Do Privacy Policies Increase Data Sharing?*

Ozge Demirci[†] Eva Ascarza[‡] Ayelet Israeli[§]

March 20, 2025

Preliminary and incomplete. Please do not circulate.
Please click here for the latest version.

Abstract

This paper studies the impact of privacy policies on consumer data-sharing behavior, focusing on recent privacy policy changes in California and Virginia that took effect on January 1, 2023. Using data from a leading customer engagement app in the US, where users share shopping information by uploading receipts, we employ a difference-in-differences strategy to analyze how these policies influence the volume and variety of information shared. Our findings show a significant increase in receipt uploads in treated states compared to control states after the policy. Beyond the volume of data shared, we also observe an increase in the variety of information, including a broader range of store types, locations, and retailer categories. To further support these results, we provide additional evidence from the Consumer Expenditure Survey, where we find that respondents in treated states became less likely to express privacy concerns and were more likely to report their expenditures. We use Google Search Volume Index data to show heightened interest in privacy topics in treated states, suggesting greater privacy awareness among residents. These findings suggest that privacy regulations can drive consumers to share more data and engage more actively with digital platforms by bringing more transparency and awareness about consumer privacy.

Keywords: Privacy, Data Sharing, Digital Platforms

*We thank the anonymous company for providing the data and Noah Ahmadi for data preparation.
[†]Imperial College London Business School, odemirci@ic.ac.uk
[‡]Harvard Business School, eascarza@hbs.edu
[§]Harvard Business School, aisraeli@hbs.edu

1 Introduction

The rise of digital platforms has brought significant privacy concerns, as vast amounts of consumer data are continuously collected, processed, and utilized for various business strategies. This issue has become particularly urgent in recent years, with growing public awareness around data misuse, unauthorized access, and inadequate transparency in data handling practices. Privacy policies have emerged as critical regulatory tools aimed at protecting consumer rights, but questions remain regarding how these policies influence consumer behavior itself, particularly their willingness to share personal data. Understanding consumer responses to privacy regulations is essential, as data-sharing behavior has implications not only for individual privacy but also for businesses and customer engagement in digital ecosystems.

This paper examines the impact of two recent privacy policy changes in the U.S., the California Privacy Rights Act and the Virginia Consumer Data Protection Act, on consumers' willingness to share data. These regulations introduce enhanced consumer rights and stricter data protection requirements, empowering individuals to exercise greater control over their personal information. We analyze the effects of these policies within the context of a leading customer engagement app where users upload their shopping receipts in exchange for points and monetary rewards. This unique setting provides an opportunity to observe granular changes in consumer data-sharing behavior in response to policy changes, allowing us to track user engagement with the platform over time. Leveraging a difference-in-differences approach, we estimate the causal effects of these policies, which became effective as of January 1st, 2023.

Our findings indicate a significant 27.1 percent increase in receipt uploads in treated states relative to control states, suggesting that users in states with stronger privacy protections are not only sharing more data but also a wider variety of information. This increase in variety in information includes receipts from diverse types of stores, geographic locations, and retailer categories, indicating that privacy policies may encourage users to engage more openly and broadly on the platform. The observed changes can imply that stronger privacy protections enhance consumer trust, reducing reluctance to share personal information and highlighting the role of privacy policies in encouraging greater data-sharing engagement.

To further support these findings, we analyze data from the Consumer Expenditure Survey, which captures consumer behavior in a broader setting. We find that respondents in treated states became less likely to express privacy concerns during survey interviews and were more likely to report their expenditures. These results align with the increased engagement observed on the platform, reinforcing the idea that privacy policies can influence

consumer perceptions of data security and their willingness to disclose information.

In addition to changes in data-sharing behaviors, we also find evidence of increased interest in privacy topics within the treated states, using data from Google's Search Volume Index (SVI) for terms related to "privacy." Our analysis reveals that privacy-related searches are notably higher in treated states compared to control states, with California displaying the most pronounced interest. These heightened search volumes align with and reinforce our main findings, suggesting that stronger privacy protections are associated with a greater public awareness of privacy issues. This increased awareness may contribute to consumers' heightened sensitivity to data-sharing practices on digital platforms, underscoring the impact of privacy regulations on both behavior and perceptions of sharing individual data within digital platforms.

This study makes several contributions to the literature on the economics of privacy, building on research that examines the economic value and consequences of personal data protection and disclosure. Existing literature has extensively investigated consumer privacy preferences and the economic trade-offs associated with data sharing, revealing how digital advancements increase both the value of personal data and concerns over privacy. Previous research by Acquisti et al. (2016) and Goldfarb and Que (2023) documents how consumers navigate privacy concerns in an evolving digital economy, emphasizing that while data disclosure supports economic gains, it also generates privacy costs. Additional studies, such as Acquisti et al. (2015) and Acquisti et al. (2020), highlight the psychological and behavioral complexities surrounding privacy, noting that consumers' desire for privacy often conflicts with the convenience and social benefits of data sharing, complicating their ability to protect personal information fully.

Building on these ideas, studies by Goldfarb and Tucker (2012) and Lin (2022) delve into consumer preferences, distinguishing between intrinsic and instrumental values for privacy. Lin (2022), for instance, demonstrates how privacy preferences vary, with some consumers valuing privacy for its own sake while others see it as a tool for negotiating better outcomes, a nuance that underscores the complex and layered nature of privacy in consumer decisions.

A substantial area of research also focuses on how privacy regulations impact firm-level outcomes, including advertising effectiveness, operational efficiency, and market structure. For instance, Goldfarb and Tucker (2011) examines the effects of privacy regulation on online advertising, finding that stricter regulations limit firms' ability to leverage data, which diminishes the effectiveness of targeted advertising. Johnson et al. (2020) further shows that privacy policies affect revenue by reducing the share of consumers available for targeted advertising. Other research, such as Aridor et al. (2020), explores the impact of the GDPR, finding a trade-off between data volume and quality, as stricter consent requirements

decrease the number of observable consumers but improve advertisers' ability to predict behavior among the remaining data. In terms of market concentration, studies by Campbell et al. (2015) and Johnson et al. (2023) reveal that larger firms adapt more easily to privacy constraints, suggesting that privacy policies may inadvertently shift competitive dynamics, as larger firms with greater resources and infrastructure can sustain data-driven operations despite regulatory limitations. Lin and Strulov-Shlain (2023) add to this discussion by investigating how choice architecture and privacy valuations influence consumers' data-sharing selections, highlighting that privacy regulations may further affect how consumers navigate these choices within digital platforms.

While this literature has provided valuable insights into the economic implications of privacy policies, less attention has been devoted to understanding how these regulations affect consumer behavior directly, particularly in terms of their willingness to share personal data. This paper addresses this gap by focusing on the consumer side of privacy regulation, examining how privacy policies shape individuals' propensity to share information on digital platforms. By shifting the focus from firm-level outcomes to consumer-level responses, our study contributes a new perspective to the economics of privacy literature, enriching our understanding of how privacy protections resonate with the individuals they aim to protect.

Our research setting provides an opportunity to investigate this issue in a highly granular manner. Leveraging detailed data on consumer interactions with a customer engagement platform, we can observe users' data-sharing behaviors over time in response to privacy policies introduced in California and Virginia. This level of granularity allows us to capture nuanced patterns in data-sharing that would be difficult to observe in aggregate data, providing a clearer picture of how privacy policies influence consumer decisions in a digital environment over time. To our knowledge, this is one of the first studies to analyze these consumer-level effects by highlighting the behavioral dimensions of privacy regulation and their potential implications for both digital platforms and regulatory frameworks.

The structure of the paper is as follows: Section 2 introduces institutional details, including recent privacy regulations in the US and the customer engagement platform used in this study. Section 3 describes our data sources and sample construction. Section 4 examines the impact of privacy policies on consumer data-sharing behavior. Section 5 analyzes how the platform adjusted its strategy in response to the policy, focusing on new partnerships and data-sharing incentives. Section 6 explores potential mechanisms underlying the observed behavioral shifts, while Section 7 investigates heterogeneity in responses across different consumer groups. Section 8 examines factors that predict consumer engagement in data sharing. Finally, Section 9 provides concluding remarks.

2 Empirical Setting

2.1 Privacy Policy Regulations in the US

In recent years, privacy regulations worldwide have evolved significantly in response to rising concerns over consumer data protection and digital privacy. The European Union's General Data Protection Regulation (GDPR), enacted in 2018, marked a milestone by establishing comprehensive data protection principles, such as the right to access, correct, and delete personal data, along with stricter consent requirements. These regulatory efforts are largely driven by the exponential growth of digital platforms and the increasing volume of personal data collected, raising concerns about consumer rights, data misuse, and transparency. In the United States, privacy discussions have primarily focused on creating state-led policies to address these challenges, aiming to build consumer trust in digital environments without comprehensive federal privacy legislation.

In the U.S., privacy regulations vary at the state level. The California Privacy Rights Act and the Virginia Consumer Data Protection Act, both effective as of January 1, 2023, represent two significant advances in data privacy regulation. The California law expands on earlier legislation by strengthening consumer rights and placing additional obligations on businesses that collect and process personal data. It enhances consumer control by granting individuals the right to correct inaccuracies in their data, limit the use of sensitive personal information, and opt out of data sales and cross-context behavioral advertising. Additionally, it created an independent enforcement agency to oversee compliance.[1]

Similarly, Virginia's law introduces key privacy protections by giving consumers the right to access, delete, and correct their data. It requires businesses to obtain consent before processing sensitive data, which includes precise geolocation, biometric data, and information related to children. It also allows users to opt out of targeted advertising, data sales, and profiling for automated decision-making. These legislative changes not only set new benchmarks for consumer rights but also challenge businesses to rethink their data management strategies. This regulatory environment encourages companies to adopt more transparent data practices, potentially leading to a more consistent approach to privacy across states as other legislatures consider similar laws. As more states explore comparable legislation,[2] these laws establish a foundation that gradually shapes privacy practices across the country, influencing data protection approaches and shaping consumer expectations in the digital

[1] Although enforcement began on July 1, 2023, many businesses, including the platform analyzed here, began implementing the policy's requirements earlier to align with legal obligations and avoid future compliance risks.

[2] For more details on ongoing privacy regulations in the U.S., please refer to https://iapp.org/resources/article/us-state-privacy-legislation-tracker/

landscape.

2.2 Customer Engagement Platform

The dataset used in this study is provided by a leading customer engagement platform based in the U.S. The platform incentivizes users to upload receipts from everyday purchases, creating a rich dataset of shopping behaviors. Users are rewarded with points for each receipt upload, with additional incentives available for purchasing specific products or reaching spending thresholds. Points accumulated through these activities can be redeemed for rewards, such as gift cards and discounts, creating a cycle of engagement that motivates consistent use and data sharing. The platform's business model is centered on partnerships with brands who pay to promote their products and gain insights into consumer behavior. This rewards-based model encourages users to participate actively, providing a comprehensive view of their spending patterns across a range of retailers and product categories.

Beyond receipt uploads, the platform offers users tailored promotions and in-app offers designed to increase engagement and direct purchasing behavior toward specific brands and products. Through partnerships with brands, the platform incentivizes particular items, rewarding users for selecting promoted products. These promotions help the platform capture a wide spectrum of user data, spanning basic purchase details to more nuanced interactions with branded promotions and offers.

Data collected on this platform generates valuable insights into consumer purchasing behaviors, which are used both to enhance the user experience and to inform marketing strategies for brand partners. The platform's personalized recommendations and offers are driven by individual user histories, making it a dynamic environment where user behavior and preferences are continuously shaped by both consumer choices and platform-driven incentives. This structure provides an ideal empirical setting to explore how digital engagement strategies, combined with privacy policies, impact data-sharing behavior among consumers.

2.3 Implications of Privacy Policies on the Platform

The privacy regulations introduced in California and Virginia impose various requirements on businesses handling consumer data. However, not all provisions directly impact the platform studied here. For instance, restrictions on processing sensitive personal data—such as health or biometric information—are not relevant, as the platform primarily collects consumer shopping behavior data rather than highly regulated personal information. Similarly, provisions related to automated decision-making and profiling have limited applicability, given that the platform's data processing focuses on consumer purchase activity rather than

predictive risk assessments.

The most significant regulatory changes affecting the platform are the requirement to provide transparency about the data collected and the requirement to offer consumers the ability to opt out of data sales. Since the platform monetizes data by sharing it with third-party advertisers and brand partners, it had to implement an opt-out mechanism, which is now available within its app, allowing users to restrict the sale of their personal information. Additionally, new transparency and disclosure requirements mandated clearer privacy policies and more explicit communication about data usage, leading the platform to update its terms of use to reflect consumer rights under the new laws. These adjustments ensure compliance while maintaining access to consumer data under the revised legal framework.

3 Data

3.1 Platform Data

Receipts data — This study utilizes a dataset from a customer engagement platform containing receipt-level data on 10,000 randomly selected users over a 21-week period from October 2022 to March 2023. The raw data from each receipt upload is aggregated at the user-week level, with the main analysis conducted on this aggregated dataset. This approach helps examine weekly trends in user engagement and data-sharing behavior. To focus on a consistent user cohort, we restrict the sample to users who downloaded the app and uploaded their first receipt in the first week of October, ensuring that all users share a common starting point within the data collection period and are the same in terms of their tenure in the app.

The main variables in the analysis, summarized in Table 1, provide key insights into user activity on the platform. Users uploaded an average of 3.45 receipts per week, and total weekly spending averaged $213.12, reflecting a range of engagement levels, with some users contributing higher spending volumes. Additionally, the average number of stores visited per week was 2.22, with users purchasing an average of 23.79 items weekly. These variables represent different facets of user engagement, from purchase frequency to various retailers and items bought, capturing a broad range of platform interactions.

In addition to the main variables, Table A1 details further engagement metrics, such as the variety of stores, unique ZIP codes variety in retailer categories visited and shared, and item categories purchased each week. These supplementary variables add depth to the analysis, capturing more nuanced aspects of platform use, such as the number of food and health-related purchases or the diversity of store visits. Demographic variables, including user age (Mean = 36.70) and gender (64 percent female), along with behavioral character-

Table 1: Summary Statistics for Main Variables

Variables	Mean	SD	Median
Number of receipts uploaded	3.45	6.08	0
Total spending	213.12	952.56	0
Number of stores visited	2.22	3.55	0
Number of items bought	23.79	54.91	0
Amount spent on awarded items	165.72	794.95	0
Points available / offers available	151.55	95.69	150.75
Recency (weeks since last upload)	2.61	3.62	1

Notes: This table presents summary statistics for the variables used in the main analysis. The dataset is at the user-week level. Total spending and amount spent on awarded items are in US dollars.

istics like total active weeks within a 21 weeks period (Mean = 10.08), add further context, allowing for an in-depth analysis by heterogeneity in user profiles.

Altogether, this dataset offers a comprehensive view of user engagement, supporting an in-depth analysis of how privacy policies and platform incentives influence data-sharing behaviors. By capturing both overall spending patterns and detailed insights into retailer diversity and item categories, it provides a valuable opportunity to examine how preferences for data sharing evolve in response to external factors, such as changes in privacy policies.

Offers data — The dataset includes all 2,832 offers available to customers during the study period, providing a detailed view of the promotional activity on the platform. Each offer is linked to a specific partner and contains key characteristics such as the dates it was available, the number of points awarded, and the U.S. states where it was accessible. This allows us to analyze both the overall volume of offers and the variation in incentives provided by different partners.

To examine how privacy regulations influenced partnerships, we constructed a dataset covering all partners creating these offers. In addition to platform-provided data, we manually collected partner-specific attributes for all 122 unique partners, including the sector in which they operate and the U.S. states where they are active. Based on this information, we categorized partners as either national or local, allowing us to distinguish differences in engagement strategies across partner types.

For the new partnership analysis, we used offer-level data and defined a new partnership based on the first week a partner appeared in the dataset with an active offer. To assess the impact of privacy regulations on data-sharing incentives, we constructed a balanced partner-week panel by filling in missing weeks with no offers for each partner with zeros, ensuring

that changes in offer activity could be accurately measured over time.

3.2 Google Search Volume Index for Privacy Related Terms

The Google Search Volume Index (SVI) provides a measure of public interest in privacy-related topics by tracking search behavior across different U.S. states. In this analysis, we use Google Trends data to examine search activity for the term 'privacy' from October 2022 to March 2023, aligning with the period covered by our platform data. To establish a control period, we also collected Google SVI data from October 2021 to March 2022, allowing us to compare search behavior one year prior to the policy change.Using 'privacy' as the primary search term enables our dataset to capture a broad range of related searches, including 'privacy act', 'privacy rights' and other terms linked to privacy concerns.

Google SVI reflects the relative search interest in privacy-related terms within each state, based on their proportion of overall search volume. The search volumes for each state are then defined relative to each other, allowing for meaningful comparisons within the selected set of states. Google Trends allows search volume comparisons across a maximum of five states at a time. Merging results from different comparisons is not feasible in this case, as Google dynamically normalizes search data based on varying reference points across queries. To ensure consistency, we focus our analysis on a set of large U.S. states—California, Virginia, Texas, Florida, and New York. These states not only capture diverse regional and demographic characteristics but also offer a robust comparison between treated and control groups, minimizing the influence of small-sample fluctuations and ensuring meaningful cross-state analysis.[3]

4 Impacts of Privacy Policies on Consumer Behavior

4.1 Empirical Strategy

To evaluate the impact of privacy policy changes on data-sharing behavior, we employ a difference-in-differences (diff-in-diff) identification strategy as the baseline approach. The main identification equation is as follows:

$$y_{it} = \gamma \text{Post}_t * \text{T}_i + \beta_l \text{X}_{i(t-1)} + \beta_c \text{X}_{it} + + \phi_i + \theta_t + \epsilon_{it} \tag{1}$$

The unit of observation is at user i and week t. y_{it} represents the outcome variables for

[3]We also conducted robustness checks using alternative control states, and the results remain consistent. Additional robustness analyses are available upon request.

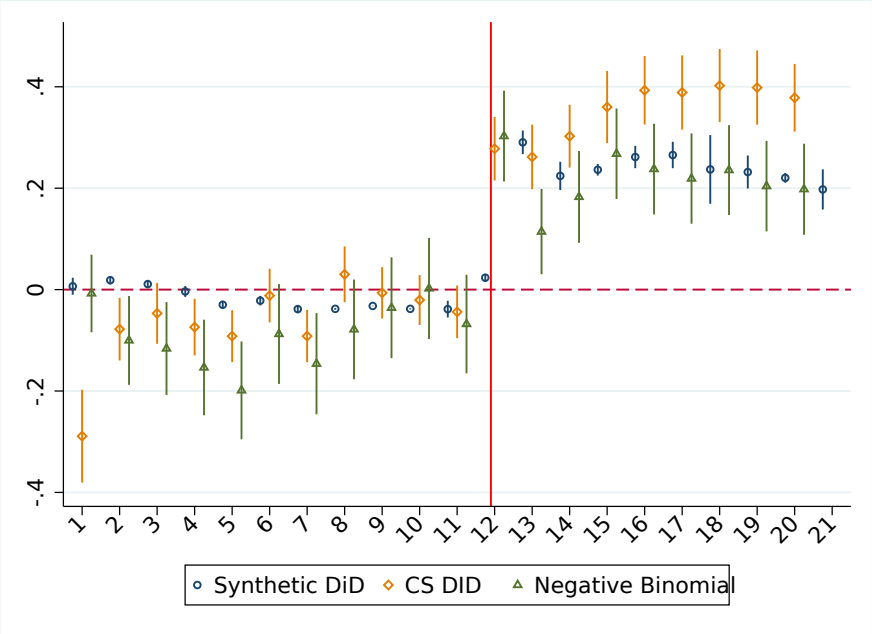
user i at week t. In the main analysis, we use the logarithm of the number of receipts uploaded as an outcome variable. The key explanatory variable, $\text{Post}_t * \text{T}_i$, is an interaction term that captures the effect of being in a treated state (California or Virginia) after the policy implementation. The coefficient γ, therefore, represents the differential impact of the policy on receipt uploads in consumers in treated states compared to consumers in control states. Control variables, denoted as $\text{X}_{i(t-1)}$, include a range of lagged indicators that capture engagement patterns and spending behavior, such as total spending, number of stores visited, number of items purchased, and spending on awarded items. Another set of control variables, represented by X_{it}, includes a ratio of points available to offers available and recency, which is defined as the number of weeks since the last upload. To account for unobserved heterogeneity, we include user fixed effects ϕ_i and week fixed effects θ_t, which control for time-invariant user characteristics and common time shocks, respectively.

An essential assumption in the diff-in-diff approach is the parallel trends assumption, which posits that in the absence of privacy policy changes, treated and untreated states would have experienced similar trends in receipt uploads. To assess the validity of this assumption, we employ an event-study framework, which allows us to observe the trajectory of receipt uploads across treated and untreated states both before and after the policy introduction. The equation for this event-study model is given by:

$$y_{it} = \sum_{j=-2}^{T_0} \beta_j \text{Pre}_j \times T_i + \sum_{k=0}^{T_1} \beta_k \text{Post}_k \times T_i + \gamma_i + \epsilon_{it} \tag{2}$$

where Pre_j and Post_k are sets of indicator variables equal to 1 when an observation is j weeks before or k weeks after the policy implementation. In our case, we observe receipt uploads both before and after January 2023, the effective date of the privacy policies. The coefficients β_j and β_k capture changes in receipt uploads relative to the pre-policy baseline. Figure 1 plots these coefficients and displays the estimated effects along with confidence intervals for three different estimation methods discussed below. Alternatively, we can interpret each estimated parameter after January 1st (shown as Week 12) in this figure as the disaggregation of the average treatment effect γ for each week t. Consistent with the assumption of parallel trends, we expect coefficients before the policy implementation to be close to zero, indicating similar pre-treatment trends between treated and untreated states. A significant deviation after the policy introduction would signal the policy's impact on user engagement.

Figure 1: Changes in Number of Receipts Uploaded



Notes: The figures plot β_j and β_k estimated from Equation 2. The red vertical line represents the week of privacy policy introduction (January 1st, 2023). Synthetic DID refers to synthetic diff-in-diff estimator (Arkhangelsky et al. 2021), CS DID is the Callaway-Sant'Anna diff-in-diff estimator (Callaway and Sant'Anna 2021).

To strengthen our analysis, we employ several alternative methods, including synthetic dif-in-diff (Arkhangelsky et al. 2021), the Callaway-Sant'Anna diff-in-diff estimator (Callaway and Sant'Anna 2021) and negative binomial. The synthetic diff-in-diff approach, which is the main estimation method used throughout this study, is used to create a synthetic control group that more closely resembles the treated units by weighting untreated units to match the pre-treatment characteristics of treated ones. This method improves the comparability between treated and control groups, especially when treated and untreated states differ significantly at baseline. The Callaway-Sant'Anna diff-in-diff (CS DID) estimator allows for treatment effect heterogeneity and provides greater flexibility in controlling for group-specific time trends. It is particularly helpful in this case, as it accommodates heterogeneous treatment effects across units, allowing for differences in treatment effects that vary across individuals or groups within the sample.

Finally, we use a negative binomial model to address the nature of the outcome variable, receipt uploads, which count data with overdispersion (Chen and Roth 2023). The negative binomial regression is well-suited for this type of data, allowing us to account for the high variance in receipt uploads that exceeds the mean. This approach provides a more accurate fit than standard linear models, offering additional robustness to our findings. By employing this variety of methods, we ensure that our analysis captures a nuanced understanding of

how privacy policies influence data-sharing behavior on the platform.

4.2 Results — Impacts on the amount of information shared

In this section, we estimate the baseline regression specified in Equation 1 to assess the impact of privacy policies on the amount of information shared on the platform, specifically focusing on the number of receipts uploaded. Table 2 presents the results of this analysis, showing estimated effects for the main variable of interest, the interaction term $\text{Post}_t * \text{T}_i$ in Equation 1, across three different estimation methods: synthetic diff-in-diff, Callaway-Sant'Anna diff-in-diff (CS DID), and negative binomial regression.

Table 2: Changes in the amount of information shared

	(1) (Synthetic DID)	(2) (CS DID)	(3) (Negative Binomial)
Treated * Post	0.241***	0.312***	0.224***
	(0.015)	(0.019)	(0.012)
Observations	210,000	210,000	210,000
Week FE	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1.

In Column (1), which reports results from the synthetic diff-in-diff model, the coefficient for the main parameter of interest γ in Equation 1 indicates a 27.1% increase in the number of receipts uploaded in treated states after the policy implementation relative to control states.[4] The standard errors for this model are constructed using bootstrapping, which provides robust inference based on many clustered bootstrap resamples from the main dataset.

Column (2) presents results from the Callaway-Sant'Anna diff-in-diff estimator, which allows for potential heterogeneity in treatment effects across units. This estimator suggests that the privacy policies led to a 36.6% increase in receipt uploads in treated states after implementation, further supporting the positive impact of the policies on data-sharing behavior. In Column (3), the negative binomial model estimates a 24.7% increase in receipt uploads following the policy change. This model is particularly suited to our log-linear outcome variable, as it accommodates the count nature and overdispersion of receipt uploads, providing a more tailored approach to modeling this type of data.

[4]The effect is calculated as $\exp(0.24)-1$, which provides the percentage change in the outcome. This interpretation is used because the regression is log-linear, where the dependent variable is log-transformed, and the independent variable is binary.

Overall, these results demonstrate a consistent positive effect of privacy policies on data-sharing behavior, as evidenced by the increase in receipt uploads across all estimation methods. The findings suggest that the introduction of privacy policies in California and Virginia may have fostered a sense of trust or transparency that encourages users to share more information on the platform. Across all three models, the significant and positive treatment effects indicate that the privacy policies have effectively promoted data-sharing, implying that regulatory interventions can have a tangible influence on user engagement in digital platforms.

4.3 Results — Impacts on the type of information shared

In addition to the increase in the volume of data shared, this section explores whether these policies also influenced the content of the information shared by users. Understanding not only how much data users share but also what types of data they are willing to disclose provides a richer perspective on the behavioral shifts affected by these privacy regulations.

To examine this, we extend our analysis by estimating the baseline regression Equation 1 with different outcome variables by using synthetic diff-in-diff, focusing on various aspects of data content. The results are organized into three main areas: the diversity of store visits, the types of retailers from which receipts are shared, and the variety of items recorded in receipts. Through this analysis, we aim to capture a more nuanced view of how privacy policies shape user engagement with data-sharing platforms.

In this analysis, retailer categories were developed by manually classifying the 207 most common retailers in the dataset, including food and beverage, department stores, health and beauty, and specialty stores (covering household, apparel, and entertainment).[5] For items, given the highly granular nature of item categorization in our dataset, we manually recoded 460 items into broader categories to facilitate easier interpretation. These categories include food and beverage, alcohol, health and beauty, household, and apparel. The retailer and item classifications were guided by the Consumer Expenditure Survey and the Nielsen retailer scanner panel classifications, following methodologies from Dubé et al. (2018) and DellaVigna and Gentzkow (2019).

Volume of store visits — The first part of the analysis examines whether privacy policies affect the diversity of store visits shared by users. Table 3 presents the results with three outcome variables: the number of unique stores visited, the number of unique ZIP codes,

[5]We focus on the 207 most common retailers due to a large number of unique retailers in the dataset. These most common retailers, covering approximately 75 percent of all retailer visits, were identified based on the frequency distribution of retailers in the dataset.

and the number of unique retailer categories.

Table 3: Changes in the type of data shared: Store visits

	(1) (Unique stores)	(2) (Unique ZIPcodes)	(3) (Retailer variety)
Treated*Post	0.179***	0.129***	0.113**
	(0.007)	(0.008)	(0.001)
Observations	210,000	210,000	210,000
Week FEs	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1.

In Column (1), the coefficient for the DiD coefficient γ indicates an increase in the number of unique stores shared by treated users post-policy, corresponding to approximately 19.5%. This suggests that users in treated states began sharing information about a broader range of stores after the privacy policies were implemented. Column (2) shows a 13.7% increase in the number of unique ZIP codes of stores visited by users in treated states, reflecting a broader geographic distribution in the data shared. This outcome implies that users are not only visiting more locations but are also more willing to share information about these diverse locations after the policy change. Lastly, Column (3) reveals an increase of approximately 11.8% in the number of unique retailer categories shared, suggesting that treated users have begun sharing information about a more diverse range of retailer types after the policy. This points to a shift in the variety of data shared, likely influenced by the increased sense of privacy protection offered by the policies.

Overall, these results indicate that the privacy policy changes not only increased the amount of information shared but also encouraged users to disclose a more diverse array of data. This suggests that users are more comfortable sharing information about their varied shopping habits and may feel more secure in revealing broader details about their consumer behavior.

Type of retailers — To explore the impact of privacy policy changes on the types of retailers for which users share data, we estimate the baseline regression Equation 1 with various retailer categories as the outcome variables. Table 4 presents the results for four retailer categories: food and beverage, department stores, health-related stores, and specialty stores.

Table 4: Changes in the type of data shared: Type of retailers

	(1) (Food&Beverage)	(2) (Department Stores)	(3) (Health)	(4) (Specialty Stores)
Treated * Post	0.146***	0.131***	0.028***	0.075***
	(0.001)	(0.008)	(0.002)	(0.004)
Observations	210,000	210,000	210,000	210,000
Week FEs	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1.

In Column (1), the positive coefficient for the DiD estimator term indicates a 15.7% increase in the number of visits to food and beverage retailers shared by users in treated states post-policy. This suggests that users are increasingly willing to disclose information about their interactions with essential goods and everyday consumption. Column (2) shows a similar effect for department stores, with a 14% increase in shared visits, indicating that users are more comfortable sharing data from general retail experiences. Health-related stores, mainly including pharmacies and drug stores, shown in Column (3), display a smaller but significant increase of 2.8%, while specialty stores covering household, apparel, and entertainment-related retailers in Column (4) exhibit a 7.8% increase in visits shared. These findings suggest that privacy policy changes encourage users to share data from a broader range of retailer types, potentially due to an increased sense of control and trust over their data privacy.

Table 5 focuses on the sharing of digital receipts. In this platform, digital receipts are uploaded automatically. Users can connect their e-commerce accounts or email accounts to the platform, enabling digital receipts from online purchases to be uploaded directly. The first column in this table estimates the effect of privacy policies on the number of digital receipts uploaded, while the second column estimates the likelihood of users connecting their digital accounts.

Table 5: Changes in the type of data shared: Digital receipts

	(1) (Number of digital receipts)	(2) (Digital account connection)
Treated * Post	0.044***	0.013***
	(0.000)	(0.001)
Observations	210,000	210,000
Week FE	Yes	Yes
Individual FE	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1.

Column (1) reveals a 4.5% increase in the number of digital receipts shared post-policy, indicating that users in treated states are more inclined to allow their online purchase history to be shared. This may reflect increased trust in the platform's handling of sensitive digital transactions. Column (2) shows a 1% increase in the probability of users connecting their digital accounts post-policy, which suggests that privacy policies may encourage users to integrate more of their digital shopping behavior into the platform. Together, these results indicate that privacy protections could lead to greater user participation in sharing digital data, further diversifying the types of information available on the platform.

Overall, the findings in Table 4 and Table 5 suggest that privacy policies impact not only the quantity of shared data but also the types of retailers and transaction methods users feel comfortable sharing. This broader range of shared data enhances the platform's understanding of user preferences, contributing to a richer data environment influenced by privacy regulations.

Type of items — Next, our analysis investigates whether privacy policies influenced the variety of item categories that users are willing to share on the platform. Table 6 presents the results, with outcome variables representing different types of items that appear on uploaded receipts: alcohol, food, health, apparel, entertainment, and household items.

Table 6: Changes in the type of data shared: Type of items

	(1) (Alcohol)	(2) (Food)	(3) (Health)	(4) (Apparel)	(5) (Entertainment)	(6) (Household)
Treated*Post	0.016***	0.141***	0.060***	0.016***	0.023***	0.047***
	(0.006)	(0.004)	(0.006)	(0.001)	(0.001)	(0.001)
Observations	210,000	210,000	210,000	210,000	210,000	210,000
Week FEs	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1.

In Column (1), we observe a small but statistically significant 1.6% increase in the sharing of alcohol-related items post-policy, suggesting that users are slightly more open to disclosing information on discretionary purchases. Column (2) indicates a more substantial effect for food items, with a 15.2% increase, implying that users are considerably more willing to share data on essential and frequently purchased goods. This pattern aligns with the results observed in in-store and retailer categories, reflecting an increased openness toward sharing everyday consumption information.

Health-related items, shown in Column (3), increase by approximately 6.2%, while apparel items in Column (4) see a smaller but significant 1.6% rise. The results for entertainment (Column 5) and household items (Column 6) show 2.3% and 4.8% increases, respectively, indicating that privacy policies may encourage users to share data across a wider range of personal and household consumption categories.

These findings suggest that privacy policy changes not only increase the amount of information users share but also expand the variety of product categories included in their shared data. This shift aligns closely with the results from the retailer type analysis, where we observed a broader range of retailer categories being shared post-policy. The increased openness to disclose information on diverse item types—ranging from essentials like food and household goods to more personal categories like health and entertainment—highlights a pattern of greater user trust in the platform. Together, these results underscore the broader impact of privacy protections in fostering a more comprehensive data-sharing environment, allowing for a richer understanding of user behavior across different consumption domains.

4.4 Supporting Evidence from Consumer Expenditure Survey

To further validate the behavioral shifts observed in response to privacy policies, we complement our platform-based analysis with evidence from the Consumer Expenditure (CE) Survey. This survey provides household-level data on expenditures, as well as metadata on

the interview process itself, including whether respondents expressed concerns about privacy and the extent to which they shared information when asked about their spending. By leveraging this data, we assess whether the implementation of privacy policies influenced consumers' willingness to disclose expenditure details in a broader setting beyond the platform analyzed in the rest of the paper.

By using Consumer Expenditure Survey data, we construct two key outcome variables. The first is an indicator of whether a respondent expressed privacy concerns during the survey. This captures instances where respondents explicitly cited privacy as a reason for hesitation, expressed broader concerns about data collection, or refused to answer expenditure-related questions due to concerns about personal information. The second outcome measures the proportion of expenditure-related questions answered relative to the total number of questions asked, including those that were refused or responded to with "do not know." A lower value for this measure suggests a greater reluctance to disclose expense details, whereas a higher value indicates a greater degree of engagement in sharing expenditure information.

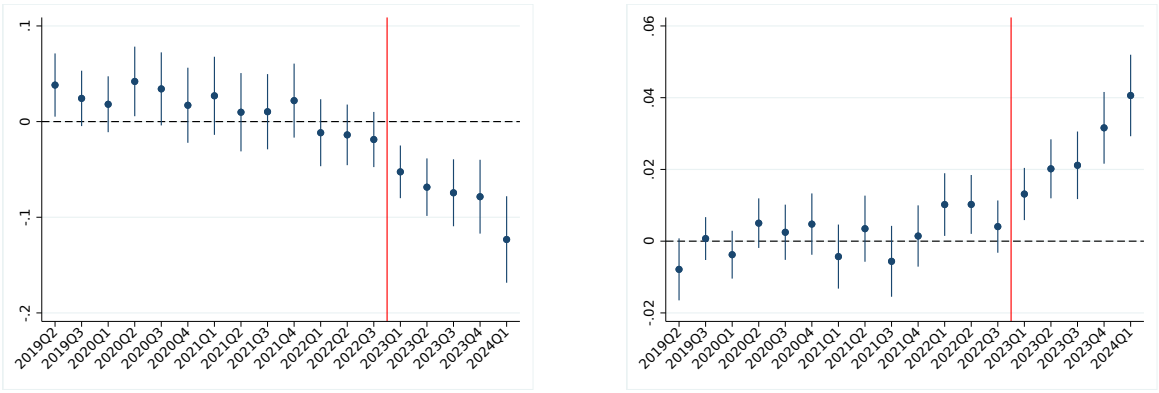
To estimate the impact of privacy policies on these outcomes, we use a difference-in-differences (DiD) approach, estimating the following specifications:

$$y_{it} = \gamma \text{Post}_t * \text{T}_i + \beta_c \text{X}_{it} + \phi_i + \theta_t + \epsilon_{it} \tag{3}$$

where y_{it} represents the two outcome variables: (i) an indicator for whether the respondent i raised privacy concerns at time t and (ii) the share of expenditure-related questions answered. The variable $\text{Post}_t * \text{T}_i$ is an interaction term that takes the value of one if respondent i is from a treated state (California or Virginia) and the interview occurs after January 1, 2023, and zero otherwise. Household fixed effects (ϕ_i) control for time-invariant respondent characteristics, and quarter fixed effects (θ_t) account for seasonality in survey responses. Additionally, we control for income (X_{it}) when estimating the ratio of shared expenditures, as income levels may influence individuals' likelihood of disclosing expenditure details.

Before analyzing the results from Equation 3, we assess whether the pre-policy trends satisfy the parallel trends assumption. The event-study estimates, presented in Figure 2, plot the evolution of privacy concerns (left panel) and the ratio of shared expenditures among all questions asked (right panel). Both series show no significant differences between treated and control states before the policy implementation, suggesting that privacy concerns and expenditure-sharing behaviors followed similar trajectories in treated and control states prior to the regulatory changes.

Figure 2: Changes in Response Behavior in CE Survey



(a) Privacy concerns **(b)** Ratio of Shared Expenditures

Table 7 presents the results examining changes in privacy concerns. Following the policy implementation, respondents in treated states became significantly less likely to express privacy-related concerns during the survey. Column (1) indicates that the likelihood of raising privacy concerns decreased by approximately 5.3 percentage points after the policy implementation. This represents a meaningful reduction, given that the baseline prevalence of privacy concerns in the sample is around 1.2 percent. The reduction in privacy concerns is particularly pronounced among those who, prior to the policy, exhibited lower levels of data-sharing, while the decline is smaller but still significant for those who were already more likely to share information. These findings suggest that individuals who were initially more cautious about privacy responded more strongly to the regulatory changes, potentially due to increased confidence in data protection.

Table 7: Changes in privacy concerns

DV: Privacy concern	(All)	(Low sharers (p25))	(High sharers (p75))
Treated*Post	-0.0528***	-0.100***	-0.0439***
	(0.012)	(0.033)	(0.013)
Observations	77,645	15,555	61,468
R-squared	0.614	0.663	0.591
Household FE	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1.

Table 8 reports the results for the ratio of shared expenditures. Respondents in treated

states increased the proportion of expenditure questions they answered relative to the total number asked. Column (1) shows that the ratio of shared expenditures increased by 1.4 percentage points post-policy. As with privacy concerns, the increase in shared expenditures is driven primarily by those who were initially more reluctant to disclose spending details, while those who were already more forthcoming exhibit a more modest change. These findings align with the behavioral shifts observed in the customer engagement platform, where data-sharing activity increased following the policy implementation.

Table 8: Changes in the ratio of shared expenses overall responses

DV: Ratio of shared expenses	(All)	(Low sharers (p25))	(High sharers (p75))
Treated*Post	0.0145***	0.0616***	0.0056**
	(0.003)	(0.017)	(0.002)
Observations	69,646	12,640	56,434
R-squared	0.664	0.684	0.633
Household FE	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1.

Overall, the results from the CE Survey provide external validation for the patterns identified in the main analysis. The observed decrease in privacy concerns, particularly among respondents who were initially more hesitant to share information, complements our findings from the platform data. Similarly, the increase in reported expenditure shares aligns with the evidence of greater engagement in data-sharing behaviors on the platform. Taken together, these results suggest that privacy policies not only shape behavior within digital platforms but may also influence broader patterns of information disclosure in other contexts.

5 Impacts of Privacy Policies on Platform Strategy

Privacy regulations introduced stricter requirements for data transparency and consumer control, but they did not directly restrict the platform's data collection, as it does not process sensitive personal information. The most immediate impact came from the requirement to provide an opt-out mechanism and clearer disclosures on collected data, prompting the platform to update its terms of use and adjust its compliance practices. These changes could have influenced how the platform structures data-sharing incentives, as well as how it engages with its business partners.

From the platform's perspective, stricter data regulations may have changed the marginal

value of consumer data for the platform. By increasing compliance obligations and making data collection more transparent, the policy could have led the platform to reconsider its approach to maintaining user engagement. If regulatory requirements raised the cost of managing consumer data, the platform may have adjusted its incentive structure to ensure continued participation. Offering greater rewards for data-sharing activities could help sustain engagement levels and offset any potential decline in data availability due to opt-outs. Even if the platform did not anticipate a major reduction in data collection, it might have still modified incentives to reinforce engagement and ensure a steady flow of high-quality data, particularly as users become more aware of their privacy rights and how their information is being used.

Besides, since the policy increased transparency and potentially built greater consumer trust, it may have led to more data being shared on the platform (as shown in the previous section). This, in turn, could have created an opportunity to expand business partnerships. With an increase in consumer data, the platform may have been in a stronger position to attract new partners by offering a larger dataset and more detailed consumer insights. Firms that rely on these insights for marketing and promotional strategies may have seen more value in collaborating with the platform simply because there was more consumer data available for analysis and targeting.

In the next section, we test these two predictions of whether these changes influenced new partnerships and how the platform adjusted its data-sharing incentives in response to the policy.

5.1 Results — New partnerships

To examine the impact of the privacy policy on new partnerships, we estimate the following difference-in-differences specification:

$$y_{it} = \beta(\text{Treated}_i \times \text{Post}_t) + \alpha_i + \gamma_t + \delta_s + \epsilon_{it} \tag{4}$$

where y_{it} is an indicator variable that takes the value of 1 if a partner appears in the offer data for the first time from the third week after the policy implementation onward, and 0 otherwise.[6] The variable treated_i is an indicator for whether partner i operates in the treated states (California or Virginia), including both national and local partners. The variable post_t takes the value of 1 for weeks after the policy implementation and 0 otherwise. The model

[6]The decision to use the third week as a cutoff is based on the average offer posting frequency among partners, which is 0.3 weeks with a standard deviation of 1.48 weeks. This threshold ensures that we do not misclassify existing partners who post infrequent offers as new partnerships. As a robustness check, Table A2 presents results using alternative cutoff weeks.

includes state (α_i), week (γ_t), and sector (δ_s) fixed effects to account for time-invariant differences across states, common temporal shocks affecting all partners, and sector-specific variations in offer strategies.

Table 9 presents the results of this estimation. Across specifications, the results suggests an increase in the likelihood of new partnerships following the policy implementation. These results indicate a change in partner engagement with the platform for firms operating in treated states. To assess the validity of these findings, Figure A1 presents an event study analysis to examine pre-trends in new partnerships. The results suggest no significant differential trend between treated and control groups before the policy implementation, supporting the parallel trends assumption.

Table 9: Results: Impacts on new partnerships

	(1)	(2)	(3)
Treated*Post	0.208***	0.0962***	0.0986***
	(0.0395)	(0.0220)	(0.0235)
Observations	2,832	2,830	2,791
R-squared	0.201	0.539	0.558
Week FE	Yes	Yes	Yes
Sector FE	No	Yes	Yes
State FE	No	No	Yes

Notes: Column 1 to 3 shows the estimation results for Equation 4. The data used in this analysis is at offer level. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

While the analysis suggests a shift in new partnerships after the policy, there are important limitations to consider. Partners are classified into three groups based on their operational scope: (i) national companies operate in the majority of U.S. states, (ii) local companies operate in a limited number of states, including California or Virginia, and (iii) another group of local companies operates in states that exclude California and Virginia. The treated group consists of firms operating in California or Virginia ((i) and (ii)), including both national and local companies, while the control group consists of local companies from non-treated states ((iii)).

An important limitation is the imbalance between treated and control groups. National companies make up the majority of the sample, leaving only a small number of firms in the control group. In particular, only six firms fall into the control group, which limits the robustness of the results. A small control group introduces challenges in identifying

treatment effects, as differences between treated and control firms may reflect underlying heterogeneity rather than the policy's impact. Given these constraints, while the difference-in-differences results in Table 9 and the parallel trends analysis in Figure Y suggest a change in partnership formation following the policy, we consider these findings as suggestive rather than establishing a definitive causal relationship.

5.2 Results — Data-sharing incentives

To examine whether the platform adjusted data-sharing incentives in response to the privacy policy, we estimate Equation 4 using different outcome variables. Table 10 presents the results of this analysis. The dependent variables include: points per offer provided by each partner in a given week (Column 1), the number of offers available in states including California and Virginia (Column 2), the number of offers exclusively available to users in these states (Column 3), and the number of offers specifically targeting food purchases (Column 4). The inclusion of the latter outcome follows from the observed increase in receipt uploads within the grocery category in Table 4.

Table 10: Results: Impacts on data-sharing incentives

	(1) (Points per offers)	(2) (Nb of Offers)	(3) (Nb of Offers)	(4) (Nb of Offers)
Treated*Post	-0.217	0.00259	-0.0328	-0.0410
	(0.145)	(0.0764)	(0.0563)	(0.0296)
Observations	2,806	2,806	2,806	2,806
R-squared	0.934	0.760	0.402	0.626
Payer FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes

Notes: Column 1 to 3 shows the estimation results for Equation 4. The data used in this analysis is at offer level. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Across all specifications, the coefficient on β is statistically insignificant, suggesting that the platform did not systematically change its data-sharing incentives following the policy implementation. The estimate in Column (1) implies no discernible effect on the number of points offered per receipt, indicating that the platform did not attempt to counteract potential user opt-outs by increasing monetary incentives. Similarly, Columns (2) and (3) show no evidence of an increase in the number of offers available to users in treated states, either for offers that included California and Virginia or those exclusive to them. Finally, Column (4) reveals no significant change in the number of offers targeting grocery purchases, suggesting that even in the category where consumer engagement increased most, the platform did not adjust its promotional strategies.

These results indicate that, despite the policy's impact on consumer engagement, the platform did not alter its incentive structures to encourage further data sharing. One possible explanation is that the observed increase in receipt uploads was due to the privacy policy, driven by consumer responses to transparency and trust rather than platform-driven interventions. Alternatively, the platform may have anticipated that the net effect of the policy on data availability would be neutral, leading to no major adjustments in its promotional strategies. Overall, these findings suggest that, at least in the short term, the platform maintained its existing approach to incentivizing consumer participation without significantly modifying its reward structures or promotional targeting.

6 Mechanisms

6.1 Interest in privacy – Google Trends analysis

A key question arising from our main findings is why individuals in treated states increased their data-sharing behavior following privacy regulations. One potential mechanism is heightened public awareness of privacy issues, which may have influenced perceptions of data protection and, in turn, sharing behavior. If privacy laws triggered greater interest in privacy-related topics, individuals might have become more conscious of their data and digital rights, potentially making them more willing to share information under a regulatory framework they perceived as safer. To test this, we analyze Google search behavior to assess whether public engagement with privacy-related topics increased around the time of policy implementation, particularly in treated states.

To investigate this, we conduct a set of analyses using Google Search Volume Index (SVI) data for the term "privacy". We test three key predictions: (i) whether treated states (California and Virginia) exhibited greater interest in privacy-related topics around policy implementation relative to control states, (ii) whether privacy-related searches increased over time within treated states, particularly as the policy implementation approached, and (iii) whether there was a differential increase in search interest for privacy in treated states compared to control states, accounting for both temporal and cross-state patterns.

It is important to note that interpreting these results requires caution, as this setting differs from a standard difference-in-differences (DiD) framework in key ways. First, privacy policies received national media attention, meaning individuals in control states may have also been exposed to discussions surrounding these policies, potentially influencing their search behavior. As a result, assuming that control states remained entirely unaffected may not be realistic. Second, defining a clear "pre" and "post" period is more complex

than in conventional causal inference settings. Given that discussions and media coverage of the policy likely intensified in the weeks leading up to implementation, we set the post-policy period to begin one month before the implementation date to capture shifts in search behavior leading up to the policy change.

To visually assess search trends, Figure A2 presents the average Google SVI values for the pre- and post-policy periods, separately for treated and control states. We observe that, while treated states already exhibited a greater interest in privacy-related topics before the policy, this interest further increased around the time of policy implementation. In contrast, the control group did not experience a similar rise; if anything, privacy-related search activity slightly declined. The increase in search volume in treated states is nontrivial—suggesting a substantial rise in public engagement with privacy issues, reinforcing the idea that individuals in these states were more attuned to privacy matters at the time of policy introduction.

To formally test for differential trends in privacy-related searches, we estimate the following regression model:

$$\text{SVI}_{st} = \beta_1(\text{Treated}_s \times \text{Post}_t) + \gamma_s + \delta_t + \varepsilon_{st} \tag{5}$$

where SVI_{st} represents the search interest for privacy topics in state s at time t. Treated_s is an indicator for whether a state was subject to privacy regulation (California or Virginia), and Post is a binary variable equal to one for the post-policy period (one month before implementation onward). The model includes state fixed effects (γ_s) to control for time-invariant state characteristics and month fixed effects (δ_t) to account for national trends in privacy-related searches. The error term ϵ_{st} captures unobserved factors affecting search interest in a given state and month.

Regression results, reported in Table 11, show that treated states exhibited significantly greater search activity for privacy-related topics compared to control states around the time of policy implementation. Column (1) estimates the effect of treated status on search interest and finds that, on average, treated states had 6.55 units higher Google SVI than control states during this period. Column (2) examines whether search activity increased over time within treated states, finding that privacy-related searches rose by approximately 1.18 units after the policy period. Finally, Column (3) confirms that the increase in privacy searches was 1.50 units higher in treated states compared to control states around policy implementation. This reinforces the idea that privacy regulations heightened public engagement with privacy-related issues.

Table 11: Results: Google Search volume Index for Privacy

	(1)	(2)	(3)
Treated	6.554***		
	(0.226)		
Post		1.175**	
		(0.563)	
Treated*Post			1.504***
			(0.195)
Mean Value of SVI	53.65	56.79	53.47
Observations	595	728	1,820
R-squared	0.833	0.098	0.924
Month FE	Yes	No	Yes
State FE	No	Yes	Yes

Notes: Column 3 shows the estimation results for Equation 5. Column 1 presents the regression of SVI terms on the treated dummy for post-period only. Column 2 corresponds to estimation results for regression of SVI term on post dummy for the treated group only. The data used in this analysis is aggregated at the monthly level. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

These findings suggest that states adopting privacy regulations saw greater engagement with privacy-related topics in online search behavior, particularly in the lead-up to policy implementation. The fact that privacy searches rose more sharply in treated states around the policy period—despite already higher baseline levels—suggests that the introduction of privacy laws generated additional awareness and engagement. This aligns with our primary findings on data-sharing behavior, providing evidence that heightened privacy awareness may have influenced consumer responses to privacy protections.

7 Heterogeneity analysis

To gain a deeper understanding of the impact of privacy policies on data-sharing behavior, we consider potential heterogeneity in user responses. Not all users may respond uniformly to privacy changes; differences in previous platform engagement or demographic factors, such as age, could influence how individuals react to enhanced privacy protections. Exploring these variations can provide valuable insights into which groups are more likely to adjust their data-sharing behaviors in response to privacy changes.

In this analysis, we examine heterogeneity in two dimensions: previous activity level on the platform and user age. Each dimension is analyzed separately to assess whether these characteristics affect how users respond to privacy policy changes.

Based on previous activity level — To investigate the impact of previous activity levels, we classify users into four groups based on their historical receipt uploads. For each user, we calculated the number of receipts uploaded before the policy change and then divided users into quartiles according to their activity level within each state. Specifically, we categorized users as falling within the bottom 25th percentile (low activity), below the median (low-to-moderate activity), above the median (moderate-to-high activity), and top 75th percentile (high activity). This classification allows us to examine whether the response to privacy policies varies by prior engagement with the platform.

For each subgroup, we perform a synthetic diff-in-diff estimation and estimate Equation 1, comparing each activity level in treated states to the overall user base in control states. This approach helps isolate the policy's impact on users with varying levels of past engagement.

Table 12: Results: Heterogeneity based on previous activity level

	(1) (Bottom p25)	(2) (Below median)	(3) (Above median)	(4) (Top p75)
Treated*Post	0.238***	0.265***	0.223***	0.128***
	(0.004)	(0.014)	(0.039)	(0.045)
Mean number of receipts	0.12	0.48	6.37	9.41
Median number of receipts	0	0	5	8
Observations	193,368	194,103	193,368	185,619
Week FEs	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1.

Table 12 presents the results of this analysis. In Column (1), we observe that users in the lowest activity quartile (bottom 25%) show a significant increase of approximately 27% in their data-sharing behavior post-policy. This finding suggests that users with previously low engagement levels are more responsive to the enhanced privacy protections. Column (2) shows that users below the median activity level also exhibit a strong response, with a 30% increase in data sharing, indicating that privacy changes similarly influence users with low-to-moderate engagement.

As we move to higher activity levels, the responsiveness diminishes. Column (3) shows a 25% increase for users above the median, while Column (4) indicates that users in the top 75% activity level group show a smaller, but still significant, increase of around 14%. These

results suggest that users with lower prior engagement are more sensitive to privacy policy changes, possibly because the policies provide a greater sense of security and trust for those who were previously more cautious in sharing data. In contrast, highly active users may already feel comfortable sharing data and, therefore, show a relatively muted response to the new privacy protections.

Overall, these findings highlight the importance of accounting for previous activity levels when evaluating the impact of privacy policies. The results indicate that privacy changes can particularly influence users who were initially less engaged, potentially broadening the platform's user base and increasing participation from previously cautious users.

Based on age — To further explore the differential impact of privacy policies, we examine heterogeneity based on user age, grouping users by generational categories: Baby Boomers (born 1946–1964), Generation X (born 1965–1980), Millennials (born 1981–1996), and Generation Z (born 1997–2012). Age may play a critical role in shaping responses to privacy policies, as different generations vary in their familiarity with digital platforms and levels of privacy concern (Goldfarb and Tucker 2012). Older users, such as Baby Boomers and Gen X, may be more cautious with data-sharing and see privacy protections as a key factor in their engagement, while younger users, especially Gen Z, may be more accustomed to data-sharing practices and potentially less sensitive to additional privacy safeguards.

Table 13: Results: Heterogeneity based on user age

	(1) (Baby Boomers)	(2) (GenX)	(3) (Millennials)	(4) (GenZ)
Treated * Post	0.264***	0.285***	0.273***	0.199***
	(0.054)	(0.070)	(0.005)	(0.007)
Mean number of receipts	4.49	4.97	3.86	2.20
Median number of receipts	3	3	1	0
Observations	180,768	183,666	186,543	191,436
Week FEs	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1.

Table 13 presents the results of the analysis for each generation. In Column (1), we observe that Baby Boomers exhibit a 30.2% increase in data-sharing behavior following the policy implementation. Similarly, Generation X in Column (2) shows a strong response, with a 32.9% increase in shared data. Millennials, shown in Column (3), also demonstrate a significant increase of 31.4%, aligning closely with the older generations in their response.

In contrast, Generation Z in Column (4) displays a smaller increase of 22%. This comparatively lower response among Gen Z users may suggest that younger users, who are generally more accustomed to sharing data online, are relatively less influenced by additional privacy protections. They may already feel comfortable sharing data, reducing the relative impact of privacy policy changes on their behavior.

Overall, these findings highlight that privacy policies have a stronger impact on data-sharing behavior among older users, who may view these protections as a key factor in their willingness to engage with the platform.

8 What predicts sharing more information?

Building on our analysis of the impact of privacy policies on users' willingness to share information, we now investigate what predicts the actual behavior of uploading receipts to the platform. Understanding the predictors of receipt uploading can provide insights into the types of incentives or behaviors that encourage users to engage with the platform. This analysis is crucial in identifying variables that correlate with data-sharing behavior, helping to consider the potential mechanisms for the impacts of privacy policies.

Baseline Estimation for the Number of Receipts — To examine the predictors of receipt uploading, we estimate the following regression model, where the dependent variable is the log of the number of receipts uploaded by each user. The estimation equation is as follows:

$$\log(\text{receipts}_{it}) = \beta_1 X_{it} + \beta_2 X_{i(t-1)} + \alpha_i + \gamma_t + \epsilon_{it} \tag{6}$$

where $\log(\text{receipts}_{it})$ represents the log-transformed number of receipts uploaded by user i in week t. The explanatory variables are grouped into two components: X_{it}, which includes contemporaneous variables such as points available per offer and recency (weeks since the last upload), and $X_{i(t-1)}$, which consists of lagged variables like weekly spending, the amount spent on awarded items, the number of items bought, and the number of stores visited.

However, variables such as points and offers available may be endogenous with respect to users' receipt uploading behavior. For instance, users who frequently upload receipts may receive more points and offers as part of the platform's engagement strategy, creating a reverse causality concern. To address this endogeneity, we re-estimate the model using

only the first two weeks of data. In the initial weeks, the platform has less information about individual user behavior, and the assignment of points and offers is likely to be more exogenous, reducing the risk of reverse causality.

Table 14: Results: Baseline estimation for receipt uploads

DV: log(receipts)	(1) (Full sample)	(2) (First 2 weeks)
Spending at t-1	-0.127***	-0.059***
	(0.004)	(0.009)
Amount spent on awarded at t-1	0.067***	0.110***
	(0.004)	(0.006)
Number of items bought at t-1	0.021***	-0.151***
	(0.004)	(0.012)
Number of stores visited at t-1	0.192***	-0.318***
	(0.007)	(0.023)
Points available / Offers available	-0.0003***	-0.0009***
	(0.000)	(0.000)
Recency	-0.121***	-1.797***
	(0.001)	(0.000)
Observations	210,000	20,000
R-squared	0.570	0.787
Week FE	Yes	Yes
Individual FE	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.1.

In Column (1) of Table 14, we observe that spending in the previous week is negatively associated with the number of receipts uploaded, with a coefficient of -0.127. This negative association suggests that higher spending in one week may reduce the need to upload additional receipts in the subsequent week, possibly due to saturation or decreased activity. However, the amount spent on awarded items shows a positive association, indicating that incentive-linked spending increases the likelihood of uploading receipts.

Other variables, such as the number of items bought and the number of stores visited, are positively associated with receipt uploads, suggesting that a higher volume and diversity of shopping activities encourage data-sharing behavior. Recency also shows a negative relationship, indicating that longer gaps since the last upload are associated with fewer uploads in the current period, likely reflecting diminishing engagement over time.

In Column (2), which restricts the sample to the first two weeks, we see a reduction in the magnitude of some coefficients, including items bought and store visits, which are less endogenous in the early stages of user interaction. The R-squared value in this restricted

sample is higher, indicating a more precise fit. This approach confirms that the initial behavior patterns differ slightly when platform data is minimal, supporting the robustness of the main results while reducing endogeneity concerns.

Instrumenting available points/offers with treatment exposure — To address potential endogeneity between the number of points per offer available and receipt uploading behavior, we employ an instrumental variable (IV) approach. Specifically, we use a variable termed "treatment exposure" as an instrument for the points available per offer. The treatment exposure variable is defined as the ratio of the count of offers available to users in the treated group relative to the total offers available. It can be a plausible instrument because it captures variations in exposure to offers tied to the treatment but is unlikely to be directly related to users' endogenous choices regarding data-sharing behavior.

The use of treatment exposure as an instrument relies on the assumption of the exclusion restriction: it influences users' likelihood of receipt uploads only through the availability of points per offer, not through other unobserved factors that could also drive data-sharing behavior. By isolating this variation, we aim to provide a more accurate estimate of how points per offer influence receipt uploading.

The two-stage instrumental variable estimation is specified as follows:

$$\text{Stage 1:} \quad (\text{points/offers})_{it} = \beta_0(\text{treatment exposure})_{it} + \beta_1 X_{i(t-1)} + +\beta_2 X_{it} + \alpha_i + \gamma_t + \epsilon_{it} \tag{7}$$

$$\text{Stage 2:} \quad \log(\text{receipts}_{it}) = \beta_0(\widehat{\text{points/offers}})_{it} + \beta_1 X_{i(t-1)} + \beta_2 X_{it)} + \alpha_i + \gamma_t + \epsilon_{it} \tag{8}$$

In the first stage represented by Equation 7, we regress the points available per offer on treatment exposure and the control variables $X_{i(t-1)}$ and X_{it} aligned with Equation 6, which include spending, the amount awarded, items, stores, and recency. User fixed effects (α) and week fixed effects (γ) are also included to control for unobserved heterogeneity. In Stage 2 in Equation 8, the log-transformed number of receipts uploaded is regressed on the fitted values of points available per offer from Stage 1, along with the same controls and fixed effects.

Table 15: Results: Instrumenting with treatment exposure

	(1) points/offers	(2) receipt uploads
Spending at t-1	-1.928***	-0.040***
	(0.387)	(0.014)
Amount spent on awarded at t-1	2.678***	-0.033***
	(0.343)	(0.012)
Number of items bought at t-1	-0.376	0.023*
	(0.303)	(0.014)
Number of stores visited at t-1	0.087	0.159***
	(0.583)	(0.021)
Recency	0.628***	-0.136***
	(0.077)	(0.003)
Treatment exposure	29.91***	
	(0.948)	
Points / Offers		0.034***
		(0.001)
Observations	210,000	210,000

Notes: *** p<0.01, ** p<0.05, * p<0.1.

Table 15 presents the results from this instrumental variable estimation. Column (1) reports the first-stage results, showing the strong predictive power of treatment exposure on points per offer. The coefficient on treatment exposure is 29.91, indicating a significant relationship and confirming the relevance of the instrument.

In Column (2), the second-stage results reveal a positive and significant association between points per offer and receipt uploads, with a coefficient of 0.034. This result suggests that, controlling for endogeneity, an increase in points per offer continues to have a positive relation with the likelihood of receipt uploading. Compared to the baseline model's estimate, this IV approach yields a slightly stronger and more robust effect, supporting the conclusion that points per offer significantly encourage data-sharing behavior.

This instrumental variable approach allows us to address potential endogeneity, yielding results that are less biased by reverse causality or omitted variable concerns. By using treatment exposure as an instrument, we obtain a clearer understanding of the relation between the points per offer on receipt uploads, confirming the effectiveness of platform incentives in encouraging user engagement.

9 Concluding Remarks

This study provides new insights into the role of privacy policies in influencing consumer data-sharing behavior. By analyzing the recent privacy regulations in California and Virginia, our results indicate that enhanced privacy protections can increase consumer trust, encouraging a substantial increase in data-sharing activities. This effect is observed through a significant rise in receipt uploads to a customer engagement platform, particularly in high-frequency purchase categories such as food and beverage. It implies that privacy protections can effectively increase consumers' willingness to share transactional data in digital ecosystems.

Beyond the volume of data shared, our analysis suggests that privacy policies influence the breadth of data shared. Treated users were more inclined to disclose diverse information types, including data from various retailers and item categories. This shift indicates that privacy policies may not only increase the data-sharing frequency but also foster more comprehensive engagement with digital platforms, allowing for a richer understanding of consumer behavior.

Our use of Google SVI data to gauge public interest in privacy topics further supports these findings. We observe increased search interest in privacy-related terms within treated states, suggesting a broader environment of privacy awareness that may reinforce consumers' responsiveness to these protections. This analysis underscores the role of privacy regulations in cultivating greater privacy awareness among users, which appears to support the observed behavioral changes.

This paper advances the literature on privacy regulation by focusing on consumer outcomes rather than firm-centric effects like profitability or targeting. Our findings demonstrate that privacy policies can foster greater transparency and trust, encouraging more active consumer participation on data-sharing platforms. By documenting these consumer-level effects, this study highlights the potential of privacy regulations to positively shape data-sharing behaviors, enriching our understanding of how privacy protections influence engagement on digital platforms in a rapidly evolving digital landscape.

References

Acquisti A, Brandimarte L, Loewenstein G (2015) Privacy and human behavior in the age of information. *Science* 347(6221):509–514.

Acquisti A, Brandimarte L, Loewenstein G (2020) Secrets and likes: The drive for privacy and the difficulty of achieving it in the digital age. *Journal of Consumer Psychology* 30(4):736–758.

Acquisti A, Taylor C, Wagman L (2016) The economics of privacy. *Journal of economic Literature* 54(2):442–492.

Aridor G, Che YK, Salz T (2020) *The economic consequences of data privacy regulation: Empirical evidence from GDPR* (National Bureau of Economic Research Cambridge, MA, USA).

Arkhangelsky D, Athey S, Hirshberg DA, Imbens GW, Wager S (2021) Synthetic difference-in-differences. *American Economic Review* 111(12):4088–4118.

Callaway B, Sant'Anna PH (2021) Difference-in-differences with multiple time periods. *Journal of econometrics* 225(2):200–230.

Campbell J, Goldfarb A, Tucker C (2015) Privacy regulation and market structure. *Journal of Economics & Management Strategy* 24(1):47–73.

Chen J, Roth J (2023) Logs with zeros? some problems and solutions. *The Quarterly Journal of Economics* .

DellaVigna S, Gentzkow M (2019) Uniform pricing in us retail chains. *The Quarterly Journal of Economics* 134(4):2011–2084.

Dubé JP, Hitsch GJ, Rossi PE (2018) Income and wealth effects on private-label demand: Evidence from the great recession. *Marketing Science* 37(1):22–53.

Goldfarb A, Que VF (2023) The economics of digital privacy. *Annual Review of Economics* 15(1):267–286.

Goldfarb A, Tucker C (2012) Shifts in privacy concerns. *American Economic Review* 102(3):349–353.

Goldfarb A, Tucker CE (2011) Privacy regulation and online advertising. *Management science* 57(1):57–71.

Johnson GA, Shriver SK, Du S (2020) Consumer privacy choice in online advertising: Who opts out and at what cost to industry? *Marketing Science* 39(1):33–51.

Johnson GA, Shriver SK, Goldberg SG (2023) Privacy and market concentration: intended and unintended consequences of the gdpr. *Management Science* 69(10):5695–5721.

Lin T (2022) Valuing intrinsic and instrumental preferences for privacy. *Marketing Science* 41(4):663–681.

Lin T, Strulov-Shlain A (2023) Choice architecture, privacy valuations, and selection bias in consumer data. *EC*, 960.

Online Appendix

A Additional Tables

Table A1: Summary Statistics for Additional Weekly Data Variables

Variable	**Mean**	**SD**	**Median**
Number of unique stores visited	2.22	3.55	0
Number of unique ZIP codes visited	1.44	2.07	0
Number of unique retailer categories visited	1.02	1.44	0
Number of digital receipts uploaded	0.25	1.81	0
Digital accounts connection status	0.13	0.34	0
Number of food retailers visited	0.87	2.10	0
Number of department stores visited	1.03	2.41	0
Number of health-related stores visited	0.08	0.49	0
Number of entertainment related stores visited	0.04	0.29	0
Number of alcoholic items bought	0.07	0.46	0
Number of food items bought	2.07	7.52	0
Number of health-related items bought	0.33	2.52	0
Number of apparel items bought	0.10	0.64	0
Number of entertainment related items bought	0.10	0.58	0
Number of household items bought	0.33	2.27	0
Total active weeks in the app	10.08	6.79	9
Female	0.64	0.48	1
Age	36.70	14.89	34

Notes: This table presents summary statistics for the additional variables used in the analysis.

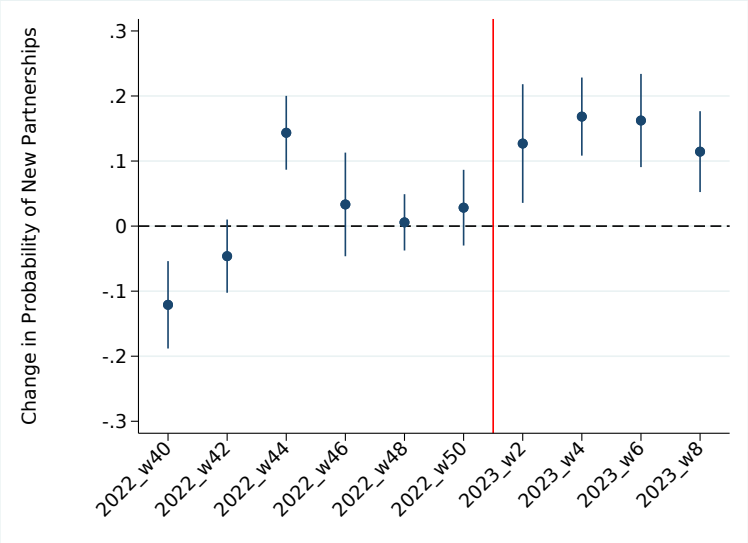
B Additional analysis: Platform strategy

Table A2: Robustness Analysis for New Partnership Analysis

	(1)	(2)	(3)	(4)
Treated*Post	0.149***	0.187***	0.236***	0.288***
	(0.0241)	(0.0282)	(0.0352)	(0.0343)
Observations	2,791	2,791	2,791	2,791
R-squared	0.485	0.456	0.423	0.440
Week FE	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes

Notes: The table shows the estimation results for Equation 4. Column (1) uses fourth week in the data as a new partnership cutoff, i.e. partners appeared in the offer data at fourth week onwards is called a new partner. Columns (2), (3) and (4) uses week 5, week 6, week 7 as a cutoff, respectively.

Figure A1: Event Study for New Partnership Analysis



Notes: The table shows the estimation results for Equation 4 by estimating an event study analysis similar to Equation 2. The new partnership dummy is regressed on the interaction of treated dummies with biweekly fixed effects after controlling for state and sector fixed effects. Given the small number of partners in this data, we plot biweek dummies rather than week dummies to avoid fluctuations. Column (1) uses fourth week in the data as a new partnership cutoff, i.e. partners appeared in the offer data at fourth week onwards is called a new partner. Columns (2), (3) and (4) uses week 5, week 6, week 7 as a cutoff, respectively.

C Analysis using Google SVI

Figure A2: Google Trends SVI for control and treated states

