-GDPR World". In: *Passive and Active Measurement* 11419, pp. 258–270. DOI: 10.1007/978-3-030-15986-3_17.

Deisenroth, Daniel, Utsav Manjeer, Zarak Sohail, Steve Tadelis, and Nils Wernerfelt (2024). "Digital Advertising and Market Structure: Implications for Privacy Rgulation". In: *NBER working paper #32726*.

Goldberg, Samuel, Garrett Johnson, and Scott Shriver (2024). "Regulating Privacy Online: An Economic Evaluation of the GDPR". In: *American Economic Journal: Economic Policy* 16.1, pp. 325–58. DOI: 10.2139/ssrn.3421731.

ISL (Dec. 2022). "K-12 EdTech Safety Benchmark: National Findings – Part 1". In: *Internet Safety Lab*. URL: https://internetsafetylabs.org/wp-content/uploads/2022/12/2022-k12-edtech-safety-benchmark-national-findings-part-1.pdf.

— (June 2023). "2022 K-12 EdTech Bechmark Findings Report 2: School Technology Practices & 3rd Party Certifications Analysis". In: *Internet Safety Lab*. URL: https://internetsafetylabs.org/wp-content/uploads/2023/06/2022-K12-Edtech-Safety-Benchmark-Findings-Report-2.pdf.

Janßen, Rebecca, Reinhold Kesler, Michael Kummer, and Joel Waldfogel (May 2022). "GDPR and the Lost Generation of Innovative Apps". In: *NBER Working Paper #30028*. DOI: 10.3386/w30028.

Jia, Jian, Ginger Zhe Jin, and Liad Wagman (Mar. 2021). "The Short-Run Effects of the General Data Protection Regulation on Technology Venture Investment". In: *Marketing Science* 40. DOI: 10.1287/mksc.2020.1271.

Jin, Ginger Z., Ziqiao Liu, and Liad Wagman (Dec. 2022). "Popular Mobile Apps in the Pandemic Era: A Glimpse of Privacy and Competition". In: *CPI Antitrust Chronicle*.

Johnson, Garrett (2024). "Economic Research on Privacy Regulation: Lessons from the GDPR and Beyond". In: *The Economics of Privacy, edited by Avi Goldfarb and Catherine Tucker, University of Chicago Press*. DOI: 10.2139/ssrn.4293618.

Johnson, Garrett A., Scott K. Shriver, and Samuel G. Goldberg (Mar. 2023). "Privacy and Market Concentration: Intended and Unintended Consequences of the GDPR". In: *Management Science* 69. DOI: 10.1287/mnsc.2023.4709.

Kollnig, Konrad et al. (Dec. 2021). "Before and After GDPR: Tracking in Mobile Apps". In: *Internet Policy Review* 10. DOI: 10.14763/2021.4.1611.

Lefrere, Vincent, Logan Warberg, Cristobal Cheyre, Veronica Marotta, and Alessandro Acquisti (2022). "Does Privacy Regulation Harm Content Providers? A Longitudinal Analysis of the Impact of the GDPR". In: *SSRN Working Paper*. DOI: 10.2139/ssrn.4239013.

Li, Ding and Hsin-Tien Tsai (2022). "Mobile Apps and Targeted Advertising: Competitive Effects of Data Sharing". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.4088166. (Visited on 03/14/2022).

Libert, Timothy, Lucas Graves, and Rasmus Nielsen (2018). *Changes in Third-Party Content on European News Websites after GDPR*. DOI: 10.60625/risj-r2ex-7248.

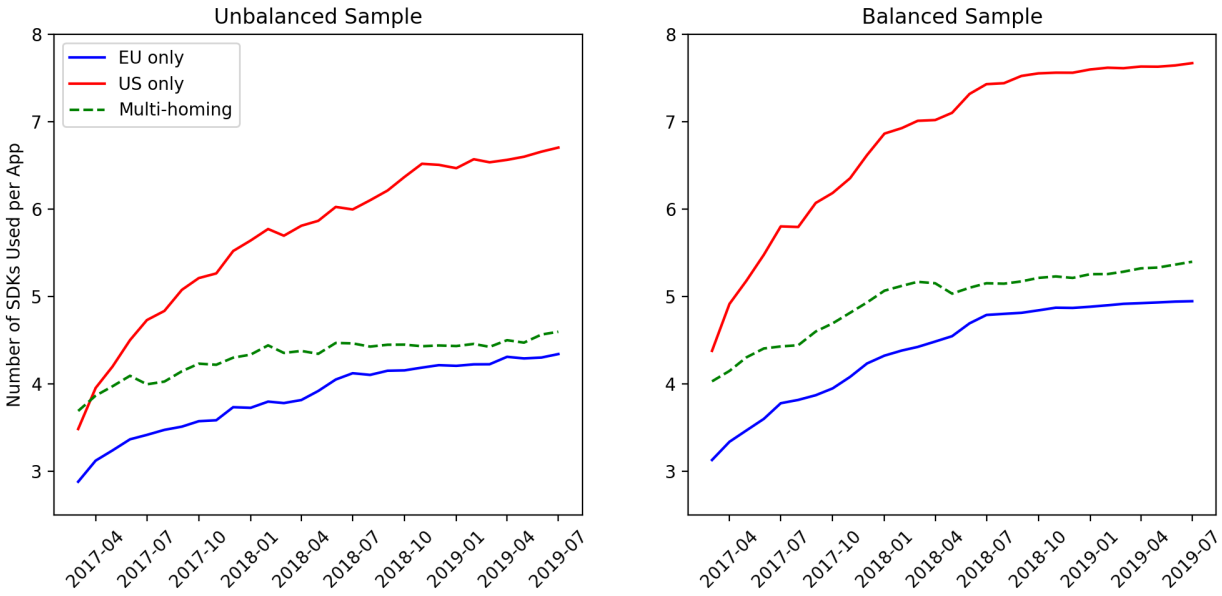
Peukert, Christian, Stefan Bechtold, Michail Batikas, and Tobias Kretschmer (Feb. 2022). "Regulatory Spillovers and Data Governance: Evidence from the GDPR". In: *Marketing Science* 41. DOI: 10.1287/mksc.2021.1339.

Zhao, Yu, Pinar Yildirim, and Pradeep K. Chintagunta (2023). "Privacy Regulations and Online Search Friction: Evidence from GDPR". In: *Marketing Science Institute Working Paper*. DOI: 10.2139/ssrn.3903599. (Visited on 10/06/2021).

Zhuo, Ran, Bradley Huffaker, kc claffy, and Shane Greenstein (Mar. 2021). "The Impact of the General Data Protection Regulation on Internet Interconnection". In: *Telecommunications Policy* 45, p. 102083. DOI: 10.1016/j.telpol.2020.102083. (Visited on 09/14/2021).

A Appendix

Appendix Figure A1: Number of SDKs Used per App over time:
compare the full sample (unbalanced) and the analysis sample (balanced)



Appendix Table A1: Summary Statistics of Apps, by Top/Bottom Panel

	Top		Bottom	
	(1) EU only	(2) US only	(3) EU only	(4) US only
Panel A. Full sample (unbalanced)				
# apps	6,442	4,006	5,601	4,521
By Genre:				
Art & Photography	27	15	24	29
Communication & Social	218	119	176	118
Education	494	129	465	251
Games & entertainment	1,706	836	1,121	1,025
Health & Lifestyle	1,598	843	1,468	1,043
Music	96	42	87	72
News	223	227	267	225
Productivity & Tools	2,080	1,795	1,993	1,758
Monthly Active Users	35,899.96	129,888.41	8,359.58	15,273.16
Downloads	8,396.44	27,417.24	1,940.23	3,801.79
Age (as of May 2017)	786.65	832.89	773.43	775.82
Panel B. Analysis sample (balanced)				
# apps	2,350	1,242	1,440	401
By Genre:				
Art & Photography	10	4	3	0
Communication & Social	73	25	30	13
Education	171	35	81	17
Games & entertainment	505	165	201	29
Health & Lifestyle	637	364	359	131
Music	37	14	26	2
News	87	117	114	49
Productivity & Tools	830	518	626	160
Monthly Active Users	54,231.54	222,013.22	12,452.65	39,014.05
Downloads	12,874.23	46,900.20	2,998.47	9,986.35
Age (as of May 2017)	799.93	895.66	792.17	772.03

Appendix Figure A2: SDID trend plots for Table