Cookie Intermediaries: Does Competition Leads to More Privacy?

Arion Cheong, Tawei Wang, D. Daniel Sokol [1]

Abstract

This study examines the relationship between market concentration in the data broker industry and the potential leakage of customer privacy information. We specifically focus on U.S. firms with an online presence and their use of first-party cookies and data trackers for customer data sharing. Our research findings indicate that firms in the concentrated market are less likely to share customer information with third-party data brokers. Furthermore, higher market concentrations for both partners and data brokers are associated with a decreased risk of customer privacy breaches. Additionally, the regulatory status of data brokers influences data-sharing practices, with firms showing a preference for registered brokers in either California or Vermont, the states that require data broker registration. Importantly, our analysis, which incorporates data collected from the dark web, highlights a key finding: registered data brokers in highly concentrated market has significantly lower risk of customer privacy information leakage. These findings emphasize the critical role of market structure and competition in shaping firms' decisions regarding privacy policies and practices.

Keywords: cookie intermediary; privacy; consumer protection; market competition

JEL Classification: M21, M37, M38, M48,

[1] Arion Cheong is an Assistant Professor of Accounting at Stevens Institute of Technology. Email: acheong@stevens.edu. Tawei (David)Wang is an Associate Dean and a Driehaus Fellow at the Driehaus College of Business and a Professor at the School of Accountancy & MIS, DePaul University. Email: david.wang@depaul.edu. D. Daniel Sokol is the Carolyn Craig Franklin Chair in Law and Professor at the USC Gould School of Law and Marshall School of Business. Email: dsokol@usc.edu. Thanks to the workshop participants at the University of Monash, UC Irvine, the University of Toronto, Seoul National University, Rutgers University, the University of Notre Dame, USC, the University of Florida, the National University of Singapore, Luohan Academy, the University of Leeds, the University of Melbourne, the Competition Market Authority (UK), and special thanks to Garrett Johnson and Ahmed Abassi.

Cookie Intermediaries: Does Competition Leads to More Privacy?

1. Introduction

Data brokers, business that gather and sell personal details of customers with whom they do not have a direct relationship, have gained public attention over the past decade (Neumann, Tucker, & Whitfield, 2019; Varnali, 2021). For instance, in 2014, major data brokers, such as Epsilon Data Management, were scrutinized for profiling financially distressed individuals on social media and sharing such information with prospective lenders (Armour, 2014). In 2018, post Cambridge Analytica scandal, Facebook announced its decision to curtail data sharing with data brokers (Seetharaman, Wells, & Vranica, 2018). Despite government reports both in the United States (FTC, 2014) and globally (UK ICO, 2020; PRIV, 2014) highlighting opaque business practices of data brokers, recent lawsuits, such as the Federal Trade Commission (FTC) against Kochava (FTC, 2022), have unveiled a complex two-tier information sharing network among data brokers and their partners, and amongst data brokers themselves (Eckchardt et al., 2019; Choi, Jeon, & Kim, 2019).

The subjects of data use, personalization, and privacy have emerged as vital topics of study in information systems (Lee, Ahn, & Bang, 2011; Sun et al., 2023). Focusing on the relationships among data brokers and partnering organizations, previous literature suggests that market competition can influence a firm's data-sharing practices, creating a decision-making process between sharing customer information and privacy protection (Casadesus-Masanell & Hervas-Drane, 2015; Chen, Choe, & Matusuhima, 2020; Fainmesser, Galeotti, & Momot, 2023). Sharing information can serve as a strategic response to market pressures, aiming to boost performance (Marthews & Tucker, 2019; Soomro, Shah, & Ahmed, 2016; Gal-Or, Gal-Or, & Penmetsa, 2018; Kox, Straathof & Zwart, 2017; De Corniere & De Nijs, 2016), while privacy protection can enhance reputation and foster relationships with stakeholders (Casadesus-Masanell & Hervas-Drane, 2015; Hagiu & Hałaburda, 2014; McWilliams & Siegel, 2011).

However, the understanding of data broker business practices remains limited, and the existing literature on the correlation between competition and privacy protection presents mixed findings. As a result, it remains unclear to what extent partner organizations in the data broker context strategically opt to share customer information in response to market competition (Goldfarb & Tucker, 2011; Adjerid et al., 2016; Buckman, Adjerid & Tucker, 2023).

More specifically, theoretical studies propose that market concentration might lead to less privacy preservation (Ke & Sudhir, 2022; Choe, Cong, & Wang, 2023), while empirical findings suggest a potential for better privacy preservation in more concentrated markets (Johnson, Shriver, & Goldberg, 2023; Jia, Jin, & Wagman, 2021). Interestingly, the enforcement of the General Data Protection Regulation (GDPR) led to increased market concentration among vendors using personal data, illustrating the complexities of stringent data regulations (Johnson, Shriver, & Goldberg 2023). The GDPR enforcement also had negative short-term effects on data-dependent technology ventures, indicating possible unintended impacts on competition and innovation (Jia, Jin & Wagman 2021)). These observations underscore the need for more research on the complex relationships between market concentration, data sharing, and privacy.

Another issue in this interface concerns the information-sharing network among data brokers themselves. A unique feature of data brokers is that they can obtain more pertinent information by sharing and selling information with each other, despite being competitors in the information market (Gu, Madio, & Reggiani, 2022). An FTC report (2014) highlights that seven out of nine data brokers engaged in data transactions with each other. For example, Acxiom had a partnership with Nielsen. This network shapes the competition dynamics among data brokers and influences the partnering organizations' decisions on which data broker to share information with.

To better understand the behavior of partnering organizations in this complex network given market competition dynamics, this paper attempts to answer the following research questions: (1) Do partnering organizations in a less concentrated industry share more customer information with data brokers? (2) Do partnering organizations share more information with data brokers who are in a less concentrated industry? This study also explores the consequences of information sharing, i.e., information security breaches, to provide further insights into the decision-making process of information sharing versus privacy protection when facing different levels of market competition dynamics.

Our study encompassed a two-fold analysis of customer privacy information leakage. First, we examine a longitudinal panel dataset from 2016 to 2021, which allows us to identify trends and examine the effects of past privacy information leakages, market concentration, and regulations over time. Subsequently, we conducted a deep-dive cross-sectional analysis using data from the dark web collected in 2020. This second analysis focuses on identifying instances of privacy information leakage, such as leaked emails, credit card information, and social security numbers. By synergizing these two analyses, we were able to construct a comprehensive understanding of the customer privacy information leakage landscape over time and its current state in the dark web.

Our findings indicate several key observations. Firstly, partners that share their customer information tend to provide a greater amount of data with data brokers that have significant market concentration, resulting in an increased number of data trackers associated with their operations. However, this trend is moderated when both partner's and data broker's market concentrations are high. Under such circumstances, partners with high market concentrations exhibit more limited data-sharing practices, even with dominant data brokers. Additionally, we observed that high market concentrations, both for partners and data brokers, contribute to decreased customer privacy breaches. On the other hand, firms with lower market concentrations face a higher risk of customer privacy breaches, particularly when coupled with lower data broker concentration.

Another significant finding is the substantial influence of a data broker's regulatory status on a partner’s data-sharing practices. Our research demonstrates that partners tend to share more data with registered data brokers, and this tendency is influenced by the level of market concentration. Furthermore, we discover that when both the partner firms and the data broker maintain high market concentration, and the data broker is registered, the likelihood of customer privacy leaks tends to decrease. These findings underscore the critical role played by the registration status of data brokers, in conjunction with their level of market concentration, in mitigating the risk of breaches in customer privacy.

Furthermore, our study, which utilizes data collected from the dark web, demonstrates that registered data brokers have a greatly lower amount of customer privacy information being leaked. These findings highlight the efficacy of regulations in preventing data breaches. The significant market share of registered data brokers reflects their proficiency in handling privacy risks and offers a secure avenue for companies to share data.

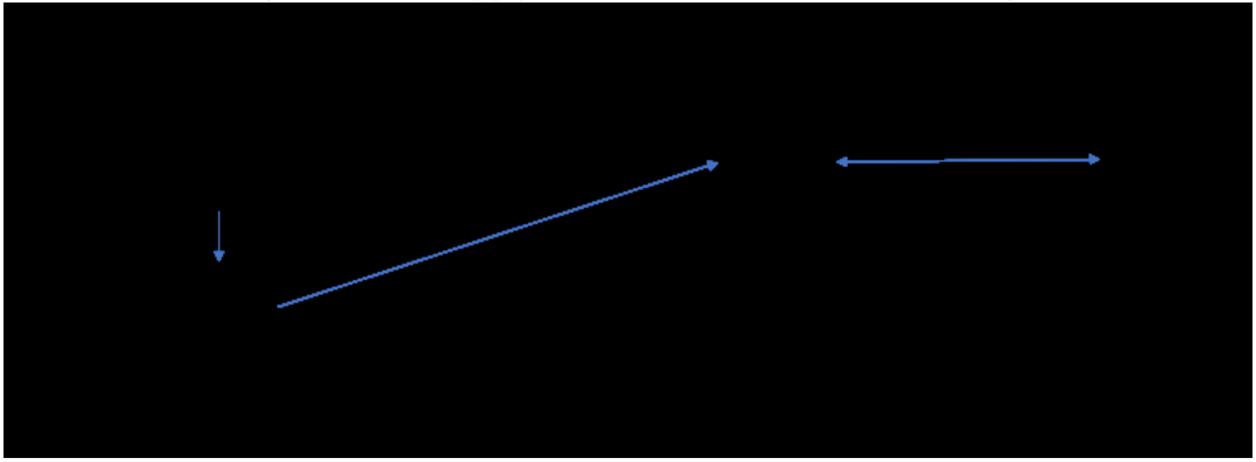
In the following sections of our study, we articulate our hypotheses, conduct an empirical analysis, and discuss our findings. We conclude by examining the policy implications of our research and proposing potential directions for future investigations in this rapidly evolving field.

2. Privacy, Market Concentration, Customer Data Sharing

Data Supply Chain

Data brokers occupy a pivotal position in the data supply chain, sourcing information from first-party data holders like Amazon and Bank of America, who maintain direct relationships with end consumers. Such first-party data holders typically operate under privacy policies that permit the sharing of personal information with their affiliates, and in some cases, non-affiliates, for the purpose of marketing and generating profit. Nevertheless, data brokers operate predominantly in the upstream market (Gu, Madio, & Reggiani, 2022; Martin, 2015a). Here, they aggregate data from a diverse range of sources, including these first-party data holders, transforming it into structured formats to distill valuable insights and actionable information (Braulin & Valleti, 2016).

Figure 1. Data Supply Chain and Customer Data Sharing



The data collected by data brokers is extensive and diverse, encompassing not only information directly shared by customers but also public records about individuals and data sold or licensed by first-party data holders (Glasgow, 2018). Data brokers utilize sophisticated techniques to analyze and organize this data, enabling them to build comprehensive profiles of individuals. These profiles include demographic information, consumer preferences, purchase history, online behavior, and other relevant data points. Such detailed profiles enable data brokers to categorize customers and place them into specific "buckets" based on their attributes, allowing for more targeted marketing efforts (FTC, 2014; Chandra & Kaiser, 2014; Goldfarb & Tucker, 2011).

Traditionally, data brokers have employed data trackers (e.g., third-party cookies) through their advertisement platform to track users' online activities across different websites and devices. However, due to growing privacy concerns, major browser companies like Google, Mozilla, and Apple have implemented changes to limit the trackers that data brokers implement. For example, Google announced plans to block the usage of third-party cookies by 2023. In response to these changes, data brokers have adapted their strategies. Data brokers now leverage certain first-party cookies, which were originally designed with the purpose of enhancing user experience and providing personalized functionality on websites.

These cookies are set by the website domain that the user is directly visiting, allowing the website to remember user preferences, store login information, and track session data. However, certain first-party cookies can be used for cross-site tracking when embedded in third-party contexts. For instance, social media log-in boxes or plugins, like those provided by Facebook, can be integrated into various websites to enable actions such as commenting or liking content (Ghosh, 2018). When users interact with these login widgets, first-party cookies are set by the widget's domain. Subsequently, these first-party cookies, functioning within a third-party context, can enable cross-site tracking.

This example highlights how first-party data holders (i.e., partners of the data brokers) voluntarily share customer information with data brokers. By implementing first-party cookies, partners actively contribute to the data supply chain, allowing data brokers to gather valuable insights and build comprehensive profiles of their own customers. While using first-party cookies presents opportunities for targeted marketing and enhanced personalization, it also underscores the importance of ensuring privacy and implementing appropriate safeguards to protect customer data.

Accordingly, data brokers and their partners form an essential link in the data supply chain, acquiring data from first-party data holders and utilizing it to provide valuable insights and targeted marketing opportunities. While this process offers benefits to businesses and marketers, privacy risks and breaches underscore the importance of implementing stringent privacy measures, adhering to regulations, and ensuring responsible data management practices throughout the data supply chain.

Market Concentration (Partners) and Customer Information Sharing

We scrutinize how the market structure influences the conduct of data sharing and the performance of privacy protection among partners. Market structure and concentration play a role in market outcomes (De Loecker, Eeckhout, and Unger 2020). The structure of the market, reflecting the level of competition and market concentration, plays an essential role in shaping the decisions and strategies related to data sharing among business partners. Our study aims to uncover the influence of market structure on data sharing and its significant welfare implications (Prüfer and Schottmüller, 2021).

Market concentration can play a major role in influencing decisions related to data sharing among business partners. These partners decide either harnessing the potential of their consumer data through partnerships with data brokers, thus enhancing performance and insights, or focusing on privacy protection as a means of fostering trust and developing substantial relationships. This decision-making process is substantially impacted by the market's competitive landscape (Marthews & Tucker, 2019; Soomro, Shah & Ahmed, 2016).

The existing body of literature provides insight into the far-reaching implications of market concentration and competition on data sharing practices (Jones & Mendelson, 2011; Savary & Parker, 1997). For instance, Sarvary and Parker (1997) draw attention to how the factors of competition and market concentration shape the tendencies of information sellers to distribute reports concerning unpredictable market conditions to other businesses. Essentially, the terms of data sharing and the propensity to engage in it are heavily dictated by the level of competition among the stakeholders involved.

Partners in a highly concentrated market are equipped to leverage their bargaining power to their advantage (Cook & Emerson, 1978). They are capable of negotiating agreements with data brokers that cater to their interests, granting them broader access to datasets. Notably, these arrangements also include a focus on heightened privacy protection, an approach to prevent customer data leakage. This strategic prioritization of trust and privacy in data sharing corresponds with the arguments put forth by Nissenbaum (2010) and Mayer-Schönberger and Cukier (2013) regarding the necessity of fostering robust privacy standards in data sharing agreements. By investing in and advocating for improved privacy standards, these market leaders can safeguard their interests and sustain their dominance.

Conversely, in markets with lower concentration, entities typically experience a reduction in bargaining power (Porter, 1979). This diminished influence often results in sub-optimal privacy outcomes in their data sharing decisions (Acquisti, Taylor & Wagman, 2016). Despite the hurdles in negotiating terms with data brokers, those entities with less dominance may still elect to share data. This strategic move allows partners to gain a competitive advantage and uncover new revenue streams (Manyika et al., 2011; Porter and Heppelmann, 2014). By sharing data, these businesses attempt to innovate and gain ground in their market, although they might be doing so at the risk of potential privacy breaches or sub-optimal data sharing agreements.

Data Sharing Conduct

We explore the significant impact of market concentration on data-sharing practices among partners. Through our analysis, we uncover how the decision to share data varies based on different levels of market concentration and, consequently, how these choices can lead to divergent privacy outcomes, which we will extensively discuss.

The level of market concentration significantly influences data-sharing practices among partners (Marthews & Tucker, 2019). In markets with high concentration, partners often have a variety of revenue channels, which encourages a more judicious stance regarding customer data sharing with data brokers (Acquisti, Taylor & Wagman, 2016). This caution originates from their ownership of vast, invaluable datasets, a product of their extensive user bases (Tirole, 1988; Mayer-Schönberger & Cukier, 2013). As a result, they might display a diminished inclination towards sharing customer information with data brokers (Schwab et al., 2011).

Conversely, organizations functioning in less concentrated markets might have limited revenue opportunities, which prompts them to view the sharing of customer data with data brokers as an alternative source of income (Manyika et al., 2011). Nonetheless, this approach necessitates forging alliances with large data brokers, demanding a more in-depth knowledge of their operational dynamics (Stahl, 2016; Tene & Polonetsky, 2012).

Table 1, for instance, gives examples of dominant data brokers, such as Google and Oracle, outlining how they disclose and utilize information collected from their partners (i.e., first-party data holders). Google, through its product Google AdSense, and Oracle both deploy first-party cookies as key elements of their data collection activities. Google AdSense primarily relies on first-party cookies when third-party cookies are inaccessible, enabling it to track and store user interactions and preferences on websites where their ads are displayed. Oracle uses first-party cookies to gather online information from individuals' activities on partner sites and third parties for cross-channel marketing. These strategies help Google and Oracle enhance their clients' first-party data, creating more comprehensive customer profiles.

Table 1. First-party Data Collection Activities by Data Brokers[2]

Data Broker	Purpose	Revenue Model
Google AdSense	▪ “Custom Search Ads (including AdSense for Search, AdSense for Shopping, and Programmable Search Engine) also uses a combination of first-party and third-party cookies. First party cookies are relied upon primarily when access to third party cookies is restricted, and are required to continue ad serving.”	▪ **Ad Revenue Multiplier** “Custom Search Ads is a Google product that lets you monetize the search results pages of your own search experience. If you don't already have a search experience on your site, consider adding an AdSense search engine, which can provide both a search experience and revenue from search ads.”
Oracle	▪ “Online information about you originates from your activities on sites operated by our online partners, such as advertising agencies and website operators … from third parties who may not have a relationship with you and who collect online information using cookies or similar technologies, such as pixels tags.”	▪ **Cross-channel Marketing** “Import DMP clients' user attributes into the Oracle Data Cloud platform and help them to leverage and enhance their first-party data for cross-channel marketing.

When partners share customer information, they essentially embark on one of two paths: either to generate revenue through what we refer to as the advertisement revenue multiplier model, or to deepen their understanding of their customers via the cross-channel marketing model, using additional attributes provided by data brokers.

[2] Google AdSense’s revenue model can be found at https://support.google.com/adsense/answer/7549925. Oracle’s revenue model is deribed at https://www.oracle.com/legal/privacy/advertising-privacy-policy.html#source.

Under the advertisement revenue multiplier model, websites providing advertising space can enhance their revenue by sharing customer information with marketers. They leverage the rich customer profiles provided by data brokers, thereby offering more targeted advertising opportunities to marketers. This leads to higher Cost-per-Click (CPC) rates. Alternatively, under the cross-channel marketing model, partners upload their customer information to the online platform of the data broker. The data broker then matches this data with their existing database, providing the partner with further customer insights. This broader perspective allows partners to form a more comprehensive understanding of their customers and devise more effective cross-channel marketing strategies.

Privacy Protection Outcomes

Market concentration is a substantial factor that impacts an entity's outcomes and influences its approach towards privacy protection. Firms in highly concentrated markets wield market power, empowering them with the resources necessary to invest in robust privacy protection measures (Armstrong, 2006). Two different theoretical lenses explain the relationship between privacy protection performance and market concentration.

The Resource-Based View (RBV) underlines the significance of unique capabilities and resources, such as data privacy protection, for a firm's performance (Barney, 1991). Firms dominating highly concentrated markets might thrive not merely due to their market dominance but also because of their distinctive abilities in managing and safeguarding data privacy. These capabilities could range from advanced algorithms and proprietary technologies to specialized human resources (Hitt, Ireland, & Hoskisson, 2014).

Simultaneously, the economics of privacy bring to light the intricacies of privacy protection and information sharing (Acquisti, Taylor & Wagman, 2016). The revelation is a complex interplay between the economic benefits that come with the disclosure of personal data and the significant detriments of inadequate data protection (Campbell, Goldfarb, and Tucker, 2015; Stone and Stone, 1990; Feri, Giannetti, and Jentzsch, 2016). Therefore, maintaining a delicate balance between privacy protection and information sharing becomes a critical consideration for all parties involved.

This nuanced understanding leads us to a critical intersection - that of market concentration and the resource-based view. Both elements play a crucial role in shaping a firm's approach to data privacy and its performance enhancement. Partners with more resource can assure customers and stakeholders of data safety, given their control over user data (Newbery, 1999). However, this concentration of control also raises potential concerns about data monopolization and possible privacy violations.

Yet, even in less concentrated markets, where resources for privacy protection might be stretched thin, privacy preservation is still a top priority (Baron, 2001; Nissenbaum, 2010). Despite their resource constraints, firms in these markets view privacy protection as an integral part of their non-price competitive strategy. By focusing on securing customer data and implementing rigorous privacy measures, these firms strive to build customer trust and enhance their competitive stance (Morgan & Hunt, 1994; Dinev & Hart, 2006). This universal acknowledgement of privacy preservation's importance underscores its significance, regardless of market concentration.

Our study is situated within this complex dynamic, drawing upon the existing literature to explore the intersection of competition and firm performance. Specifically, we investigate how the degree of market concentration affects partners' data collection and sharing practices. Every firm must navigate the delicate balance between safeguarding customer privacy and sharing customer data, in line with their competitive landscape (Leiponen, 2008). To provide an empirical examination of these dynamics, our primary research question probes the correlation between the market structure of cookie intermediaries and violations of customers' privacy (Goldfarb & Tucker, 2011; Mayer-Schönberger & Cukier, 2013).

Research Question 1: Within a more concentrated industry, do partners share more customer information with data brokers, and does this result in more significant customer information leakage?

Market Concentration (Data Broker) and Customer Information Sharing

Data brokers operate in a complicated landscape marked by competition and collaboration. Despite being competitors, data brokers often engage in collaborative relationships and information sharing practices, enriching their data pool, leveraging shared expertise, and enhancing their service offerings. This trend of collaboration, as outlined in the 2014 Federal Trade Commission (FTC) report, showcases that most data brokers engage in data transactions with one another, underlining the coexistence of competition and collaboration (FTC, 2014; Martin, 2015b).

In this market, the level of concentration significantly impacts data acquisition and sharing strategies. High market concentration can lead to power imbalances, potentially promoting anti-competitive behaviors such as data hoarding or monopolistic pricing (Lamdan, 2022; Puaschunder, 2021). Conversely, high market concentration can also foster efficiencies. For example, larger data brokers, due to their scale and resources, can streamline data exchange processes and lower transaction costs, enabling partner firms to manage operational expenses more effectively (Kim & Mahoney, 2005; Argyres & Liebeskind, 1999).

Drawing on Transaction Cost Economics (TCE), data brokers in a less concentrated market often strategically align and share customer information within their networks to reduce transaction costs and enhance their data assets (Ghosh, 2018; Ranganathan & Brown, 2006). By doing so, they are able to streamline data exchanges and lower management costs, thereby boosting their competitive advantage in the marketplace (Kim & Mahoney, 2005; Argyres & Liebeskind, 1999). The appeal of decreased transaction costs and the chance to tap into the network effect also motivate these data-sharing practices among competitors, leading to broader and more insightful data pools (Bakos & Brynjolfsson, 1999).

Contrastingly, in a highly concentrated market, dominant data brokers may be less inclined to share their data broadly within their networks. Following the RBV approach, these larger brokers may perceive their amassed data as a unique, strategic resource that offers a competitive edge and differentiates them in the market (Barney, 1991; Wernerfelt, 1984). As a result, they may tend towards retaining their proprietary data rather than risking its dilution through wide dissemination.

It is important to note that while sharing customer data can lead to cost efficiencies and potential benefits, it also comes with the inherent risk of consumer privacy breaches. In both scenarios, whether it's in a less concentrated market where data is more widely shared or in a more concentrated market where data is viewed as a unique resource, consumer information could potentially become dispersed among multiple entities, escalating the risk of privacy leakage (Glasgow, 2018). Thus, the level of market concentration in the data broker industry remains a critical factor in the discussion of consumer privacy risk.

Research Question 2: Does data broker market concentration affect partner information sharing and customer information leakage?

Data Brokers and Regulation

The ascendancy of data brokers in the data economy has triggered significant concerns around consumer privacy and data protection. These concerns are primarily attributed to the opacity of data broker operations and the potential misuse of consumer data (Marthews & Tucker, 2019;

Goldfarb & Tucker, 2011; Adjerid et al., 2016). Consequently, there are increasing calls for more stringent regulation of the data broker industry. Additionally, studies suggest that data brokers are prone to data breaches, leading to the potential distribution of consumers' personal information on the dark web (Ponemon Institute, 2014). This evidence underscores the need for amplified regulation to ensure the security of consumer data.

In an effort to enhance regulatory effectiveness and foster transparency in the data broker industry, the Federal Trade Commission (FTC) has recommended enforcing the disclosure of data collection practices and granting consumers access to their data (FTC, 2014). More recently, in a 2021 report to Congress, the FTC has proposed legislation that would broaden its regulatory authority concerning privacy and data security. Furthermore, two states, Vermont and California, have taken strides to regulate data brokers. Vermont implemented Data Broker Regulations in 2018 (9 VSA § 2430) requiring data brokers to annually register with the Secretary of State. California also compels data brokers to disclose information on their websites. In Vermont, a 'data broker' pertains to any business or entity that knowingly collects, sells, or licenses personal information of consumers with whom the business has no direct relationship. In California, the term is defined as a business that knowingly collects and sells personal information of consumers without establishing a direct relationship (Cal. Civ. Code § 1798.99.80).

Despite these efforts, the lack of enforcement actions and negligible penalties for non-registration pose challenges to the regulation of data brokers. Under California law, registered data brokers are obligated to issue a pre-collection notice before selling or sharing any collected personal information. However, registration is often seen as a strategic means for data brokers to obtain an exemption from issuing pre-collection notices, thereby increasing the likelihood of extensive data sharing activities and potential privacy breaches.

In our study, we pose a research question to investigate whether partners are more likely to share customer information with registered data brokers as opposed to unregistered ones. This query stems from the assumption that registered data brokers, who are required to disclose their data collection practices and provide consumers access to their data, appear to be more transparent and accountable entities. Conversely, unregistered data brokers may lack such transparency and accountability measures, potentially making them less appealing partners for firms intent on protecting their customers' privacy. Thus, we propose the following research question to further scrutinize the data sharing behaviors of firms in relation to registered and unregistered data brokers:

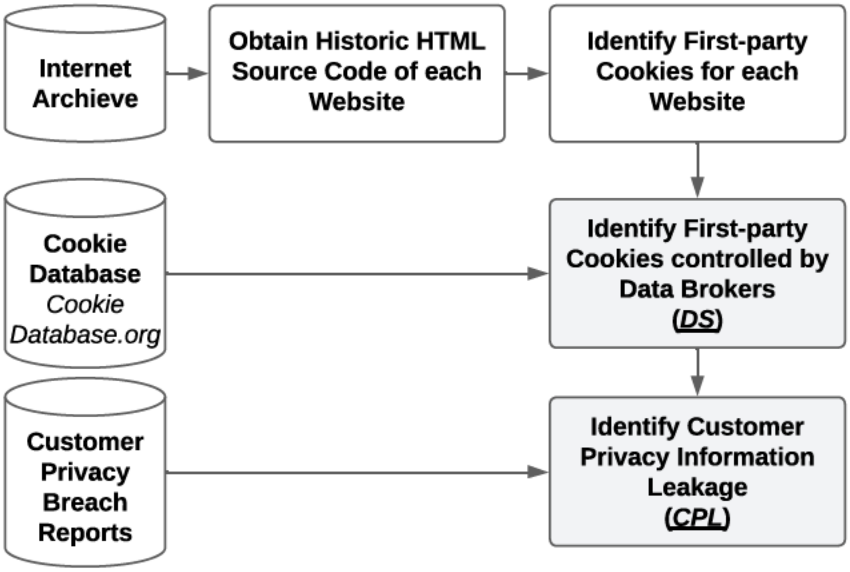
Research Question 3: What is the relationship between market concentration and customer information leakage when sharing data with registered data brokers versus unregistered data brokers?

Data and Measurement

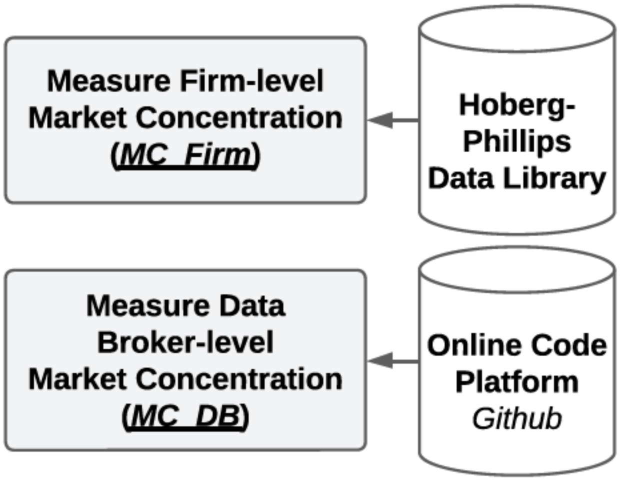
In our empirical analysis, we utilize multiple unique data sources. Figure 2 illustrates our data collection and variable measurement process.

Figure 2. Data Collection and Variable Measurement

Panel A. Measuring Data Sharing Activities and Customer Privacy Information Leakage


Internet Archieve
Obtain Historic HTML Source Code of each Website
Identify First-party Cookies for each Website
Cookie Database
Cookie Database.org
Identify First-party Cookies controlled by Data Brokers
(*DS*)
Customer Privacy Breach Reports
Identify Customer Privacy Information Leakage
(*CPL*)

Panel A. Measuring Firm-level and Data Broker-level Market Concentration


Measure Firm-level Market Concentration (MC_Firm)
Hoberg-Phillips Data Library
Measure Data Broker-level Market Concentration (MC_DB)
Online Code Platform Github

Data Sharing Activities

First, we identify 1,046 U.S. public firms where websites are captured and stored by *Internet Archive*, which websites actively engage in e-commerce. *Internet Archive* primarily archives and backs up websites with significant traffic and active e-commerce activities. This allows us to not only observe previous versions of websites but also retrieve the initial source code of websites that may no longer be directly accessible to the public. Next, we exclude firms that do not have any financial information provided by *Compustat* and market concentration measured by Hoberg and Philips (2016). Our final sample includes 470 unique firms that results in 2,206 firm-year observation between 2016 and 2021.

Table 2. Sample Selection

	Number of Firm-year Observations	*Number of Firms*
Firms with Available Website URL Address	**51,608**	**8,803**
Less: No historical websites available in Archieve.org	(45,820)	(7,757)
Less: No market concentration measure available from Roberg and Philip's Database	(2,918)	(471)
Less: No financial information available from Compustat	(664)	(105)
Total Sample	**2,206**	**470**

Next, we retrieve the list of first-party marketing cookies and the data brokers that controls information collected by those cookies from *CookieDatabase.org*. Each cookie that a user encounters on a website can provide information to third-party organizations that are listed in *CookieDatabase.org*. These organizations are commonly referred to as data brokers, who are deemed as "data controllers" due to their exclusive control over the information collected from users through first-party cookies. The data collected from first-party cookies are often encrypted using techniques such as hashing before being sent to the data brokers' servers via secure channels. In our study, we define these third-party organizations identified in *CookieDatabase.org* as data brokers (both registered and unregistered).

We count the number of first-party marketing and analytic tracking cookies that are implemented on each firm's website, which is labeled as *DS*. After reviewing our list, we have identified the majority of the advertisement firms, such as Kentico, and e-commerce data analytics service providers like Shopify. We have excluded data brokers based on their highest and lowest *DS*, eliminating the top 2.5%, and those with names that consist of commonly used dictionary words. As a result of this filtering process, we have obtained a final sample of 75 data brokers.

Customer Privacy Breaches

In our study, we investigate the occurrence of customer privacy information leakage (*CPL*) resulting from data-sharing practices. To accurately measure *CPL*, we utilize data breach reports sourced from *Audit Analytics*, a reputable provider of comprehensive cybersecurity incident information. We specifically capture instances of customer privacy information leakage. *Audit Analytics* collects breach reports from multiple reliable sources, including the U.S. Department of Health and Human Services, as well as the State Offices of the Attorney General in Washington, Oregon, and California. From our sample of 470 firms, we have identified that 47 of them have experienced at least one instance of customer privacy information leakage.

Market Concentration between Industry and Data Brokers

In order to measure the level of market concentration in different industries, we utilize Hoberg and Philips' (2016) Text-based Network Industry Concentration (TNIC) Data, which provides the Herfindahl-Hirschman Index (*HHI*) for each industry. This measure is a more current and relevant measure of industry concentration as it is updated yearly based on the business description disclosed in Item 1 of the 10-K filing. The resulting measure of market concentration, labeled *MC_Firm*, ranges from 0 to 1 in the sample, with a median of 0.195. Descriptive statistics are presented in Table 3. The TNIC-HHI measure is used to proxy the level of competition faced by the firms in the respective industries.

Table 3. Descriptive Statistics

Panel A. Variable Description

Variable	Description
Partner-level	
CPL	Counts of customer privacy leakage incidents in a fiscal year period.
DS	Number of data trackers controlled by third-party data brokers.
MC_Firm	Market concentration measured by Hoberg and Philips (2016)
MC_DB_Average	Average level of data broker's market concentration measured for *Github*
Reg	Data Broker Registration in California or Vermont (1 if the partner shares with at least one registered data brokers, 0 if not)
NI	Logarithmic value of total revenue
AT	Logarithmic value of total Assets
Intan	Logarithmic value of intangible assets
ROA	Return on assets
ROE	Return on Equity
Neg	Negative income (1 if negative income, 0 otherwise)
Emp	Logarithmic value of the number of employees
Data Broker level	
MC_DB	Market concentration measured from *Github*
Reg	Data Broker Registration in California or Vermont (1 if registered, 0 if not)
CPL_P	Counts of user account information leakage in the dark web
CPL_C	Counts of credit card information leakage in the dark web

CPL_S	Counts of social security number leakage in the dark web
Visit	Monthly average visits to site including desktop and mobile web measured by *SemRush*
Past_Breach	Past cybersecurity breach history before 2021
EU	Data Broker headquartered in EU member nations (1 if EU member, 0 otherwise)
Rev	Estimated revenue range for both private and public firms estimated by *Crunchbase*
Age	Data broker firm age

Panel B. Descriptive Statistics

Variable	**Count**	**Mean**	**Sd/ Mean**	**Min**	**1st Quartile**	**Median**	**3rd Quartile**	**Max**
CPL	2,206	0.028	0.179	0.000	0.000	0.000	0.000	2.000
DS	2,206	0.130	0.609	0.000	0.000	0.000	0.000	8.000
MC_Firm	2,206	0.331	0.291	0.020	0.110	0.223	0.466	1.000
MC_DB_Avg	2,206	0.005	0.049	0.000	0.000	0.000	0.000	1.000
Reg	2,206	0.086	0.281	0.000	0.000	0.000	0.000	1.000
NI	2,206	4.265	2.126	-5.521	2.826	4.354	5.690	10.29
AT	2,206	7.082	2.198	0.646	5.531	7.255	8.689	13.70
Intan	2,206	5.024	2.919	-6.215	3.032	5.320	7.271	12.64
ROA	2,206	-0.051	0.406	-12.853	-0.045	0.025	0.071	1.121
ROE	2,206	-0.080	0.532	-11.007	-0.043	0.021	0.053	7.517
Neg	2,206	0.376	0.485	0.000	0.000	0.000	1.000	1.000
Emp	2,206	13.92	37.80	0.000	0.416	2.903	11.005	543.0
CPL_P (Thousands)	57	2.450	14.158	0.000	0.000	0.000	0.000	105.0
CPL_C	57	226.5	1,479	0.000	0.000	0.000	0.000	1,109
CPL_S	57	10.52	62.44	0.000	0.000	0.000	0.000	454.0
Visit (Millions)	57	0.137	505.6	0.000	0.093	0.920	37.69	3,313
MD_DB	57	0.183	0.285	0.000	0.004	0.027	0.248	1.000
Past_Breach	57	0.157	0.367	0.000	0.000	0.000	0.000	1.000
EU	57	0.210	0.411	0.000	0.000	0.000	0.000	1.000
Reg	57	0.140	0.350	0.000	0.000	0.000	0.000	1.000
Rev	57	2.596	1.962	0.000	1.000	2.000	4.000	7.000
age	57	13.701	6.298	4.000	10.00	13.00	16.00	46.00

Note: For our analysis on the darkweb, we focus on 57 data brokers that have financial information provided by Crunchbase.

To estimate the market concentration level of data brokers, we suggest the measure of market share of each data broker, which is referred to as *MC_DB*. We count the number of code repositories on *Github* that mentions the name of the data brokers to calculate this measure. *Github* is a leading platform where developers and companies create, develop and maintain software. We use the Search API provided by *Github* to identify the number of topics that mention the name of the data broker in both public and private repositories. *MC_DB* measures the relative market share of data brokers, where a higher value of *MC_DB* indicates a larger market share of data brokers in the data broker industry. The study found that *MC_DB* varies between 0 to 1 in their sample, with a median of 0.027. Furthermore, we construct *MC_DB_Avg* as the firm-level average of *MC_DB* specifically for those data brokers with whom they share data.

In addition to our main effect variables, we consider firm-specific variables, such as the logarithm amount of net income (*NI*) and total assets (*AT*), performance measures such as Return on Assets (*ROA*) and Return on Equity (*ROE*), negative income (*Neg*), and the age of the firm (*Age*). In alignment with the methodologies used in preceding studies examining the impact of firm-level characteristics on cybersecurity policy (Gordon et al., 2010) and the consequences of cybersecurity (Wang et al., 2013), we have chosen to control the firm-specific effects in our research.

Panel B of Table 3 provides the descriptive statistics of our study. The *CPL* is generally low, albeit with a few notable exceptions where it reaches up to a value of 2. *DS* is also typically limited, though there is a substantial range in the data, peaking at a value of 8. *MC_Firm* hovers around one-third, indicating a moderate level of market concentration across the firms in the study. However, some firms have reached a maximum value of 1, pointing towards a higher market concentration in certain instances. *MC_DB_Avg* is markedly low, suggesting a lack of dominant power among data brokers in the market. *Reg*, representing the registration status of data brokers, unveils that only a small percentage of data brokers are registered, with a mean value close to 0.086. This implies that the majority of data brokers, with whom firms share customer data in our study, operate without formal registration.

Dark Web Analysis

As part of our investigation, we perform a cross-sectional analysis to explore and quantify the degree of customer privacy information leakage within the dark web. Our focus is on uncovering Personally Identifiable Information (PII) associated with each data broker operating in this covert online space. For this study, we utilized an extensive dataset of 297,935 darknet market posts collected from 714 distinct websites. This data was sourced from DarkOwl, a leading dark web monitoring service provider, over a span of six months, from February 1 to July 31, 2020.

We processed the posts, filtering out those identified to contain PII or malicious keywords. Utilizing textual analysis, we homed in on posts that mentioned the data broker's name and contained specific user information, such as email IDs and passwords. For example, posts that listed a data broker's name alongside leaked email addresses and respective passwords were singled out. Each unique email-password pairing was documented as an individual instance of consumer privacy leakage, denoted as *CPL_P*. In addition, we differentiated and quantified instances of credit card information leakage (*CPL_C*) and social security number leakage (*CPL_S*), creating a comprehensive view of customer privacy information leaked on the dark web.

We incorporate several control variables into our model. *Visit* denotes the average number of monthly visits to a data broker's website, providing a measure of the broker's visibility or popularity provided by SEMrush. *Past_Breach* is a binary variable indicating whether the broker has experienced any cybersecurity breaches before 2021, thereby reflecting the broker's historical security performanceEstimated revenue (*Rev*) and *Age*, signifies the financial capacity of a firm, which could impact its investment in cybersecurity infrastructure and consequently the possibility of data breaches.

The descriptive statistics provided in Panel B of Table 3 show that, on average, there have been 2.45 thousand instances of personal identifier leakage, 226.5 cases of credit card information leakage, and 10.52 instances of social security number leakage, even though the majority of firms show no leakage. The average website traffic stands at 137 thousand visits per month, but this number ranges dramatically up to 3.3 billion visits. The firms in the dataset, on average, have a low market dominance, with a prior history of cybersecurity breaches in about 15.7% of the cases. The firms have an average estimated revenue level of approximately 2.6, and they have been operating for around 13.7 years on average.

3. Analysis and Results

The Effect of Market Concentration on Data Sharing and Customer Privacy Information Leakage

The first research question aims to determine whether partners in less concentrated industries share more information with data brokers than those in more concentrated industries. Our second research question is related to the first and investigates whether firms in less concentrated markets are more likely to experience privacy breaches when they share data through first-party cookies.

We employ a two-stage least squares (2SLS) framework with a control function approach to address endogeneity and omitted variable bias. In the first stage (2SLS-1), the estimation centers around capturing the relationship between the endogenous variable, *DS* (number of data trackers controlled by third-party data brokers), and the exogenous variables, including market concentration (*MC_Firm* and *MC_DB_Avg*) for each firm *i* in year *t*. Notably, the inclusion of the interaction term (*MC_Firm*×*MC_DB_Avg*) allows for the examination of the joint impact of market concentration on *DS*.

In the second stage (2SLS-2), the analysis shifts to investigating the impact of *DS*, instrumented in the first stage, on the dependent variable *CPL* (counts of annual customer privacy leakage incidents). Particular attention is given to the coefficient of the interaction term, *MC_Firm×MC_DB_Avg*. By including it in the second stage equation, the model ensures the comprehensive examination of the joint influence of market concentration on *CPL*.

$$DS_{it} = \alpha_0 MC_{Firm_{it}} + \alpha_1 MC_{DB_{Avg_{it}}} + \alpha_2 MC_{Firm_{it}} \times \ MC_{DB_{Avg}}{}_{it} + \ \sum\alpha Control\ Variables_{it} + \sum\alpha Fixed\ Effectsi_t + \epsilon_{it}. \qquad (2SLS - 1)$$

$$CPL_{it} = \beta_0 DS_{it} + \beta_1 MC_{Firm_{it}} \times \ MC_{DB_{Avg}}{}_{it} + \ \sum\beta Control\ Variables_{it} + \sum\beta Fixed\ Effects_{it} + \mu_{it}. \qquad (2SLS - 2)$$

The first column in Table 4 presents the results of the first-stage model as an OLS regression alone, while the second column corresponds to the results obtained using an entropy-balanced sample. The results demonstrate a positive and statistically significant relationship between data broker's market concentration (*MC_DB_Avg*) and the number of data trackers controlled by third-party data brokers (*DS*) among firms engaged in data sharing with data brokers. This indicates that partners tend to share more information with data brokers that operate in more concentrated markets. In other words, partners are more likely to engage in data sharing with giant data brokers, resulting in a higher number of data trackers associated with their activities.

Table 4. Market Concentration and Data Sharing

Dependent Variable	**(1) *DS***	**(2) *DS***	**(3) *DS***	**(4) *DS***
MC_Firm	-0.010	0.026	-0.018	0.006
	(-0.41)	(0.20)	(-1.22)	(0.18)
MC_DB_Avg	15.622^{***}	11.219^{***}	12.465^{***}	10.157^{***}
	(8.77)	(15.24)	(6.64)	(18.61)
MC_Firm ×MC_DB_Avg	-9.838^{***}	-3.623^{***}	-8.004^{***}	-3.777^{***}
	(-7.27)	(-4.54)	(-5.12)	-5.09
Reg			1.017^{***}	-0.980^{***}
			(10.06)	(14.70)

Dependent Variable	(1) ***DS***	(2) ***DS***	(3) ***DS***	(4) ***DS***
MC_Firm × MC_DB_Avg × Reg			0.084	0.167
			(0.79)	(1.01)
NI	-0.010	-0.033	-0.802	-2.899
	(-2.14)	(-1.05)	(-2.13)	(-1.21)
AT	-0.028	0.042	-1.963	3.351*
	(-0.80)	(1.26)	(-1.14)	(1.79)
Intan	0.017	-0.015	0.880	-0.320
	(1.49)	(-0.50)	(1.10)	(-0.27)
ROA	-0.004	-0.072	-0.503	-4.961
	(-0.42)	(-0.42)	(-0.69)	(-0.40)
ROE	-0.006	0.044	-0.108	4.681
	(-1.20)	(0.85)	(-0.50)	(1.26)
Neg	-0.008	-0.013	-1.789	-0.125
	(-0.71)	(-0.18)	(-1.06)	(-0.02)
Emp	0.001	0.000	0.042	0.002
	(0.371)	(0.25)	(0.89)	(0.05)
Constant	0.227	0.467*	0.131	0.018
	(0.85)	(1.83)	(0.82)	(0.13)
Fixed Effects				
Firm	Yes	No	Yes	No
Year	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes
Observations	2,206	2,206	2,206	2,206
R-squared	0.700	0.751	0.908	0.900

Note: ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. *t*-statistics are reported in parentheses.

Furthermore, the negative coefficient of the interaction term (*MC_Firm* × *MC_DB_Avg*) reveals a significant moderating effect of market concentration (*MC_Firm*) on the relationship between *MC_DB_Avg* and *DS*. Specifically, when both market concentration and data broker concentration are high, the joint impact suggests that partners with high market concentration level share less data with third-party data brokers. This implies that in markets with high concentration levels, partners may be more reluctant about sharing their data with data brokers, even if the data brokers themselves have high market concentration. These results shed light on the complex dynamics between market concentration, data sharing, and the control of data trackers by third-party data brokers among firms.

Panel A in Table 5 presents the results of our first-stage regression, mirroring the findings presented in Table 5, with data sharing (*DS*) serving as the dependent variable. In contrast, Panel B delineates the findings from our second-stage regression, wherein the dependent variable is customer privacy leakage (*CPL*). Here, we observe a significant negative coefficient for the interaction term (*MC_Firm* × *MC_DB_Avg*), implying a reduction in customer privacy leakage when both the firm and the data brokers possess high market concentrations. Although the *DS* variable does not exhibit significance, it is noteworthy that *MC_DB_Avg* exceeds zero solely when the firm engages in data sharing with data brokers.

Table 5. Market Concentration, Data Sharing, and Customer Privacy Information Leakage

Panel A. First-stage regression

Dependent Variable	**(1) *DS***	**(2) *DS***
MC_Firm	-0.000	-0.001
	(-0.00)	(-0.20)
MC_DB	12.217^{***}	10.314^{***}
	(13.82)	(16.76)
$MC\_Firm \times MC\_DB$	-4.307^{***}	-4.254^{***}
	(-4.62)	(-5.32)
Reg		0.983^{***}
		(15.65)
$MC\_Firm \times MC\_DB \times Reg$		0.169
		(0.99)
Constant	0.227	-0.010
	(0.85)	(-0.45)
Fixed Effects (Firm / Year / Indsutry)	Yes	Yes
Observations	2,206	2,206

Panel B. Second-stage regression

Dependent Variable	**(1) *CPL***	**(2) *CPL***
DS	0.000	0.005
	(0.08)	(0.94)
MC_Firm × *MC_DB_Avg*	-0.067^{***}	-0.070
	(-2.63)	(-0.82)
MC_Firm × *MC_DB_Avg* × *Reg*		-0.402^{***}
		(-3.44)
NI	-0.338	-0.339
	(-0.83)	(-0.83)
AT	0.262	0.253
	(0.76)	(0.74)
Intan	0.292	0.294
	(1.32)	(1.33)
ROA	-0.073	-0.059
	(-0.15)	(-0.12)
ROE	0.159	0.151
	(0.53)	(0.50)
Neg	0.686	0.699
	(0.84)	(0.85)
Emp	0.074^{***}	0.074^{***}
	(2.63)	(2.64)
Fixed Effects (Firm / Year / Indsutry)	Yes	Yes
Constant	-0.003	-0.002
	(-0.21)	(-0.15)
Observations	2,206	2,206
Adj R-squared	0.032	0.032

Note: ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. *t*-statistics are reported in parentheses.

To delve deeper into the interaction effects between the market concentrations of data brokers and partners on customer privacy information leakage (*CPL*), we proceed with a mean comparison analysis. Table 6 presents the results of the marginal analysis conducted to examine the variations in the mean customer privacy information leakage (*CPL*) based on different levels of market concentration. In this analysis, market concentration measures *MC_DB_Avg* and *MC_Firm* are categorized into high and low groups, considering values higher than the mean.

Table 6. Margin Analysis on Data Sharing and Customer Privacy Information Leakage

Dependent Variable (*DS*) (*N* = 2,206)		**(1) Margin**	**(2) Std. Err**
MC_Firm			
Low		0.068^{**}	0.032
		(2.12)	
High		0.005	0.043
		(0.12)	
MC_DB_Avg			
Low		0.044	0.030
		(1.49)	
High		0.029	0.444
		(0.77)	
MC_Firm	***MC_DB_Avg***		
Low	*Low*	0.070^{**}	0.034
		(2.12)	
High	*Low*	0.003	0.045
		(0.08)	
Low	*High*	0.020	0.043
		(0.46)	
High	*High*	0.043	0.062
		(0.70)	

Note: ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. *t*-statistics are reported in parentheses.

The null hypothesis in the marginal analysis is that the customer privacy information leakage (*CPL*) measure is not equal to zero, which represents a higher amount of leakage. By conducting the analysis, we aim to examine whether there are significant differences in the mean *CPL* across different levels of market concentration. The results reveal that firms with lower market concentration (*MC_Firm*) exhibit a significantly higher mean *CPL* compared to zero. This suggests a substantial presence of customer privacy information leakage in firms characterized by lower market concentration.

Furthermore, our investigation of the interaction between lower market concentration (*MC_Firm*) and lower data broker concentration (*MC_DB_Avg*) provides additional insights. Among firms that demonstrate both lower market concentration and lower data broker concentration, the mean *CPL* is significantly higher compared to zero. This specific subgroup of firms shows a notable occurrence of customer privacy information leakage.

These findings underscore the significance of lower market concentration, particularly when combined with lower data broker concentration, in contributing to a higher likelihood of customer privacy breaches. Firms operating with a lower market concentration and limited engagement with data brokers are more prone to experiencing customer privacy information leakage.

Data Broker Registration and Customer Privacy Information Leakage

In this section, we probe the role of market concentration - both at the firm level and among data brokers - along with the regulatory environment, and their impact on the quantity of customer data shared (*DS*) and ensuing leakage of customer privacy (*CPL*). We again adopt a two-stage least squares (2SLS) regression model to address potential endogeneity issues.

In our first-stage model (2SLS-1), we incorporate the data broker registration status (*Reg*) variable. To explore potential interactions between these variables, we introduce corresponding interaction terms, thereby enabling us to scrutinize how the effects of market concentration might differ depending on data broker registration status, and vice versa. In the second-stage model (2SLS-2), customer privacy leakage (*CPL*) is characterized as a function of the predicted data sharing (*DS*) obtained from the first stage, as well as the interaction terms that reflect the market concentration levels of data brokers, partners, and the data broker registration status.

$$DS_{it} = \alpha_0 MC_{Firm_{it}} + \alpha_1 MC_{DB_{Avg_{it}}} + \alpha_2 Reg_{it} + \alpha_3 MC_{Firm_{it}} \times MC_{DB_{Avg}}{}_{it} + \alpha_4 MC_{Firm_{it}} \times MC_{DB_{Avg}}{}_{it} \times Reg_{it} + \sum \alpha Control\ Variables_{it} + \sum \alpha Fixed\ Effectsi_t + \epsilon_{it}. \quad (2SLS - 1)$$

$$CPL_{it} = \beta_0 DS_{it} + \beta_1 MC_{Firm_{it}} \times MC_{DB_{Avg}}{}_{it} + \alpha_4 MC_{Firm_{it}} \times MC_{DB_{Avg}}{}_{it} \times Reg_{it} + \sum \beta Control\ Variables_{it} \quad + \sum \beta Fixed\ Effects_{it} + \mu_{it}. \quad (2SLS - 2)$$

Table 4, specifically columns 3 and 4, delve into the determinants that shape a firm's decision to engage in data-sharing practices, with an emphasis on the impact of data brokers' registration status and their level of market concentration. Column 3 presents the results derived from the standard model, while Column 4 illustrates the findings from an entropy-balanced sample.

The variable of interest is the *Reg* coefficient, which represents data brokers' registration status. A positive and statistically significant coefficient in both columns (3) and (4) suggests that, all else being equal, firms are more likely to share customer information with registered data brokers as compared to unregistered ones. This result holds even after entropy balancing in column (4), reinforcing the robustness of this finding.

Importantly, the coefficients for the interaction terms *MC_Firm* × *MC_DB_Avg* × *Reg* are not statistically significant in both columns (3) and (4). This indicates that the registration status of data brokers doesn't significantly alter the relationship between market concentration (both of the firm and of data brokers) and data sharing. In other words, firms prefer to share data with registered data brokers, irrespective of the market concentration levels of either the firm itself or the data brokers.

This result could be potentially explained by registered data brokers being perceived as more trustworthy or accountable. The registration requirement implies adherence to certain standards and regulations, which could enhance their appeal to firms, particularly when dealing with sensitive customer information. This finding aligns with our understanding of firms prioritizing customer privacy and potentially avoiding partnership with less transparent, unregistered data brokers.

The first-stage regression results in Table 5 are aligned with our previous findings shown in columns 3 and 4 of Table 4, which implies consistency in our data analysis process. Turning our focus to the second-stage regression results in the second column of Panel B, the three-way interaction term (*MC_Firm* × *MC_DB_Avg* × *Reg*) is negative and statistically significant value (-0.402). This suggests that when both the firm and the data broker have high market concentration and the data broker is registered, customer privacy leakage tends to decrease. This indicates that the registration status of data brokers, combined with market concentration levels, can play a critical role in mitigating customer privacy leakage.

In sum, our findings suggest that in more concentrated markets, firms tend to share more data with dominant data brokers, thereby increasing the number of data trackers associated with their operations. Interestingly, this trend is moderated when both market and data broker concentrations are high. Under such conditions, firms with high market concentrations become more conservative in their data-sharing practices, even with dominant data brokers. We also find that high market concentrations, for both firms and data brokers, result in a decrease in customer privacy breaches. Conversely, firms with lower market concentrations face an elevated risk of customer privacy breaches, especially when paired with lower data broker concentration.

Another notable insight is the significant influence of a data broker's regulatory status on firms' data-sharing practices. We found that firms share more data with registered data brokers, and this tendency is modulated by the level of market concentration. We also discover that when both the company and the data broker maintain high market concentration, and the data broker is registered, the likelihood of customer privacy leaks tends to decrease. This finding emphasizes that the registration status of data brokers, combined with their level of market concentration, plays a crucial role in mitigating the risk of breaches in customer privacy.

Data Broker Information Leakage on the Dark Web

Table 7 presents the results of customer privacy information leakage on the dark web. This is gauged through different measures: *P* (*CPL_P*), *P_C* (Sum of *CPL_P* and *CPL_C*), and P_C_S (Sum of *CPL_P*, *CPL_C,* and *CPL_S*). Columns 1-3 display the results without considering the regulatory effects, while Columns 4-6 incorporate the impact of GDPR and state-level data broker registration in the United States (specifically, in Vermont and California).

Table 7. Customer Privacy Information Leakage in the Dark Web

Dependent Variable (*CPL*)	**(1) *P***	**(2) *P_C***	**(3) *P_C_S***	**(4) *P***	**(5) *P_C***	**(6) *P_C_S***
MC_DB	0.062	0.062	0.062	0.003	0.003	0.003
	(0.47)	(0.47)	(0.47)	(0.07)	(0.07)	(0.07)
Past_Breach	0.844***	0.845***	0.845***	-0.041	-0.040	-0.040
	(4.83)	(4.83)	(4.83)	(-0.45)	(-0.44)	(-0.44)
EU	0.119*	0.119*	0.119*	-0.070**	-0.070**	-0.070**
	(1.86)	(1.86)	(1.86)	(-2.42)	(-2.42)	(-2.42)
Reg	0.106	0.106	0.106	-0.020	-0.020	-0.020
	(1.50)	(1.50)	(1.50)	(-0.49)	(-0.49)	(-0.49)
MC_DB × *Past_Breach*	-1.029***	-1.029***	-1.029***	0.179	0.178	0.178
	(-3.43)	(-3.43)	(-3.43)	(1.23)	(1.23)	(1.23)
MC_DB × *Past_Breach* × *EU*				6.057***	6.055***	6.055***
				(15.33)	(15.33)	(15.33)
MC_DB × *Past_Breach* × *Reg*				3.953***	3.954***	3.954***
				(3.03)	(3.03)	(3.03)
Rev	0.011	0.011	0.011	-0.005	-0.005	-0.005
	(0.78)	(0.78)	(0.78)	(-0.94)	(-0.94)	(-0.94)
Age	-0.006	-0.006	-0.006	-0.007	-0.007	-0.007
	(-1.25)	(-1.25)	(-1.25)	(-3.52)	(-3.52)	(-3.53)
Constant	0.008	0.008	0.008	0.105***	0.105***	0.105***

Dependent Variable (*CPL*)	**(1) *P***	**(2) *P_C***	**(3) *P_C_S***	**(4) *P***	**(5) *P_C***	**(6) *P_C_S***
	(0.12)	(0.12)	(0.12)	(3.49)	(3.49)	(3.49)
Observations	57	57	57	57	57	57
R-squared	0.332	0.332	0.333	0.887	0.887	0.887

Note: ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. *t*-statistics are reported in parentheses.

In columns 1 to 3, it is clear that companies with a history of data breaches typically experience an increase in customer privacy leakages (*CPL*) on the dark web. Significantly though, our data also shows that firms commanding larger market concentration experience a decrease in *CPL*, even in the context of past breaches, which is represented as a negative and significant coefficient of the interaction term (*MC_DB* × *Past_Breach*). This aligns with our previous analysis, suggesting that greater market dominance may potentially buffer companies against the negative impacts of cybersecurity incidents, thus mitigating the extent of customer information leakage in the dark web.

In order to consider regulatory impacts, we incorporated an additional variable, *EU*, which indicates whether a data broker is headquartered in the European Union. Being based in the EU brings these companies under the scope of Article 3(1) of GDPR. According to this rule, the regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, irrespective of whether the processing takes place within the Union or not (EU GDPR, 2018). As a result, data brokers headquartered in the EU must adhere to this comprehensive regulation, signifying a more stringent regulatory framework for data handling and privacy protection (although GDPR reaches global firms as well, the preoccupation with GDPR may be more significant for European firms).

From our result in columns 4 to 6, *Reg* has a positive but statistically insignificant coefficient, suggesting that it does not have a significant impact on customer privacy leakage measures. Similarly, the inclusion of *EU* does not yield statistically significant results. However, when examining the interaction of market concentration, past breaches, and *REG* (and also *EU*), the coefficient is highly statistically significant and positive. This suggests that where data brokers, especially those not officially registered as data brokers, are likely to see more significant customer privacy leaks, based on dark web data. Essentially, it shows that higher market concentration combined with a past history of breaches, without the protective effect of broker registration, correlates strongly with an increase in customer privacy information leakage on the dark web.

4. Conclusion

In our interconnected world, the value of data is undeniable as businesses rely on it to fuel targeted advertising and drive revenue. However, this exchange of data raises concerns about privacy breaches and the delicate balance between financial growth and customer privacy. Our study delves into this issue, focusing on how market competition influences data-sharing practices among first-party data holders and the associated risk of consumer privacy breaches.

Our research reveals several key findings. Partners tend to share more data with data brokers in more concentrated markets, leading to an increase in data trackers linked to their operations. Interestingly, when both the market and data broker concentrations are high, even partners with significant market power become more cautious in their data-sharing practices with data brokers in more concentrated markets. This highlights a balancing effect between market concentration and data sharing behavior. Additionally, higher market concentrations, both for firms and data brokers, correlate with a decrease in customer privacy breaches. Conversely, firms with lower market concentrations face a higher risk of customer privacy breaches, particularly when combined with lower data broker concentration.

Another significant finding is the influence of a data broker's regulatory status on firms' data-sharing practices. We observe that firms are more inclined to share data with registered data brokers, and this inclination is influenced by the level of market concentration. Furthermore, when both the company and the data broker maintain high market concentration, and the data broker is registered, the likelihood of customer privacy leaks tends to decrease. This underscores the importance of data broker registration and their market concentration in mitigating the risk of customer privacy breaches.

Moreover, our analysis, based on data obtained from the dark web, indicates that registered data brokers have significantly lower amount of customer privacy information leakage. These findings underscore the effectiveness of regulations in preventing data leaks. The substantial market share of registered data brokers reflects their expertise in managing privacy risks and provides a secure platform for data sharing among companies. These results emphasize the critical need for robust data-sharing policies and stringent regulations in today's digital landscape, where the dark web poses a substantial threat to the exposure of sensitive customer information.

However, our study does have several limitations. Our conclusions are drawn from data obtained from various sources such as online code repositories, breach reports, and a dark web monitoring firm and limitations as to these data sources might not fully reflect dynamics across the data supply chain. While these data sources shaped our analysis, we have ensured the robustness and consistency of our findings, making them relevant and applicable.

In essence, our study provides valuable insights into the intricate relationship between market competition, data sharing practices, and consumer privacy breaches. These findings highlight the necessity for industry-specific strategies in managing privacy risks and underscore the importance of considering market structure and competition when assessing the privacy risks associated with data sharing.

References

Acquisti, A., Taylor, C., & Wagman, L. (2016). The Economics of Privacy. Journal of Economic Literature, 54(2), 442–492.

Adjerid, I., Acquisti, A., Telang, R., Padman, R., & Adler-Milstein, J. (2016). The impact of privacy regulation and technology incentives: The case of health information exchanges. Management Science, 62(4), 1042-1063.

Argyres, N. S. (1996). Evidence on the Role of Firm Capabilities in Vertical Integration Decisions. Strategic Management Journal, 17(2), 129-150.

Argyres, N. S., & Liebeskind, J. P. (1999). Contractual Commitments, Bargaining Power, and Governance Inseparability: Incorporating History into Transaction Cost Theory. Academy of Management Review, 24(1), 49-63.

Armour, S. (2014). Data brokers come under greater scrutiny. The Wall Street Journal. https://www.wsj.com/articles/SB10001424052702303874504579377164099831516

Armstrong, M. (2006). Competition in Two-Sided Markets. RAND Journal of Economics, 37(3), 668–691.

Bakos, Y., & Brynjolfsson, E. (1999). Bundling Information Goods: Pricing, Profits, and Efficiency. Management Science, 45(12), 1613-1630.

Barney, J. (1991). Firm resources and sustained competitive advantage. Journal of Management, 17(1), 99-120.

Barney, J. B. (1986). Strategic factor markets: Expectations, luck, and business strategy. Management Science, 32(10), 1231-1241.

Baron, D. P. (2001). Private politics, corporate social responsibility, and integrated strategy. Journal of Economics & Management Strategy, 10(1), 7-45.

Bergemann, D., & Bonatti, A. (2019). Markets for information: An introduction. Annual Review of Economics, 11, 85-107.

Bonardi, J. P., & Keim, G. D. (2005). Corporate political strategies for widely salient issues. Academy of Management Review, 30(3), 555-576.

Braulin, F. C., & Valletti, T. (2016). Selling customer information to competing firms. Economics Letters, 149, 10-14.

Buckman, J. R., Adjerid, I., & Tucker, C. (2023). Privacy regulation and barriers to public health. Management Science, 69(1), 342-350.

California Civil Code, Division 3, Part 4, Title 1.81.5, Chapter 2, Section 1798.99.80 (2018).

Casadesus-Masanell, R., & Hervas-Drane, A. (2015). Competing with privacy. Management Science, 61(1), 229-246.

Cavoukian, A. (2010). Privacy by Design: The 7 Foundational Principles. Information and Privacy Commissioner of Ontario, Canada.

Chandra, A., & Kaiser, U. (2014). Targeted advertising in magazine markets and the advent of the internet. Management Science, 60(7), 1829-1843.

Chen, Z., Choe, C., & Matsushima, N. (2020). Competitive personalized pricing. Management Science, 66(9), 4003-4023.

Choe, C., Cong, J., & Wang, C. (2023). Softening competition through unilateral sharing of customer data. Management Science.

Choi, J. P., Jeon, D. S., & Kim, B. C. (2019). Privacy and personal data collection with information externalities. Journal of Public Economics, 173, 113-124.

Cook, K. S., & Emerson, R. M. (1978). Power, Equity and Commitment in Exchange Networks. American Sociological Review, 43(5), 721–739.

De Corniere, A., & De Nijs, R. (2016). Online advertising and privacy. The RAND Journal of Economics, 47(1), 48-72.

De Loecker, J., Eeckhout, J. Unger, G. (2020) The rise of market power and the macroeconomic implications. Quarterly Journal of Economics 135(2), 561-664.

Eckhardt, G. M., Houston, M. B., Jiang, B., Lamberton, C., Rindfleisch, A., & Zervas, G. (2019). Marketing in the sharing economy. Journal of Marketing, 83(5), 5-27.

EU General Data Protection Regulation (GDPR). (2018). Territorial scope - Article 3. GDPR.eu. https://gdpr.eu/article-3-territorial-scope/

Fainmesser, I. P., Galeotti, A., & Momot, R. (2023). Digital privacy. Management Science, 69(6), 3157-3173.

Federal Trade Commission (FTC). (2022). FTC v. Kochava Inc. Retrieved from https://www.ftc.gov/legal-library/browse/cases-proceedings/ftc-v-kochava-inc

Federal Trade Commission. (2014). Data brokers: A call for transparency and accountability. Retrieved from https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

French, J. R. P., & Raven, B. (1959). The Bases of Social Power. Classics of Organization Theory, 7, 311-320.

Gal-Or, E., Gal-Or, R., & Penmetsa, N. (2018). The role of user privacy concerns in shaping competition among platforms. Information Systems Research, 29(3), 698-722.

Ghosh, A. (2018). The impact of digital technologies on routine tasks: Do labor policies matter? Journal of Economic Perspectives, 32(3), 31-50.

Ghosh, D. (2018). Facebook is changing how marketers can target ads. What does that mean for data brokers? Harvard Business Review.

Glasgow, J. B. (2018). Data brokers: Should they be reviled or revered? The Cambridge Handbook of Consumer Privacy, 25-46.

Goldfarb, A., & Tucker, C. (2011). Privacy Regulation and Online Advertising. Management Science, 57(1), 57-71.

Gordon, L. A., Loeb, M. P., & Sohail, T. (2010). Market value of voluntary disclosures concerning information security. MIS Quarterly, 567-594.

Gu, Y., Madio, L., & Reggiani, C. (2022). Data brokers co-opetition. Oxford Economic Papers, 74(3), 820-839.

Hagiu, A., & Hałaburda, H. (2014). Information and two-sided platform profits. International Journal of Industrial Organization, 34, 25-35.

Harrigan, K. R. (1981). Barriers to entry and competitive strategies. Strategic Management Journal, 2(4), 395-412.

Hitt, M. A., Ireland, R. D., & Hoskisson, R. E. (2014). Strategic Management: Concepts and Cases: Competitiveness and Globalization. South-Western College Pub.

Hoberg, G., & Phillips, G. (2016). Text-Based Network Industry Classifications: A Versatile Framework for Research and Policymaking. Journal of Economic Perspectives, 30(4), 59-88. doi: 10.1257/jep.30.4.59

Ichihashi, S. (2021). The economics of data externalities. Journal of Economic Theory.

Jia, J., Jin, G. Z., & Wagman, L. (2021). The short-run effects of the general data protection regulation on technology venture investment. Marketing Science, 40(4), 661-684.

Johnson, G. A., Shriver, S. K., & 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, G. A., Shriver, S. K., & Goldberg, S. G. (2023). Privacy and market concentration: Intended and unintended consequences of the GDPR. Management Science.

Jones, G. R., & Mendelson, H. (2011). Information Goods vs. Industrial Goods: Cost Structure and Competition. Management Science, 58(1), 166-180.

Ke, T. T., & Sudhir, K. (2022). Privacy rights and data security: GDPR and personal data markets. Management Science.

Kim, J., & Mahoney, J. T. (2005). Property rights theory, transaction costs theory, and agency theory: An organizational economics approach to strategic management. Managerial and Decision Economics, 26(4), 223-242.

Kox, H., Straathof, B., & Zwart, G. (2017). Targeted advertising, platform competition, and privacy. Journal of Economics & Management Strategy, 26(3), 557-570.

Lamdan, S. (2022). Data cartels: The companies that control and monopolize our information. Stanford University Press.

Lee, D. J., Ahn, J. H., & Bang, Y. (2011). Managing consumer privacy concerns in personalization: A strategic analysis of privacy protection. MIS Quarterly, 35(2), 423-444.

Leiponen, A. (2008). Competing through cooperation: The organization of standard-setting in wireless telecommunications. Management Science, 54(11), 1904-1919.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.

Marthews, A., & Tucker, C. (2019). Privacy policy and competition. Brookings Paper.

Martin, K. (2015a). Ethical issues in the big data industry. MIS Quarterly Executive, 14, 2.

Martin, K. (2015b). Privacy notices as tabula rasa: An empirical investigation into how complying with a privacy notice is related to meeting privacy expectations online. Journal of Public Policy & Marketing, 34(2), 210-227.

Mayer-Schönberger, V., & Cukier, K. (2013). Big Data: A Revolution That Will Transform How We Live, Work, and Think. Eamon Dolan/Houghton Mifflin Harcourt.

McWilliams, A., & Siegel, D. S. (2011). Creating and capturing value: Strategic corporate social responsibility, resource-based theory, and sustainable competitive advantage. Journal of Management, 37(5), 1480-1495.

Morgan, R. M., & Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. Journal of Marketing, 58(3), 20-38.

Neumann, N., Tucker, C. E., & Whitfield, T. (2019). Frontiers: How effective is third-party consumer profiling? Evidence from field studies. Marketing Science, 38(6), 918-926.

Newbery, D. M. (1999). Privatization, Restructuring, and Regulation of Network Utilities. MIT press.

Nissenbaum, H. (2010). Privacy in Context: Technology, Policy, and the Integrity of Social Life. Stanford University Press.

Office of the Privacy Commissioner of Canada (PRIV). (2014). Data Brokers: A Look at the Canadian and American Landscape. Retrieved from https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2014/db_201409/

Ponemon Institute. (2014). The economics of the dark web.

Porter, M. E. (1979). The structure within industries and companies' performance. The Review of Economics and Statistics, 61(2), 214-227.

Porter, M. E. (1981). The contributions of industrial organization to strategic management. Academy of Management Review, 6(4), 609-620.

Porter, M. E., & Heppelmann, J. E. (2014). How smart, connected products are transforming competition. Harvard Business Review, 92(11), 64-88.

Prüfer, J., & Schottmüller, C. (2021). Competing with big data. The Journal of Industrial Economics, 69(4), 967-1008.

Puaschunder, J. (2021). Towards a utility theory of privacy and information sharing and the introduction of hyper-hyperbolic discounting in the digital big data age. In Research anthology on privatizing and securing data (pp. 68-111). IGI Global.

Ranganathan, C., & Brown, C. V. (2006). ERP Investments and the Market Value of Firms: Toward an Understanding of Influential ERP Project Variables. Information Systems Research, 17(2), 145-161.

Sarvary, M., & Parker, P. M. (1997). Marketing information: A competitive analysis. Marketing Science, 16(1), 24-38.

Schwab, K., Marcus, A., Oyola, J. O., Hoffman, W., & Luzi, M. (2011). Personal data: The emergence of a new asset class. In An Initiative of the World Economic Forum (pp. 1-40). Cologny, Switzerland: World Economic Forum.

Seetharaman, D., Wells, G., & Vranica, S. (2018). Facebook limiting information shared with data brokers. The Wall Street Journal. https://www.wsj.com/articles/facebook-says-its-ending-use-of-information-from-outside-data-brokers-for-ad-targeting-1522278352

Solove, D. J. (2005). A Taxonomy of Privacy. University of Pennsylvania Law Review, 154(3), 477–564.

Soomro, Z. A., Shah, M. H., & Ahmed, J. (2016). Information security management needs more holistic approach: A literature review. International Journal of Information Management, 36(2), 215-225.

Stahl, B. C. (2016). Responsible research and innovation: The role of privacy in an emerging framework. Science and Public Policy, 43(6), 708-716.

Sun, T., Yuan, Z., Li, C., Zhang, K., & Xu, J. (2023). The value of personal data in internet commerce: A high-stakes field experiment on data regulation policy. Management Science.

Tene, O., & Polonetsky, J. (2012). Privacy in the Age of Big Data: A Time for Big Decisions. Stanford Law Review Online, 64, 63–69.

Tirole, J. (1988). The theory of industrial organization. MIT Press.

UK Information Commissioner's Office (UK ICO). (2020). Investigation into data protection compliance in the direct marketing data broking sector. Retrieved from https://ico.org.uk/action-weve-taken/investigation-into-data-protection-compliance-in-the-direct-marketing-data-broking-sector/

Varnali, K. (2021). Online behavioral advertising: An integrative review. Journal of Marketing Communications, 27(1), 93-114.

Vermont Statutes Annotated, Title 9, Chapter 62, Section 2430 (2018).

Vives, X. (2008). Innovation and Competitive Pressure. The Journal of Industrial Economics, 56(3), 419-469.

Wang, T., Kannan, K. N., & Ulmer, J. R. (2013). The association between the disclosure and the realization of information security risk factors. Information Systems Research, 24(2), 201-218.

Williamson, O. E. (1981). The Economics of Organization: The Transaction Cost Approach. The American Journal of Sociology, 87(3), 548-577.

Zajac, E. J., & Olsen, C. P. (1993). From Transaction Cost to Transactional Value Analysis: Implications for the Study of Interorganizational Strategies. Journal of Management Studies, 30(1), 131-145.